Where are we now ?
Student achievement in Mathematics, Nepali and Social Studies in 2011

Jari Metsämuuronen & Bhoj Raj Kafle (Eds)

National Assessment of Student Achievement
Education Review Office
Ministry of Education, Kathmandu
Nepal
2013

Jari Metsämuuronen & Bhoj Raj Kafle (Eds)

Where are we now ?

Student achievement in
Mathematics, Nepali and Social Studies
in 2011

Where are we now ?
Student achievement in Mathematics, Nepali and Social Studies in 2011

Edited by Jari Metsämuuronen & Bhoj Raj Kafle

Contributed also by Shyam Prasad Acharya, Sunita Shakya, Devi Ram Acharya, Siddhi Prasad Koirala, Hemanga Adhikari, Shree Ram Upadhyaya, Sushan Acharya, Tulsi Thapaliya, Hari Aryal, Maria Metsämuuronen

Government of Nepal

MINISTRY OF EDUCATION

Tel.: {

Ref. No.:

**Singh Durbar
Kathmandu, Nepal**

Foreword

The result of National Assessment of Student achievement (NASA) is instrumental to initiate reform activities in different aspects of education. School Sector Reform Plan (SSRP) has visualized undertaking two cycles of NASA for grade 3, 5 and 8 by 2015. In the process of addressing this SSRP intent, Ministry of Education (MoE) established NASA Unit under Education Review Office (ERO). In this connection, the MoE felt the need of technical expertise and discussions were made regarding different possibilities. To carry out assessment activities, the government of Finland provided technical expert as a consultant to support the program. The MoE prepared a roadmap to address the achievement of the test of grade 3, 5 and 8 students' in two cycles by 2015. This is a comprehensive report of the grade 8 NASA test which includes the first cycle of assessment.

Findings of the study have pointed out various important aspects like educational policy, administration, implementation and teaching learning activities in the schools. I hope, all those concern agencies will pay due attention to address the serious issues and concerns pointed out by the study.

I would like to thank students, teachers, head-teachers, focal persons, district education officers and all ERO personnel who contributed in the study. From the very beginning the financial and technical supports have been provided by the Government of Finland and the Development Partners (DPs) and continuous guide in the process of the study by the technical committee of ERO. I would like to express my sincere gratitude to both the Finnish government and all the members of technical committee.

Lastly, in the process of educational reform, NASA is one of the most important and regular activity of MoE, in this connection we will be highly grateful to receive constructive feedback regarding the process and quality of the NASA report.

Narayan Gopal Malego
Secretary
The Government of Nepal
Ministry of Education

Contents

Mr. Shyam Prasad Acharya, Dr. Jari Metsämuuronen, & Ms T. Maria Metsämuuronen:

Dr. Tulshi Thapaliya & Dr. Jari Metsämuuronen:

Dr. Jari Metsämuuronen:

Dr. Jari Metsämuuronen, Mr. Shyam Prasad Acharya & Mr. Hari Arial:

Dr. Bhoj Raj Kafle & Dr. Jari Metsämuuronen:

Preface and acknowledgements

This report is, in many ways, a joint effort to publish high-level accurate information on the national achievement level at grade 8. First, by using international assistance within the Ministry of Education (MoE), a semi-independent unit was set up within the Education Review Office (ERO). This was a result of long discussions between the Development Partners (DP) and the MoE, without which the product would not have started nor nurtured.

Equally, without the mutual agreement between the MoE and the Ministry for Foreign Affairs in Finland (ForMin), it would not have been possible to send a Finnish expert to manage the process. From Finland's side, Ms. **Helena Kuivila** in ForMin and Ms. **Satu Pehu-Voima** and Mr. **Bhola Prasad Dahal** in Finland's Embassy in Nepal were the key persons to initiate the negotiations. Key actors in the MoE were Secretary Mr. **Sankar Prasad Pandey**, Joint Secretaries Dr. **Lava Awasthi**, Mr. **Janardan Nepal,** Mr. **Mahasram Sharma**, and Under Secretaries Dr. **Nakul Baniya,** Mr. **Divakar Dhungel,** and Mr. **Kamal Pokharel** in the ERO.

Third, the work was strongly supported by the DPs, from the beginning to the end. In particular, strong support was given by Finland's Embassy to address many administrative challenges. This help proved invaluable. The study visit to Finland was also supported by the Embassy. Similarly, the support from the World Bank (WB), in the first phase by Dr. **Venkatesh (Venki) Sundararaman** and at a latter phase by Dr. **Saurav Dev Bhatta** and Ms. **Shradda Shah**, and from the Asian Development Bank (ADB) in the first phase by Ms. **Smita Gyawali** and at a latter phase Dr. **Bhuban Bajra Bajracharya** and Dr. **Pramod Bhatta** was also indispensable. Not only was their role remarkable in the negotiating of the permanent staff, but, for example, in commenting on the drafts of this report before finalizing it. The final part of finalizing this report was completed with financial support from the WB (by paying the compensation for the consultant) and Finland's Embassy (by paying for the final editing, layout and printing of the report). In the Embassy, Mr. **Jussi Karakoski**, Ms. **Marianne Kujala-Garcia** and Ms. **Indra Gurung** were mainly responsible for this phase.

Fourth, the work was supported in many ways by the Steering-, Technical-, and Subject committees. The Steering committee made the policy decisions and was led by the current Secretary. The Technical committee, led by Dr. **Ishwar Prasad Upadhaya**, made the final decisions on the assessment program and the subjects to be assessed. It took the principle responsibility for this task. The practical responsibility was given to the

Subject committees lead by Prof. Dr. **Hari Upadhaya** (Mathematics), Prof. Dr. **Kadhar Prasad Sharma** (Nepali), and Prof. Dr. **Bhuchanda Baidya** (Social Studies). The item writers were led by Prof. Dr. **Hari Upadhaya** (Mathematics), Prof. Dr. **Madhav Paudel** (Nepali), and Prof. Dr. **Shareeram P. Upadhaya** (Social Studies). During the process, their role was very important, not only in taking care that the items fulfill the content-wise validity and the process of item writing itself but also in discussions when altering the items before the pretest. The Subject committees took an active role in preparing and approving the final tests as well as giving support in commenting on the texts.

Fifth, There was remarkable contribution of ERO chiefs who were working during the process – **Arjun Bahadur Bhandari, Chitra Prasad Devkota, Divakar Dhungel, Kamal Prasad Pokhrel, Kamleshwor Kumar Sinha** and **Jivan Sharma Paudel.** Without a strong team spirit within the ERO, the work would have been difficult to carry out. All the officers were pre-testing the items and monitoring the process all across the country. They helped in many ways in the item writers' seminars as well as the difficult language issues by critically reading the papers and letters. The small and active team within the National Assessment of Student Assessment (NASA) unit under the ERO – Mr. **Hari Aryal,** Mr. **Shyam Prasad Acharya**, and **Gyanmani Nepal** succeeded by Ms. **Sunita Shakya**, Mr. **Devi Ram**, and Mr. **Prakash Karel** – did a remarkable job creating the system of collecting information in the Nepalse context and organizing item writers' seminars, pretesting processes, and the final testing as well as preparing this final report.

Sixth, the process involved hundreds of teachers and head teachers and tens of District Education Office civil servants. Without them, no reports would be written. Remarkable work was also done by the private company, led by Dr. **Tirtha Raj Khaniya**, which was responsible for the data collection, marking and data entry. At the writing phase, Dr. **Sushan Achrya**, Prof. Dr. **Hemanga Adhikari**, Dr. **Siddhi Koirala**, Ms. **Maria Metsämuuronen**, and Dr. **Tulshi Thapaliya** gave valuable insights through their contribution. The final proof-read of the final text was done by **Jovan Ilic** from the British Council.

Many more names could have been mentioned starting from the countless resource persons in the ERO. They all did more or less invisible work in order to help the main team in NASA. We are cordially thankful for all of them.

20.10.2013

NASA, ERO, MoE,

Sanothimi, Kathmandu, Nepal

Dr. **Jari Metsämuuronen**

Adjunct Professor

Dr. **Bhoj Raj Kafle**

Under Secretary

Executive Summary and Discussion for Implementation

Context

At the beginning of 2012, the unit for National Assessment of Student Achievement (NASA), within the Educational Review Office (ERO) in the Ministry of Education (MoE) in Nepal, assessed the learning outcomes of grade 8 students in three subjects: Mathematics, Nepali, and Social Studies. The learning outcomes concerned the fiscal year 2011, and altogether 48,682 students, 1,156 subject teachers, and 1,158 head teachers from 1,201 schools (randomly selected and stratified) participated in the assessment. Of the students: 23,743 were boys and 24,818 girls. Out of the 75 districts of Nepal the dataset represents a random selection of 25 with all Development Regions (Eastern, Central, Western, Mid-Western, Far-Western, and Kathmandu Valley) accounted for, 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 represented so that the results of the assessment can be credibly extended to the whole student and school populations of Nepal.

Three versions of the final tests in each subject were administered, with the final scores equated by IRT modeling. The reliability of the tests is high and validity is assured by using the specification grids of the local curriculum as prepared by the National Examination Board. From a methodological standpoint, the process and practices of the inquiry successfully followed the procedures as used in Finland, but contextualized to Nepalese reality. The results were linked to the set of results from the 2008 assessment as well as to the international item banks of TIMSS (Trends in Mathematics and Science Study) and PISA (Programme in Student Assessment).

The results are reported in the main 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%. The Mathematics test, however, proved to be more difficult (average 43%) than the other tests (average 49%), so in the final large student dataset the scores were given an additional equating by shifting the means to the same level following a procedure of 'mean equating', with the final average in the common dataset set at a result of 46.9%. These figures set the norm from which

the different student subpopulations are compared. For the Nepali language assessment, the Common European Framework of Reference for Languages (CERF) was also used to obtain the absolute level of the students from a language achievement level point of view.

List of Major Findings

- Differences: There are great differences in achievement between the students, schools, districts, and developmental regions.

- Institutional/Community Schools: The students in institutional schools perform well. Within the community schools, there is a wide range in average student achievement. The difference between the low and high performing community schools is remarkable.

- Cognitive skills: Students are apt in tasks related to memorization and recall, but are not effective in skills requiring application or ability at a higher cognitive level.

- Socio-economic background: The low educational and social background is directly and strictly related to low results.

- Caste/ethnicity/home language: Achievement level still depends on caste/ethnicity as well as on the home language of the student.

- Similarities: There are no remarkable differences between boys and girls, rural and urban schools, and across ecological zones.

- Three content areas: In Nepali, achievement in reading and writing is low in absolute terms – an average student of grade 8 cannot read and write sufficiently well to manage higher studies for example. Achievement in Mathematics is not distributed normally, and in Geometry and Sets it is remarkably low when compared with the other content areas. In Social Studies, the achievement levels in Politics, History, and Civics is lower than the other content areas.

- 2008/2011: The results in Mathematics have slightly declined from those of 2008 while the results in Nepali have increased. Reading skills in Nepal are lower than at the international level.

Major findings and discussion areas for implementation

1. Differences: There are great differences in achievement between the students, schools, districts, and developmental regions.

The study confirms that there are great differences between the achievement levels across students and schools in Nepal. Some students in the survey did not

give a correct answer to any single test item (0%) while the best students gained more than 95%. The average achievement of students in the lowest performing schools was below 15% while the students in some of the highest performing schools gained, on average, over 90%. When knowing that the average score in a school is below 15%, it means that in those schools many students fall below that level. From an equal opportunities viewpoint, it is not a good sign that 'low' and 'high' performing students are concentrated in certain schools.

Not only is the variation great amongst students and schools, but it is also significant across the 25 districts covered by the study. In the five lowest performing districts (Bhojpur, Jajarkot, Morang, Kanchanpur, and Ilam), the average performance was below 40%, while in the three highest performing districts (Lalitpur, Kathmandu, and Bhaktapur) it exceeds 55%. The students in the Kathmandu Valley schools (59%) exceed the average students in the other regions (41–49%) across numerous indicators. The achievement level of the students from the Eastern (41%), Far-Western (42%), and Mid-Western (43%) development regions is far behind that of the students in the Kathmandu Valley. Once again from an equal opportunities standpoint, this is not a positive sign.

Further analysis of the highest and lowest performing community schools gives some hints as to the possible reasons behind the deviating results. (1) The number of study days in the highest performing schools is higher than in the lowest performing schools. (2) The higher performing schools are smaller (388 students or less) than the lowest performing schools (461 students or more). (3) The student/teacher ratio is smaller in the highest performing schools. (4) The highest performing students spent more time on homework, they were less absent from school, more of them did not need to work while studying, they were younger, and their attitude towards the subject was positive. (5) Parents' illiteracy – especially that of fathers – is prevalent within the lowest performing students.

Discussion and suggestions: Such a significant difference in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. Though lots of work has been done to reduce the gap, there still would appear to be a substantial amount of work to do. As a minimum, the confirmed state of educational inequity across the country across most indicators should at least initiate discussions, within and between the stakeholders, as to how the difference in average achievement among the schools can be successfully reduced.

The possible reasons behind the deviation indicated above, provide a starting point for such discussions. The lower number of study days challenges the stakeholders to consider whether the main reason behind this indicator could be the number of strikes and disturbances in these regions due to which the schools remained closed for several days. If this was the reason for the low number of

study days, it would urge teachers' unions and other partners to find more positive ways to seek justice without harming the opportunities of the children to gain constant access to the education system. It also challenges all district education offices to detect those schools which are at danger of becoming one of the poorest performing schools. The possible reasons for the low performance should be identified and actions should be taken to address this. If more resources are needed for raising the standard in these lowest performing schools, it might be recommendable to create a system within the Ministry of Education to distribute grants to these schools, in a spirit of positive discrimination, for example.

2. Institutional/Community Schools: The students in institutional schools perform well. Within the community schools, there is a wide range in average student achievement. The difference between the low and high performing community schools is remarkable.

The average achievement in institutional schools is remarkably higher (63%) than in the community schools (44%). The comparison is, nevertheless, unfair because the average socio-economic status (SES) is more than twice as high in the private schools (74%) as compared with the students at the community schools (34%). This means that the educational, social and economic capacity in these families to provide private tuition, for example, is much higher in the private schools. Hence, the higher results can be explained by a strict selection of the students and by tuition given by some external source outside of the schools, and not necessarily by the better processes within the school itself.

What is important is that there are a significant number of community schools where the average results are at the same level as in the private schools even though the SES is radically lower. In these schools, either the processes are more effective than in the private schools or the students are of the same ability as those in the private schools and are not adversely affected by the processes within the school or their socioeconomic status.

The analysis based on comparing the highest and lowest performing community schools showed that in the highest performing schools teachers are using more homework evaluation and classroom evaluation, they have more interaction with other teaching staff, and they have a more realistic perception of their students learning capacity. In the highest performing schools, there was more interaction with the community, the student/teacher ratio was smaller, there were fewer incidents of being late in schools as well as destroying school property, and there was more interaction with teaching staff with regard to improving teaching and learning processes. The students in the highest performing schools spend more time on homework, they are less absent, they do not spend too much time playing and doing sports, or on home chores, and they do not undertake paid work, nor are they bullied as much as their counterparts in the lowest performing schools.

An analysis of all community schools shows that the better results were achieved when the student/teacher ratio was lower than 31 students, the number of study days was 200 or more, when there are computers for mathematics teaching and staff for maintenance and technical assistance, when there are few or no incidents of being late for school, or of the destruction of school premises, when there are interactions with the teaching staff on improving teaching and learning processes more than seven times a year, and when there are classroom evaluations more than twice a week.

Discussion and suggestions: A major lesson which can be learnt from the findings is that many of the community schools are capable of maintaining a high level of achievement despite the poor socioeconomic status of their students. These schools seem to use more effective processes than the private schools because they can reach similar levels of achievement with significantly fewer resources. On the other hand, there are many community schools where the average performance is very low. These schools with a similarly low socioeconomic status need to learn from these other comparable schools about the processes they are utilizing to achieve these high results.

A comparison of the highest and lowest performing schools gives some hints as to the areas to focus on: the low number of study days, absenteeism in schools, students' paid work, less effective practices in home work and classroom evaluation, as well as low communication between the community and teaching staff.

These kinds of activities and checklists provided in the report may be used by the district officers when monitoring the schools. Not only the schools but also all those concerned with educational development and management including policy makers, implementing and monitoring agencies, curriculum developers, training providers and managers, should, at least, be aware of these differences between the highest and lowest performing community schools.

What activities are carried out and how learning takes place in such schools is a matter of discussion at the policy making level. It is urgent that such issues be discussed at the policy making level and solutions to the problems sought. All possible efforts should be made to reduce the differences between these schools.

3. Cognitive skills: Students are apt in tasks related to memorization and recall, but are not effective in skills requiring application or ability at a higher cognitive level.

Students are comparatively poorer in the ability to solve problems, to analyze, deduce logic, generalize, justify an argument or viewpoint, and in the ability to transfer learning from one context to another. A remarkably high number of students were able to solve only 15% or less of the practical problems, (17% of the students in Mathematics, 13% in Nepali, and 8% in Social Studies).

In Mathematics, students are able to do basic calculations, but are weak in reasoning, problem solving, plotting, 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.

In Nepali, students performed well when called upon to recognize the correct answer, in recalling simple facts from the texts, fundamental thinking, the basic interpretation of a paragraph, tables, charts, and in logical thinking that required only a few steps. However they are much weaker in producing fluent texts or letters, and in preparing synthesis and abstracts from a text. In Nepali, the students did attempt open-ended tasks but the skills were not high enough for the highest marks.

In Social Studies, students are good in recognizing the correct answer, in very fundamental knowledge/content, in true or false questions, matching texts, and in the selection of words for gap fill activities. They are much weaker in reasoning, problem solving, and in constructing arguments.

Contemporary thinking is that constructive and child-centered methods may help with higher cognitive tasks. The comparison of Nepali and Finnish teachers' perceptions of good teaching shows that the teachers in Nepal are aware, at least in theory, of the same constructivist principles of teaching and learning as the Finnish teachers. However, in Nepal, the classroom reality of large class numbers, sometimes in excess of 100 students in a single class, and excessive classroom sessions per day do not encourage teachers to use student centered teaching methods, which may in fact be a necessity for students to develop their higher ability skills'. The reality of the classrooms in many cases is that teachers lecture, although interestingly the teachers themselves do not necessarily see it this way, suggesting a possible lack of understanding of the constructivist principles despite their claims to be aware of them.

Discussion and suggestions: The educational system in Nepal would appear to be geared toward remembering items rather than solving problems creatively. This begs the following questions: What are the reasons for this culture? How is it nurtured? What could be done to change this culture? What kind of citizen is the education system producing?

Traditionally, in South Asia, wisdom has been transmitted orally with the onus on remembering by heart in order to pass knowledge on. Once introduced, the written text was not taken to be as reliable as that of the spoken word, and hence, the tradition continued to favour remembering teachings by heart. This long-lived oral tradition connected with the deep-rooted culture of idolism, has given the teacher the position of a 'trusted' guru. The students are taught at a very early age to trust the teacher without question. Students are expected to receive

this knowledge passively without opportunity for inquiry or exploration on their own account (in the modern constructive psychological manner: children are not given possibilities to create their own realities).

Consistently large classroom sizes of 40–100 students, as well as the heavy-loaded curricula, nurture teacher-led teaching methods. To complete the whole curriculum is not an easy task if the teacher is expected to cover all areas.

The analysis of effective class size in Nepal shows that the optimum student/teacher ratio, from an achievement point of view, is 22 students or less per teacher (mean score 55%), the lowest results come from schools where the student/teacher ratio is more than 31 students (43%).

As a comparison, in Finland, the student/teacher ratio is 13.5 and the maximum class size is usually 25 students. A median classroom size for 1st–6th graders is 60 m^2 for 25 students, which equals 2.5 m^2 for each student to work and move in. Practically speaking, all Finnish students have their own desk with half a meter of space around their work area in which to move. To change the size of classroom is not an easy task. It requires more teachers, new school buildings, new ways of thinking of the students' learning environment, as well as the proper maintenance of premises.

Similarly, though the teachers claim to know the meaning of continuous evaluation in class it is doubtful to what extent it could really be applied when there are so many students within one classroom.

The curricula, also seems to be loaded with detailed tasks to fulfill. If this is feeding the teacher-centered methods in schools, it may be worth rethinking the role and form of the curricula. Perhaps the curricula could be shorter with more guidelines as it is in Finland rather than an exhaustive teachers' guide as it is in Nepal. Or training on how to employ the curricula in a creative way may be necessary. Either way, the pedagogy and its practical application in the classroom should be changed to allow for more creative thinking, reasoning, and application.

4. Socio-economic background: The low educational and social background is directly and strictly related to low results.

The dataset gives a strong indication that Socio Economic Status plays a significant role in the educational processes in Nepal. The difference in the achievement level of the students from the lowest (41%) and highest SES group (64%) is remarkable. One of the indicators for the SES was parents' educational level and specifically parents' illiteracy. In the dataset, out of 1200 schools, there are 73 schools where more than 40% of grade 8 students reported to have an illiterate father and 609 schools where more than 40% had an illiterate mother. 82 schools have a mothers' illiteracy rate of more than 80%, and there were 18 schools where all the mothers were reported to be illiterate. Many students from illiterate

families can reach a high achievement level, however, when the parents are, or the father alone is illiterate, the probability of being within the highest performing students is quite low (39% or the highest level students) compared with the situation when both the parents are literate and they are working outside of agriculture (64% of the students).

When there are more than 10% of students with an illiterate father in the school, the average result of the school is significantly lower (46%) than if there was less than 3% illiterate fathers (62%). Correspondingly, when there are more than 54% of the students with an illiterate mother in the school, the average result of the school is significantly lower (43%) than if there was less than 16% illiterate mothers (60%). When both parents were literate, the achievement level was 63%.

Another aspect for the impact of the illiteracy is that by using the illiteracy rate in the school, it is quite easy to provide a rough indicator for assessing whether the school belongs to the lowest level schools (the lowest quintile). If there were 8% of the students with illiterate fathers in a school, the probability of it belonging to the lowest quintile is 65%. In parallel, if 40% of the students had an illiterate mother, the probability of belong to the lowest level schools was 38%. If more than 60% of the students had an illiterate mother, the probability would be 60%. That is, by simply asking the students how many of them have an illiterate father, we have a clear indicator of whether the school is one of the lowest performing. If one out of ten students raise their hand, the school most probably is one of those with very low results. The same can be inferred by asking the number of illiterate mothers in the school; if six out of ten students raise their hand, the school most probably is one of those with very low results.

Poverty is seen in another way in the data. The students who need to carry out paid work during their school time perform statistically lower (40–41%) than those without a need for paid work (50%).

Discussion and suggestions: The educational and social capacity of the family may impact on the child in several ways. First, the parents' education and especially their literary skills lay the ground for the child's early development, in particular on their range of vocabulary, and on the child's motivation to learn. Second, more educated parents are most probably more motivated to push their child to study harder in order to improve, or at least to maintain their social position. Third, educated parents understand the value of education in raising the standard of their child's future. Fourth, educated parents can give tuition to their children in the early years of schooling which may help the children radically in these years. Several other advantages could also be found, however, it is not easy to change the social structure in a society without realistic possibilities of finding employment or a place of study after primary education.

Parental illiteracy is most probably malignantly tied to the occupational background of the parents, a low motivation to educate their children, poverty, or, in some cases, even to the caste of the family. All possible efforts should be made to raise the level of literacy among the parents of the young children.

Another question related to illiteracy, is how to reduce the negative effect of poverty on educational achievement. It is assumed that families with very low incomes are not motivated to send their children to school because the children are needed in making a living. If the children were taken care of, by offering them, for example, one warm lunch and a small snack during the school day, it may reduce the need for paid work and so motivate the parents to send the children to school. All the children in all the schools would need this support regardless of the achievement level of the schools. However, it may be reasonable to start the aid in those schools which are detected as the lowest level schools. The district offices could detect these schools and national grants could be addressed to them.

5.   Caste/ethnicity/home language: Achievement level still depends on caste/ ethnicity as well as on the home language of the student.

The achievement level of the Dalit students is found to be lower in all 3 subjects. Dalit girls especially seem to be low-achievers in Mathematics (35%), Nepali (35%), and Social Studies (45%), as compared with the Dalit boys (39%, 40%, and 48% respectively). It is worth noting that in the Valley, Dalits' results are much higher than the average for the country.

Madhesi students' performance in Mathematics is higher than the average (49%), but the results in Nepali (45%) and Social Studies (47%) are lower than the average. In the Madhesi population, the girls seem to be behind the boys especially in the Mountain zone (44% against 50%), in the Hill zone (42% against 50%), and to some extent in the Tarai zone (42% against 45%).

With regard to the Nepali language, using the standard setting of the CEFR levels, reading proficiency was remarkably lower in the Madhesi population than in all other castes. 40% of the Madhesi' students fell into the category of "lower than A2.2". This is a remarkably high value when compared with the other castes with 21–28% of students in this category. This indicates that they cannot even "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)" as defined in the nuanced CEFR classification. The result means that almost half of the Madhesi students are at the level where their Nepali reading ability is limited to "simple texts containing the most common vocabulary". Also, in the language groups of Limbus and Tharus there are far too many students at this low level (42% and 35% respectively) and a high proportion in the Magar (30%) and Gurung (28%) groups too.

With regard to writing skills, it is alarming that 51% of Madhesi students are classified at the level A2.1 or lower. This means that they can (only) write "brief, simple messages (personal letters, notes), which are related to everyday needs" in Nepali, but they could not, for example, "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) (level A2.2). At this low level there are an astonishing 100% of Yadavs, 40% of Limbus, 36% of Tharus, and 33% of Rais. The number of Yadavs and Rais are, however, very small in the data and, hence, no deep analysis or interpretation can be done.

Discussion and suggestions: The skills of reading and writing are essential for reaching a study place in Higher Education. In all language groups, especially among Newars, Tamangs and Urdus, there are many exceptionally good writers. However, it is evident that a notable percentage of students have not reached the required level of reading and writing of Nepali to be able to continue in Higher Education in the Nepali educational system. From an equality viewpoint, the educational system in Nepal is not able to offer equal opportunities for all students to reach the same level in Nepali as already discussed above.

There may be natural reasons for this, as historically the Madhesi population have had a very strong connection to the Indian Madhesi population and they may think that learning Nepali is not useful when most connections are to be with Indians. However, there are many other small populations which evidently need more support in their language learning. In any case, from a national cohesion viewpoint, it is important to think of possibilities of integrating the strong Madhesi population into Nepali society. One option is to offer the minority populations possibilities to study mainly with their own language and to study the Nepali language as an obligatory second language. This is the manner in, for example, Finland, where there are three official languages and where the instruction language in the schools depends on the official language of the school. This suggestion is not easy to apply because it will lead to a situation where there should also be study places in the Higher/tertiary education with the minority languages, too. Hence, this would prove quite a radical change in society. The advances and risks of the change should be carefully surveyed by the National Planning Commission, as the number of non-Nepali speaking groups is so large that providing mother tongue language instruction in schools to some or one of them and not others will have potentially massive political repercussions.

6. Similarities: There are no remarkable differences between boys and girls, rural and urban schools, and across ecological zones.

Though the boys are slightly out-performing girls in all content areas of Mathematics the differences are not significant. Generally speaking, in all subjects, there seems to not be any remarkable difference in achievement between the boys (48%) and girls (46%) who have reached grade 8. Although the results are more positive in

cities (52%, without Kathmandu Valley 47%) than in the rural areas (45%) the difference again is not remarkable. With the exception of Kathmandu Valley, there are no differences between the average achievement levels of the students from the Mountain (46%), Hill (45%), and Tarai (45%) ecological zones. In fact they are remarkably consistent. From an equity point of view the signal is positive, although there is work to be done to reduce the gap between Kathmandu Valley and the other areas.

Discussion and suggestions: The results indicate positive signs with regard to educational equity between sexes, the locality of the school, and ecological zones. The results do not, however, mean that there is no need to follow-up on these indicators in the future. Another aspect is that there might be remarkable deficiencies in equity when it comes to certain subgroups in ecological zones and school location as well as between sexes. One of these issues is that of the ethnic groups; for example, among Madhesi students, the achievement of the girls is noticeably lower than that of the boys. It is also known that many non-Nepali speaking children drop out in early grades. The main suggestion is that, in the future, gender equity is stressed more at the lower grades rather than at the higher grades. However, it is still important for the Curriculum Development Centre to review and analyze its educational materials from a gender equity viewpoint. Likewise, current efforts for teacher training also require a review from the gender perspective. Each school needs to analyze the annual achievement level from a gender perspective and to adopt measures for improvement. Teachers need to be sensitive towards this issue and carry out appropriate learning activities.

The educational opportunities are much greater in the Kathmandu Valley than in the remaining part of the country. It would appear that the main intellectual capacity of the country is concentrated in the Valley, mainly because of the wide possibilities to both study in the universities and work in high level business. Hence, it is wise to think how the intellectual capacity could be protected and enhanced all over the country and not only in the Valley. If educated parents in other areas find out that their children are able to get the same high scores as are received in the Valley, it may halt the uncontrolled shift of the population to the Valley.

7. Three content areas: In Nepali, achievement in reading and writing is low in absolute terms – an average student of grade 8 cannot read and write sufficiently well to manage higher studies for example. Achievement in Mathematics is not distributed normally, and in Geometry and Sets it is remarkably low when compared with the other content areas. In Social Studies, the achievement levels in Politics, History, and Civics is lower than the other content areas.

In Nepali, the Common European Framework of Reference for Languages was employed to identify the absolute achievement of students in the language skills of

reading and writing. It was found that, although there are many good readers and writers in the sample, an average 8th grader reader of Nepali cannot read and understand newspapers independently. The risk of being a weak reader is higher in the language groups of Tamangs, Tharus, Newars and Limbus, and especially so amongst the Madhesi population. An average 8th grader writer of Nepali cannot make, for example, lecture notes or brief summaries independently.

8% of the students are exceptionally weak in writing with the risk being very high once again in the language groups of Tharus, Limbus, and Gurungs, and especially so in Madhesi societies.

In Mathematics, the population is not normally distributed as it is in the Nepali and Social Studies assessments. The distribution suggests that there are three distinctive student populations: 'low', 'mediocre', and 'high' performing students. The low-performing students where the majority of the students lie, achieve on average, a score of 20–25%, the medium-performing students 40–50%, and the high-performing students as high as 70–80%. Of the content areas, the achievement is lower in Geometry (37%) and Sets (38%) than in the other areas (48–49%). In Social Studies, the achievement is lower in Politics (46%), History (50%), and Civil Society (51%) as when compared to Economics (58%) and Geography (57%).

Discussion and suggestions: Grade 8 students have not developed their reading and writing skills in Nepali to a level required to access Higher Studies. In particular, they are weak in writing essays, reports, and summaries, and in the ability to give an opinion, and to express their own views and those of others. It is clear that many of these students may not be able to access Higher-level studies due to this lack of skills. From an equality viewpoint, the educational system is not able to offer equal opportunities for all students to reach the same level in Nepali. This has a strong predictable implication for the enrollment rate in Higher Education, as at present from certain minority populations it is practically impossible to rise to the level of studying in a Nepali Higher Education institution.

On the basis of the pretests for the reading test, students were not, in general, used to reading longer texts or in expressing their opinions on the basis of the read text. At present only 10% of the time allocation at grade 8 is given to reading whereas 65% is given to writing, and 25% for grammar and vocabulary combined. The validity of this weighting therefore needs review.

As Nepali is a second language for more than 30% of the students in the sample, it seems appropriate for the pedagogy to be developed accordingly, such that Nepali is not taught as a mother tongue to these students but as a second language, else these language minorities will continue to suffer from poor academic performance. Learning Nepali may in many cases be too demanding for a non-Nepali student if the teaching methods used are directed at Nepali speakers, and

do not cater for second language learners. In many cases, this is the reality. In addition, there is the issue of the capability of the teacher as this challenges those teachers who themselves are not speaking Nepali as their native language as it is significantly more challenging to teach children to speak, read, and write Nepali if the teacher is not himself fluent.

In the Mathematics test, 5% of the students scored less than 15%. These students will have severe challenges in their possibilities to enter Higher Education, not to mention their everyday life numeracy. The same students almost certainly scored very low at the lower grades too, suggesting a need to identify low-achievers early on and to make additional arrangements to cater for their particular needs.

8. 2008/2011: The results in Mathematics have slightly declined from those of 2008 while the results in Nepali have increased. Reading skills in Nepal are lower than at the international level.

The average achievement level in Nepali was 49%, which is 2% higher than in the year 2008. In Mathematics, the score of 43% was 4% lower than in 2008. The Social Studies score of 49% is not comparable with the 2008 results because of obvious differences in the marking schemes of the linking items.

In all content areas, Mathematical skills are slightly lower in the year 2011 as compared with 2008. This is clearly seen in most linking items, where the average performance in 2011 is 8% lower than in 2008. The Nepali results have increased, on average by 2% within the four years. The level has especially risen in Writing (+16) and Reading (+6) but it has dropped dramatically in Vocabulary (-22). On the basis of the dataset it is not possible to say anything firm with regard to development in Social Studies between the years 2008 and 2011, due to the apparently different marking schemes.

The dataset gives a signal that the average reading proficiency in Nepal is much lower (-0.88) than the international average (0.00). This is supported by the fact that the CEFR levels show a low achievement in reading. According to the dataset, the average achievement of Mathematics and Geography in Nepal seems to be slightly better than the international average (+0.27 and +0.31 respectively). Though the results are obtained by using a good number of linking items to the international item bank, there are mitigating factors, such as the poor level of India in the TIMSS 2011 survey, so more rigorous studies are needed to confirm the results.

Finally, a comparison of the same test items for the 2008 and 2011 samples shows that the students' ability to solve these problems has reduced. This is possibly an indication that there are more weak students enrolling in studies because of a successful "Education for All" program.

Discussion and suggestions: The change between 2008 and 2011 does not necessarily need any action. It is, though, a good sign that the reading and writing

proficiency levels seem to have been increased within these three years. It is more difficult to explain why the results in Mathematics have decreased. It may be possible to explain part of this by a stricter marking scheme in 2011, but it is too early to give any recommendations concerning the result. NASA 2013 testing will give more comparable evidence of the change; as then the 2013 results can be compared with both the 2011 and 2008 results.

It is worth mentioning that the equating procedures used in Nepali, Mathematics, and Geography was identical. Hence, there is no reason related to the equating process which would explain the deviating results between the subjects. However, one technical aspect may explain the incredibly high results in Mathematics and Geography as compared with Nepali. Namely, hindsight shows that the selection of the items from the international item bank differed, maybe crucially, between the subjects. In Nepali, the reading items were selected without any connection to curriculum while the Mathematics and Geography items were selected on the basis of the local curriculum. The consequence seems to be that, in the Nepali reading content area, the equated achievement level corresponds better to the international level while in Mathematics and Geography, the obtained results are better than if items were randomly selected from the international bank. This may be a remarkable lesson to learn for the next NASA rounds.

Suggestions and recommendations classified by stakeholder

This section comprises the suggestions and recommendations and presents them as classified by the relevant stakeholders.

National planning commission

- The educational system in Nepal is geared towards the art of recall rather than higher cognitive processes such as problem solving. *What kind of citizens is the Education System producing?*
- The large class sizes of 40–100 students, as well as the heavy-loaded curricula, nurture a teacher-led or teacher-centered teaching methodology. *What kind of teachers is the Education System producing?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*
- The educational opportunities are much higher in the Valley than in the remaining part of the country. The main intellectual capacity of the country is at present concentrated in the Valley. *How to address this imbalance?*
- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

- Parents' education and especially their literary skills would appear to be a significant factor in achievement levels in Nepal. *How to raise the literacy levels among the parents of the poor performing children, as well as the children themselves?*
- It is assumed that families with very low incomes are not motivated to send their children to school because the children are needed to help make a living. *If the children were taken care of by offering them, for example, one warm lunch and a small snack during the school day, it may reduce the need for paid work and so keep these children in school?*
- From the national cohesion viewpoint, it is important to discuss the possibilities of integrating the strong Madhesi population as well as the large Tharu society into common Nepali society. *One option is to offer these minority populations the possibility to study mainly in their own language at the lower grades and to study the Nepali language as an obligatory second language?*

Department of education (DOE)

- Low achievement level of many students in Mathematics. *Need for a supportive national program for the slow learners?*
- Teaching pedagogy and practices in the classroom focus on learning by rote teacher-centred activites. *How to allow for a more student-centred approach focusing on areas such as creative thinking, problem solving and deductive reasoning?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*
- Low achievement level of many community schools. *Need for a supportive national program to positively discriminate in favour of these schools?*
- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

Teacher educators

- Low achievement level of many students in Mathematics. *Need for a supportive national program for the slow learners with a new pedagogy?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

Curriculum developers

- Low reading and writing skills on average in Nepali, and specifically in certain language minorities. *How to reduce the gap between different groups in language proficiency?*
- Students are not used to reading longer texts or to expressing their opinions on the basis of the read text. *Should the allocation for reading be higher than the 10% currently specified?*
- Very low results of the main student population. *Is the curriculum so loaded with material it is causing difficulties for the teachers?*
- The educational system in Nepal is geared toward remembering things rather than solving novel problems. *Should the curricula be shorter, with more guidelines rather than an exhaustive teachers' guide?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*
- No remarkable differences in achievement between boys and girls. *Still important to review and analyze educational materials from the gender equity viewpoint to maintain this?*
- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

Teacher trainers

- Low reading and writing skills on average in Nepali, and specifically in certain language minorities. *How to reduce the gap between different groups in language proficiency?*
- Nepali is a second language for more than 30% of the students in the sample. *Should the pedagogy be developed, such that Nepali is not taught as a mother tongue to these students but as a second language?*
- Very low results of the main student population. *Should there be more emphasis on support with remedial teaching for those who are in danger of dropping out of Mathematics in the early grades?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*
- No remarkable differences in achievement between boys and girls. *Still important to review and analyze educational materials from the gender equity viewpoint to maintain this?*

- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

School leave certificate (SLC) board

- The educational system in Nepal is geared towards the art of recall rather than higher cognitive processes such as problem solving. *Should the SLC Examination be rewritten to promote higher cognitive processes?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

Non-formal education centre

- Parents' education and especially their literary skills would appear to be a significant factor in achievement levels in Nepal. *How to raise the literacy levels among the parents of the poor performing children, as well as the children themselves?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

District education offices (DEOs)

- Low reading and writing skills on average in Nepali, and specifically in certain language minorities. *How to reduce the gap between different groups in language proficiency?*
- Low level of Nepali in certain language minorities. *How to support teachers with low levels of Nepali to teach in the medium of Nepali?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*
- Low achievement level of many community schools. *Need for a supportive national program to positively discriminate in favour of these schools?*
- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

- It is assumed that families with very low incomes are not motivated to send their children to school because the children are needed to help make a living. *If the children were taken care of by offering them, for example, one warm lunch and a small snack during the school day, it may reduce the need for paid work and so keep these children in school?*

Teachers, teachers' unions, and head teachers

- Low achievement levels in Nepali in certain language minorities. *How to reduce the gap between different groups in language proficiency?*
- Very low results of the main student population. *Should there be more emphasis on support with remedial teaching for those who are in danger of dropping out of Mathematics in the early grades?*
- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*
- Low number of study days. *Are bandhs useful, how to seek justice without harming the opportunities of the children to access educational provision?*
- No remarkable differences in achievement between boys and girls. *Still important to review and analyze educational materials from the gender equity viewpoint to maintain this?*
- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

Section **1**

1. National Assessment of Student Achievement in the Context of Nepal

Dr. Bhoj Raj Kafle[1]

1. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal

This first Section situates and justifies why a National Assessment of Student Achievement (NASA) 2011 is needed in the Nepalese context. It attempts to establish the context in which the current study has been initiated by exploring the situation of the various diverse groups' of Nepal in terms of achievement, geography, ethnicity, gender, language and economic status. An overview of previous studies regarding national assessment of student achievement in Nepal is also given. Then it is discussed why and how the permanent NASA unit has grown in the Ministry of Education formal stem under the Education Review Office (ERO) and elaborates on the distinguishing features of NASA 2011. Finally, the objectives of the study are given along with some overarching comments on the other Sections in the report.

1.1   The Context

In this globalized world, the importance of formal education has been increasing gradually. Schools have become a major place not only for learning the cognitive aspects of knowledge as set by the curriculum, but also as a place for learning the non-cognitive aspects of knowledge and behavior. What students learn in the school is, therefore, a matter of great importance. As we all know, this is an age of the knowledge economy. Knowledge has become the most important factor in economic development. Education is associated with productivity, the service sector, business, and industry. The function of education is to enable an individual or the society as a whole to produce knowledge, and use it for sustained economic growth so as to lead the members of its society to have better living standards. In this regard, it is very important to look at what our children are learning, how they are making progress, how our system of education is functioning and the extent to which it is

serving what we want it to serve. There is, therefore, increasing importance on information regarding the progress students make in learning achievement as students need higher levels of knowledge and skills than in the past if they are to participate meaningfully in the world of work in the future (Greaney & Kellaghan, 2008a).

Student achievement is seen as the most important indicator of national development. A country cannot develop its socioeconomic situation without enhancing student achievement (Educational and Developmental Service Centre [EDSC], 2003; EDSC, 2008). As has been argued by Greaney and Kellaghan (2008b) increased student achievement is the key to alleviating poverty and improving economic competitiveness. As a matter of fact, to get the benefit of technological advancement and have access to the development of the modern world, a country depends, by and large, on the skills and abilities of its people. Without enhancing the quality of education, a country cannot enhance the ability and skills of its people. It is therefore necessary to consistently monitor how the system of education is functioning and whether or not it is contributing to improving the learning achievement of the students so as to make them able to cope with the emerging needs of the society. For the last a couple of decades, National Assessment of Student Achievement (NASA) has emerged as a competent tool for examining whether the system of education is working towards enhancing learning achievement or causing the deterioration of the students' learning. As a matter of fact, NASA has been widely accepted as an extremely popular tool for determining the quality of education in the 1990s and 2000s (ibid) by several governments and international development agencies (e.g. The World Bank).

1.2 National Assessment of Student Achievement

Quality of education is the major concern in the field of education. The quality is itself a vague and abstract term. There is lack of congruent information to measure the educational quality. Nevertheless, all over the world, the mostly used indicator of educational quality is scores of the students' achievement tests (TIMSS & PIRLS, 2008). Most information regarding educational quality exists for those countries which have participated in the international tests such as the Program for International Student Assessment (PISA) or the Trends in International Mathematics and Science Studies (TIMSS) (TIMSS & PIRLS, 2008; OECD, 2007; 2003)**.** These tests provide a direct measure of student achievement and are typically used to indicate educational outputs and, consequently, the quality of each educational system. In the same vein, student achievement at a national level is used to measure educational quality from one country to another in many developed and developing countries (World Bank, 1996).

National assessment provides basic information for policy makers, politicians, and the broader educational community. National assessment is generally conducted

for providing information on the quality of education for policy makers, planners, managers and for other educational stakeholders (Fullbright, 2008). "It provides data for a type of national education audit carried out to inform policy makers about the key aspects of the system" (Greaney & Kellaghan, 2008b, p. 7).

Greaney & Kellaghan (2008b) further argue that the achievement of the students in a curriculum area be aggregated to provide an estimate of the achievement level in the education system as a whole at a particular age or grade level. This can then be interpreted on the basis of what skills and abilities students are expected to have achieved, i.e., the learning outcomes defined by the curriculum. The national assessment data can also be used to know how changes in enrollment and budgetary conditions, students' home environment, and socioeconomic conditions impact on the students' learning.

It is used to monitor the achievement over time, the effects of policy decisions relating to educational inputs such as budget, curricula, textbooks, teaching materials, and teacher development. So, it is useful for policy decisions especially when decisions are to be made in relation to the optimum utilisation of resources (EDSC, 2008). The national assessment results can help governments identify the strength of the association between the student achievement and various factors over which they have some control for example, the availability of textbooks, class size, and number of the years of teacher training. It is "systematic, regular measures of learning achievement in a country that is designed to assist policy making" (Lockheed *et al*. cited in EDSC, 2008, pp. 19). It is also carried out to compare factors such as geography, religion, ethnicity, gender, culture, language, disabilities and economic status. Moreover, the national assessment is widely used to compare the performance of units within the system of education such as schools, districts, regions, states, provinces or [even] nations (EDSC, 2008). It tries to gauge the difference in achievement between the rural and urban areas as well as between the public- (community-) and private (institutional) schools. Hence, it is meaningful for the National Assessment of Student Achievement (NASA) in a particular country, to address these various factors so as to determine to what extent each of them impact on student performance.

1.3   The Diversity and Issue

Nepal is a landlocked country situated in South Asia between China in the North and India towards the South, East and West. Nepal has a great variation in geographic structures, languages, ethnic groups, religions, culture and occupation. It has an enormous physical diversity with its unique geographical position and altitude variation. The elevation of the country ranges from 60 meters above sea level to the highest point on earth, Mount Everest at 8,848 meters, all within a distance of 150 kilometers resulting in climatic conditions from Sub-tropical to Arctic. The

Himalayas stretch across the northern section, eight of the ten highest peaks in the world are located and most are covered with permanent snowfall.

The area of Nepal is sparsely populated. It can be described in terms of three main geographical regions: Tarai (plains), Hills and Mountains each stretching from East to West. The country also has many ever flowing rivers. About 45 percent of the total population speaks Nepali, there are more than 123 language groups. Religiously, Nepal can be divided roughly into Hindu peoples (who live mainly in the lowlands) and Buddhists, who live in mountain villages close to Tibet. Hindus, who make up 81.3 percent of the population, dominate political and religious life (CBS, 2011).

In the same vein, wide diversity (variations) can be observed in terms of educational access and quality. This includes the magnitude of the problems of illiteracy, non-enrollment and school dropout rates, which vary by the region, gender, social group, remoteness, gender, ethnicity (in particular the Dalits) and economic status. While viewing the challenges at the macro perspective, i.e. from the national perspective, the literacy rate of the county is 65.9% (CBS, 2011), and of the total primary school age children 4.7 percent never enroll in school, and about 16 percent of the children enrolled in primary schools dropout without completing grade five. Many students begin school older than they should (at ages nine or ten); and a considerable number among them left school after completing only one year (MOE, 2012). The literacy rate of women and girls is much lower than that of men and boys and the discrepancy is worse among lower caste groups and rural people. Illiteracy limits the opportunities for millions of women and girls to learn a full range of positive life skills. There are serious gaps between the rural and the urban, ethnic groups and the Dalits.

At present, the serious diversity and variation has also been created because of urbanization, with education remaining largely urban-biased. The majority of education institutions, particularly better quality institutions, were found in urban areas. In rural areas where schools were set up, the quality of instruction was inferior, facilities were very poor, and educational materials were either difficult to find or virtually unavailable. Consequently, if rural families were serious about the education of their children, they were forced to send them to urban areas, which is a very expensive proposition that the vast majority of rural households would not be able to afford (Parajuli, 2008). The inequity in terms of affordability has resulted in a proliferation of private schools, especially in urban areas, both in terms of number of schools and students. [3]

---

[3] According to the data of the Ministry of Education (MOE 2012), there are about 15 percent private schools in Nepal. The ratio of public and private schools is expanding considerably. For example, in 1984 the ratio of private and public schools at lower secondary and secondary level was 30:1 and 5:1 respectively (Caddle, 2009). According to the latest MOE data, the ratio is in 2012 4:1; 5:1 and 4:1 respectively for primary, lower secondary and secondary schools.

1.4 Development Partners' Assistance in Education

Nepal has made significant progress on educational reforms with the proper mobilization of its internal and external resources. There have been crucial contributions from different government and other agencies from abroad for the educational development in Nepal from the very beginning. International Technical assistance was recorded while National Education Commission was formulated in 1953 for the intensive study and recommend formulating the educational policy in Nepal after the dawn of democracy. Formal and systematic assistance from the International Development Partners/DPs began in the education sector of Nepal from 1982. The pioneer initiative was the Education for Rural Development Project in Seti zone popularly known as the Seti Project. The Seti project was started with the technical and financial assistance from the United Nations Education, Science and Cultural Organization (UNESCO), United Nations Development Programme (UNDP) and United Nations Children's Fund (UNICEF). The primary aim of the project was to transform the rural communities into conscious and productive communities through education. Within the framework of this angle, the project took the initiatives for quality primary education including adult literacy programs (Santona Memorial Academy, 2011). Then after, Primary Education Project (PEP) was launched in 1983 with the assistance from the World Bank for developing school level curriculum, textbooks, teacher training and construction of school building etc. (EDSC, 2008).

In 1991 Basic and Primary Education Project (BPEP) was implemented with the support from various Development Partners (DPs) supporting different program components. For example, the World Bank supported for school construction, recurrent costs and Resource Center (RC) Construction; Danish International Development Agency (DANIDA) supported to RC development, women education and non-formal education; UNICEF supported to curriculum / textbook development and non-formal education; Japanese International Cooperation Agency (JICA) contributed on school construction materials (Skar & Cenderroth, 1997). According to Skar and Cenderroth (1997), the BPEP had a good impact on educational development in a short span of time. It provided additional classroom space for about 300 thousand students and non-formal education for more than 170 thousand adults and children. With such achievements, the government seemed enthusiastic to initiate second phase of this project. The first phase of the project was implemented in 40 districts of Nepal and the second phase covered all the 75 districts of the country.

The Basic and Primary Education Program second phase or BPEP II (1999–2004) tried to initiate the sector-wide approach to funding in the education sector. The main aim of the sector wide approach was to establish a unified financing mechanism (i.e. basket) in order to channel donor support to an agreed core investment program within the basic and primary education sub-sector ensuring better coordination and harmonization (Acharya, 2007). The major five external

supporting agencies namely, DANIDA, Norwegian Agency for Development Cooperation (NORAD), European Union (EU), Finnish International Development Agency (FINNIDA) and International Development Association (IDA) were involved in the basket system of funding. Moreover, JICA, UNICEF and the Asian Development Bank (ADB) were involved for funding to some of the specific projects. In the same way, ADB funded education projects such as, Primary Education Development Project (PEDP) and Secondary Education Development Project (SEDP) were focused in the field of primary, lower secondary and secondary education. The BPEP II was converted into the program approach and was managed under the direct administration of the regular structure of the Ministry of Education and Sports.

As a continuation of BPEP II, Education for All (EFA) 2004–2009 was implemented under the similar modalities (Sector Wide Approach (SWAp) modality in a School Sector Reform Plan/SSRP) in order to include interested DPs contributing to the program. In the beginning phase, it was initiated by the Government of Nepal and four bilateral donors, namely, the governments of United Kingdom (UK), Finland, Norway, IDA and Denmark. Later the ADB, the Government of Australia and the EU joined. The JICA and UNICEF were non-pooling donors for the EFA (Cambridge Education Ltd and METCON Consultants, 2009). Significant gains in the targeted areas like a pro-poor focus, gender mainstreaming and social inclusion, good governance and decentralization were remarkable. The government and the DPs jointly initiated a joint evaluation of the EFA Programme 2004–2009 after the termination of the programme period in 2009. The main purpose of the Evaluation was to provide information about the outcomes of Education for All 2004–2009 so that the Ministry of Education and Sports and DPs and other education stakeholders can use for policy reform and in designing the SSRP (Cambridge Education Ltd and METCON Consultants, 2009).

Apart from the above-mentioned projects and programmes, some other specific projects and programmes were launched with the technical and financial support from the DPs. Secondary Education Support Project/SESP (2003–2008) was one of them, which was launched by the government of Nepal with the financial assistance from ADB and the DANIDA. The programme aimed to expand equitable access to secondary education and improve the quality of education (MOES, 2002). The next was Teacher Education Project/TEP (2002–2007), which was implemented with the support from ADB to enhance the quality of primary education providing training for all the primary school teachers in Nepal (MOES, 2002).

Having experienced from all of these past efforts, School Sector Reform Plan SSRP, 2009–2015 has been designed and implemented as a continuation of EFA 2004–2009. The aid management under the SSRP follows the basic principles and guidelines as laid down in The Paris Declaration on Aid Effectiveness and Accra Agenda for Action (AAA) emphasizing on ownership, alignment, harmonization, managing results and mutual accountability. Based on this principle, the supports from the DPs were crucial in the areas of Government's policies,

systems, rules and regulations. In this connection, the government of Nepal has prepared the National Plan of Action for donor harmonization (MOE, 2009). The Plan for donor harmonization endorsed by the Nepal Development Forum (NDF) represents a guiding framework and reference point that provides an opportunity for the MOE to develop a detailed implementation plan for the education sector. Hence, moving towards the second phase of EFA under the SSRP framework, the SWAp and Joint Funding Agreement (JFA) process has been adopted for the effective mobilization of foreign assistance.

Similarly, the modality of incorporating Technical Assistance (TA) and Direct Funding (DF) mechanism has been taken into consideration. Moreover, the educational reforms under SSRP are being made with the continuous discussion among the ministry system and the DPs. The major donor partners under SSRP involving in the pooling funding are the World Bank, ADB, DANIDA, EU, UNICEF, DFID, the government of Finland, Australia and Norway. The DPs under non-pooling system are United Nations' World Food Program (WFP), UNESCO, UNDP, United States Agency for International Development (USAID), and other various INGOs

Frequent discussions have also been effective for prioritizing the educational activities, responding to emerging needs and carrying out innovative activities for achieving the SSRP objectives. While being more particular, such regular discussion has proven the considerable effort to add value in implementing the plan. The discussion held continuously with DPs has capacitated to the implementing agencies and authorities. It will also be contextual to mention that different DPs have different good experiences and expertise in different specific areas. National Assessment of Student Achievement (NASA) is one of the most crucial learning areas under SSRP.

1.5 The NASA and Nepalese Context

1.5.1 A historical perspective to NASA activities

The Government of Nepal has made a commitment to achieve Education for All (EFA) and Millennium Development Goals (MDGs) within the time period specified. Several interventions have been carried out to address these commitments. The EFA National Plan of Action has been developed and implemented so as to improve the access and equity, and the quality and efficiency of basic education. The Secondary Education Support Program (SESP) was implemented in 2003 to improve the learning environment, quality, management efficiency, and even access to secondary education. The Teacher Education Project (TEP) was also implemented in 2002 and all teachers were trained throughout the country. A dozen national assessment projects (BPEP, 1997; EDSC, 1997; BPEP 1998; PEDP, 1998;

EDSC, 1999; CERID, 1999; EDSC, 2001; EDSC, 2003; CERSOD, 2001; EDSC, 2008; Fullbright, 2008) have been carried out to study student achievement.

With these interventions, several gains were evident in the education sector. Increased student enrollment, improved promotion rate and learning environment, the increased percentage of trained teachers, and revised and more relevant curricula are some of the gains to mention. However, there are problems in the field of access to and quality education. About 34% of the people of the country are still illiterate. Around 4.7% children of primary level age group have not got an opportunity for schooling (studying). The groups which have not been able to access education came from hard-to-reach groups. The quality aspect of education is felt to be rather poor in community (public) schools. The community schools are heavily criticized for offering low quality education despite the fact that these schools have trained and qualified teachers, a reasonable level of infrastructure, and a guarantee of the salary, etc. the low quality is generally attributed to over the politization and mismanagement of schools.

Keeping all these educational scenarios in perspective as well as considering the changed political situation, international situation, the responses of stakeholders and expectations of Nepalese people and utilizing the experience gained on EFA, SESP, TEP and other efforts and initiatives, the Ministry of Education (MOE) developed the School Sector Reform Plan (SSRP) in 2009 and launched it in the fiscal year 2009/10.

The SSRP is perceived as a major intervention in school education which will integrate different levels of educational structure at the school level like primary, secondary and higher secondary levels into a single structure "school sector education". In terms of functionality, what it would mean is that unlike the past, the government, the Ministry of Education would see this sector as a whole rather than seeing it as different pieces, parts and layers of education. The government machinery would own the sector as a whole and provide assistance to be utilized for the whole. Historically, it has been quite some time since Nepal put much emphasis on primary education alone and allocated almost 55% of its budget to it alone.

In this sense, the SSRP can be best understood as the structural and functional integration of the education system. As it appears, the SSRP intends to improve the internal efficiency of the education system and to harmonize policy functions in the sectors with a focus on ensuring quality and excellence in education. It must be noted that the main focus of the education reform is concentrated on the improvement in the student learning. It is widely accepted by the persons involved in the reform process that unless the students who receive the reformed education demonstrate better performance in terms of learning outcomes, the reform, no matter how much is poured into it and who are involved, it would make not much sense. Since the primary purpose of education is concerned with the student learning, the most important measure of an education system is the enhanced

level of student achievement. In this regard, the approved SSRP document has also made NASA an integral part of it which mentions that the assessments of student achievement will be conducted periodically for students enrolled in grades 3, 5 and 8. It is therefore argued that the proposed national assessment studies will help to set norms and standards of the quality of achievement of our students on the one hand and the system of education as a whole on the other. Taking all these concerns into account, the Nepal Government has made a meaningful decision of establishing the NASA Unit within the regular structure of the Ministry of Education.

1.5.2 Establishment of NASA unit under ERO

The School Sector Reform Plan (SSRP) has envisioned the establishment of the Education Review Office (ERO) as an independent institution in order to ensure the accountability of the institutions and quality of education. In accordance with the spirit of the SSRP, an Education Review Office has been established within the regular structure of the Ministry of Education. It is understood that, for the time being, it is under the Ministry of Education, but soon will be given an independent status so that it can work autonomously without any pressure from the Ministry of Education. The office is responsible for conducting the external audits of agencies at different levels based on set norms and standards by using appropriate measures. To provide additional management support to the ERO, the Ministry of Education has issued a directive (Directives for the Management of Education Review Office 2010). According to the directive, a NASA unit has been established under the ERO. The main purpose of establishing the NASA unit is to support the government to accomplish its educational objectives by getting feedback from the studies on the student learning achievement.

As has been envisaged by the SSRP, the NASA programmes are supposed to be regular activities of the government. Within this framework, the present study, the first one after the establishment of the NASA Unit would determine the current level of student achievement. This study would be followed by another study of its type. By this statement, it could mean that the forthcoming study would be carried out to look into the extent to which the system of education is functioning well so that the learning achievement levels of the students are improved. Or, it could also show that the learning achievement level is not improving or deteriorating or remaining static. What is important to note here is that the NASA exercise is carried out to examine the overall health of the system of education, and whether or not it is functioning towards accomplishing what we want it to accomplish. In this context, the newly established NASA unit is intensively involved in carrying out assessment activities. The NASA unit has prepared a roadmap for further works as shown in Table 1.

Table 1.1 Long-term plan (roadmap) of NASA for Nepal

Grades	2008	2009	2010	2011	2012	2013	2014	2015	2016
Grade 3 & 5	√[1]				√		√		√
Grade 8	√[1]			√		√		√	

1) administered before SSRP
   Source: NASA brochure, the Nepal Government, Ministry of Education, ERO

1.5.3 Characteristics of NASA 2011

Based on the roadmap, in the first cycle of the programme, the NASA study administered questionnaires and achievement tests involving three subjects (Nepali, Mathematics and Social Study) to grade 8 students. As has been discussed earlier, the present study is not the first national assessment study carried out in Nepal under the auspices of the Ministry of Education. Several studies were carried out in the past. However, there are some special features of the NASA 2011 study which make it appear different to those studies carried out in the past.

Large scale and wider coverage: In terms of the size of the sample, NASA 2011 is very large in comparison with the samples used before. According to Greaney & Kellaghan, (2008b), it is always wise to follow the strong theoretical background while selecting the sample of a study which is carried out for important decisions. Since the present study was methodologically highly complex and requiring intensive collaboration among many stakeholders, it was decided that a sample as large as possible should be involved. National assessment is a systematic study that follows standard procedures. It is a regular study in a sense that it is not a one shot activity, rather an activity to be carried out at regular intervals so that the impact of reform initiatives on the performance of the system of education can be ascertained (SEDP, 2008). National assessment studies normally involve the administration of achievement tests either with a sample or a population of the students, usually focusing on a particular sector in the system (such as the fifth grade or 13-year-old students). In addition, teachers and others (for example, parents, principals, and students) may be asked to provide background information, usually in the form of a questionnaire to provide insights on how student achievement is related to factors such as household characteristics, levels of teacher training, teachers' attitudes toward the curriculum areas, teacher knowledge, and availability of the teaching and learning materials (ibid).

It seems difficult to gauge overall levels of student achievement, to assess the relative performance of particular subgroups, and to monitor changes in performance over time (Greaney & Kellaghan, 2008b). It is difficult to determine

what is adequate or sufficient as a representative sample of population. The available evidence indicates that few developing countries have systematically addressed such issues. In our case in the past, the national assessment studies involved about 2% of the total population. It is difficult to say whether it is sufficient or not to accept as a representative sample. There are also cases which have used even a lower sample size. Keeping this in mind, in this study, the ERO has decided to involve 5 % of the total schools as a sample accepting the total number of schools as the population and a school as a basic unit. In comparison to the studies carried out so far, this size is quite large. From our point of view, this sample size (1201 schools, and Nepali language (16350 students, 418 schools, 415 Head Teachers (HT) and 415 teachers), Mathematics (16033 students, 421 schools, 407 HTs and 407 teachers) and Social Studies (16299 students, 414 schools, 402 HTs and 402 teachers) would be enough to make the findings of the study generalizable.

As the NASA is a continuous process, it was planned that within three rounds of the students' assessment all the 75 districts would be covered. In the first cycle, 25 districts were selected by following random sampling procedures. The other districts were expected to be covered by the studies to be carried out next year and the year to follow. In this sense, the present study is supposed to be a part of the broad design of the national assessment study.

 Use of item response theory (IRT): While viewing the comparison perspective of assessment results, one of the modern ways of conducting the item analysis and test construction is using Item Response Theory (IRT) modeling. This is the statistical way of equating (comparison) several datasets. IRT is essential to validly monitor performance over time and to describe growth between grades and between groups of interest; this modeling also allows the description of growth in terms of skills and understandings, and is therefore a powerful diagnostic support for policy makers and curriculum developers (Metsämuuronen, 2002).

IRT modeling is used two ways. First, the item difficulties are calibrated to the same scale by using linking items and a certain mathematical model. Second, the test scores of the different version or different years can be equated, that is, the scores can be adjusted comparable. Hence, IRT modeling can be used to compare two consecutive years' data by using the help of calibration of item difficulties. It can be expected that by using IRT modeling, valid and reliable information is available for monitoring changes in levels of student achievement (trends) over time. In this study IRT has used to equate and compare the study data of 2008 and 2011, and the 2011 result with some international studies.

Comparison with 2008 study: As mentioned above, NASA is not a new initiative in the context of Nepal. About a dozen other studies have been carried out. Nevertheless, the study data from these has not been permanently established as

an institutional memory. No linkage between one assessment test and the other has been established. Also the processes of these studies differ. Hence, no basis was found to compare achievements among one study with others. The only meaningful comparison able to be made was between the 2008 and 2011 studies; this comparison revealed a number of important findings.

International flavor: This study is different from other studies in a sense that the present study bears some international experiences. Most national assessment studies carried out in the past involved only national consultants. In the present study, an international consultant provided by the Finland Government for Nepal who was associated with the Finish National Board of Education was involved as an expert. In a way, the involvement of an expert, who has a rich experience in such studies at national and international levels and has strong professional strength, has made this study more professional. His involvement has given this study an international flavor. It must be appreciated that his involvement offered a new experience to those who are involved in this study from the ERO.

A number of countries in the world have been participating in international tests such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Studies (TIMSS)**.** These tests provide a direct measure for student performance and are typically used to indicate educational outputs and, consequently, quality and accountability of each educational system. Such initiatives can directly link with the national studies which ultimately would help to know the national standard to the international standard and to take necessary improvement. This study has no direct bearing in the international studies. However, it gives some international flavor in a sense that some test items were developed gaining insights on those studies and throughout the study these tests were kept in perspective. In a way, this is an indication that Nepal has the willingness to work in close communication with international trends and align itself to the examples of best practice.

Item bank: One of the distinguishing features of the 2011 study is that a special effort has been made to prepare the item banks. A large number of items was prepared and pretested. After the item analysis, all the items were banked for later use.

Item-wise analysis: Analysis of each and every item and getting output is yet another unique and compressive work in this study. Around forty-nine thousand student booklets and individual questions have been analyzed separately to identify the content wise achieve situation. This effort has helped to generalize and recommend specific areas to be improved on in the curricular process.

Shared approach: Another thing which makes this study different from previous studies is the way it was carried out. In the past, all the studies were carried out by external consultancy agencies recruited, on a competitive basis, by the

Department of Education, and Ministry of Education. There was no direct involvement of the government agency or individuals in carrying out any activity apart from getting involved in the steering committee and monitoring the quality of the work. But unlike this, after the establishment of ERO/NASA, it carried out by itself several activities of the present study such as training on general test item analysis, item writing, item selection, item decision, developing questionnaires and ability test items, pretest, pretest analysis, final selection of items, typing and preparing of the camera ready copy (CRC) of the test booklets. After completing all these activities, it hired a national consultancy firm to carry out the remaining activities such as the printing of the booklets, the delivery of the booklets to schools, test administration, the collection of the booklets, marking/scoring the items (booklets) and data entry. After that NASA again took responsibility for the data analysis and production of the report. In this way, the present study is a result of a joint venture of the internal exercise of the regular structure of the Ministry of Education, Education Review Office and a national consultancy firm as an external agency.

1.6   Objectives of the Study

The main objective of this study was to present the national scenario of grade 8 student achievement by providing evidence on the student learning in the education system. More specific objectives were as follows:

- To determine the current national level of achievement of eighth-grade students in Mathematics, Nepali, and Social Studies;
- To determine variations in student achievement between different ecological zones, Developmental regions, districts, school location (rural/urban), ethnicity, gender, language groups, socioeconomic conditions as well as school type (community/institutional schools);
- To examine extent to which the school, teacher, home background, and other pupil factors influence learning learning achievement;
- To compare student learning achievement in the current study with that of the 2008 study;
- To compare the student learning achievement in Nepal with that of international studies: PISA (Reading) and TIMSS (Mathematics and Geography).
- To create reliable baseline data for the future study in 2013;
- To generate recommendations for policy making to improve educational quality as well as for other stakeholders to improve their respective working areas including schools.

1.7 Overview of the Report

This report consists of five main Sections. In this introductory Section, an overview has been made regarding the context of national student assessment in general and how the context appeared in Nepali specifically. While discussing the Nepali context, a brief discussion has been made on the issue of its national characteristics, i.e. plurality/diversity (geographic, religious, ethnic, gender, language, etc.) assuming that such discussion would help to pave some way out for identifying the areas of results in the Section for the basic results (Section 3). Moreover, discussion has been made about the need for establishing a National Assessment Unit under MOE system, including the special characteristics of the NASA 2011 for grade 8.

Section 2 addesses the methodological aspects of the study and presents in detail the procedure used to carry out the national assessment. This Section will contribute to gaining basic skills in key technical aspects. This Section has sections on 1) sampling, item writing, 2) background questionnaires, 3) IRT modeling and equation of the test scores, 4) IRT modeling, 5) criterion-based assessment in Nepali language and CEFR and 6) used statistical methods. Section 3 presents the subject-wise results for Mathematics, Nepali, and Social Study. For each subject, Section includes 1) the main findings (results), 2) diversity-wise deepened results, and 3) the selection of other related factors, such as socioeconomic status, explaining the results. Section 4 offers a set of deepening analysis in various aspects on the basis of the whole data of around 50,000 students and 1200 schools: Diversity effect, Teacher effect, School effect and Individual effect. These Sections try to explain why the poor schools are poor and the good schools are good. Finally, Section 5 reflects in the results and offers recommendations for the future.

References for Section 1

Acharya, S. (2007). *Social inclusion: Gender and equity in education SWAPs in South Asia, Nepal case study.* Kathmandu: UNICEF, Regional Office for South Asia.

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

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

Caddle, M. (2009). Private schools as battlefields: Contested visions of learning and livelihood in Nepal. In **B**. **Promod**, *Education in Nepal: Problems, reforms and social change*. Kathmandu: Martin Chautari.

Cambridge Education Ltd and METCON Consultants (2009). *Joint Evaluation of Nepal's Education for All 2004-2009 Sector Program. Oslo:* Norad (Norwegian Agency for Development Cooperation).

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

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

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

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

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

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

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

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

Fullbright (2008). *A final report on national assessment of grade-V students*, Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education. Fullbright Consultancy & Cooperative Hands in restoration (CHIRAG).

Greaney, V. & Kellaghan, T. (2008a). *Implementing a national assessments of educational achievement*, Vol 3, the World Bank.

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

**Lockheed, Marlaine E., Verspoor, Adriaan M., and associates (1996*).* Improving Primary Education in Developing Countries*. New York. Oxford University Press.

Metsämuuronen J (2002). *Small book of test theory and test construction*. Helsinki: International Methelp Oy

MOE (2009). The school sector reform plan, 2009-2015. Kathmandu: The government of Nepal, Ministry of Education MOE.

MOE (2012). *Flash I report, 2012*. Kathmandu: Ministry of Education.

MOES (2002*). Secondary education support programme 2003*. Kathmandu: Ministry of Education and Sports, DANDDA/DOE

MOES (2003). *Education for all 2004-2009*: Core document. Kathmandu: Ministry of Education and Sports/DOE

OECD (2003). Learning for tomorrow's world: First results from PISA. OECD, Paris.

OECD (2007). The Programme for International Student Assessment (PISA) 2006: Science Competencies for Tomorrow's World Executive Summary. OECD, Paris.

Parajuli, K. (2008). The Valley's Relentless Growth. *Himal South Asian*, 21(10/11), 34–41.

PEDP (1998). Primary Education Development Project: An assessment. 1998. Kathmandu.

Santona Memorial Academy (2011). Role of resource center for improving quality education in schools. Kathmandu.

Skar, H. O. & Cederroth, S. (1997). *Development aid to Nepal: issue and option in energy, health, education, democracy and human right*. Copenhagen: Nordic Institute of Asian Studies (NIAS).

TIMMS & PIRLS (2008). Appendix C: The test-curriculum matching analysis. Trends in Mathematics and Science Study (TIMMS) & Progress in International Reading Literacy Study (PIRLS) International Study Center. Retrieved from: http://timss.bc.edu/timss2007/PDF/T07_M_IR_AppendixC.pdf

World Bank (1996). *National assessments: Testing the system*. Washington: The International Bank for Reconstruction and Development / The World Bank.

Section 2

Methodological Solutions in NASA 2011

Dr. Jari Metsämuuronen[1] & Mr. Shyam Prasad Acharya[2]

1. Helsinki University, Faculty of Behavioral Sciences, Finland
2. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal

The National Assessment of Student Achievement (NASA) for grade 8 in the year 2011 (in what follows, NASA 2011) is the largest national assessment administered during the history of several assessment projects in Nepal. The core functions (sampling, developing the tools, and data analysis) was made by the Educational Review Office (ERO)/NASA unit with the help of an international expert. This section describes the methodological solutions of NASA 2011. These methodological solutions – sampling, item writing and test construction, background questionnaires, equating of the test scores over three versions used in the final testing, as well as the statistical methods used in getting the result or findings (Sections 3.1. to 3.3) are common to all three subjects (Mathematics, Nepali language, and Social Studies) and hence they are handled jointly in this separate section. Section 2.1 concentrates on sampling; Section 2.2 on the item writing and test construction; Section 2.3 on the background questionnaires; Section 2.4 on the variables used in the analysis; Section 2.5 on the equating of the test scores over three versions used in the final testing; Section 2.6 on the principles of criterion based assessment used in the assessment of Nepali proficiency; and Section 2.7 on the statistical methods used in getting the result or findings in Sections 3.1. – 4.6.

2.1 Sampling

2.1.1 Strata

Because the main interest in the National Assessment is not the individual students but the system itself, the basic unit for the sampling is the school. Thus the schools should represent the country as widely as possible and the pupils are selected to

represent the schools as well as possible. The most usable sampling strategy is therefore the Proportional stratified sample with random selection. The strata are:

1) Ecological zones[1] (Mountain, Hill, and Tarai),
2) Developmental regions (Eastern, Central, Western, Mid-Western, and Far-Western),
3) Districts (75 altogether),
4) School type (Community and Institutional),
5) School location (Rural and Urban).

Kathmandu Valley can be taken as a single 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, also the mixed ethnicities, weather conditions, economic activities, aggressive development, as well as the dense human capacity make the Valley a unique fourth geographical area in the analysis. Hence there were 16 basic areas in the sampling ($3 \times 5 + 1$). Such type of information as School Size, Ethnic group and School Language were also used in selection, too, however, not as strata. Selecting all 75 districts with a representative number of schools from each would have automatically lead to a good coverage of both the Developmental regions and the Ecological zones, but to do so would have entailed selecting far too many schools to be cost-efficient from the assessment point of view. Instead it was decided that for the first phase of the sampling, 25 Districts were to be randomly selected to represent each of the 16 geographical strata (see Figure 2.1). The selected districts, numbers of schools and numbers of the students are condensed in Table 2.1
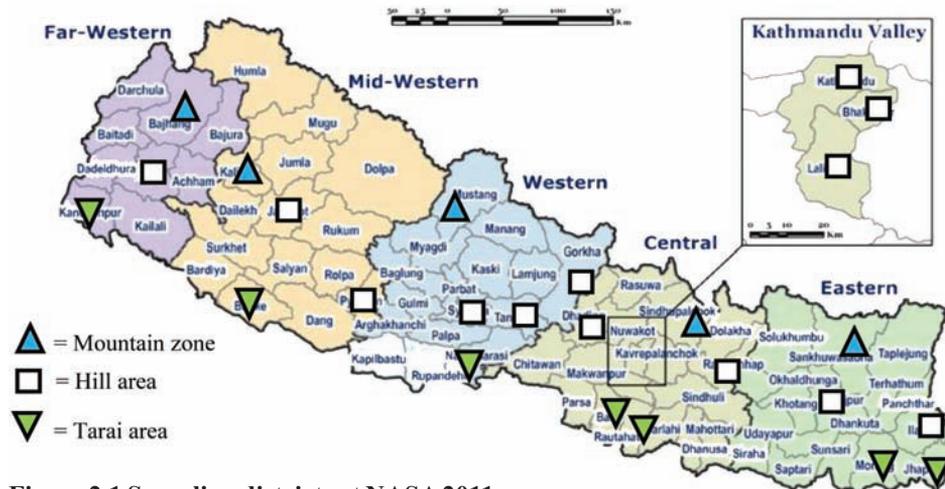
Figure 2.1 Sampling districts at NASA 2011

---

[1] Within the text the Ecological regions or -belts are called Ecological zones to make a distinction to Developmental belts which are called Developmental regions.

Table 2.1 Basic information on the Sample of NASA 2011

District	N. of Schools			N. of the students		
	Mathematics	Nepali Language	Social Studies	Mathematics	Nepali Language	Social Studies
1. Sankhuwasab	8	8	8	241	279	242
2. Ilam	20	20	20	896	927	1055
3. Bhojpur	13	13	13	484	532	471
4. Jhapa	20	20	20	1100	1169	1230
5. Morang	19	18	18	1041	975	921
6. Sindhupalch	11	11	11	370	384	352
7. Ramechhap	15	15	15	581	687	597
8. Dhading	22	23	22	782	886	872
9. Rautahat	13	13	13	775	873	880
10.Bara	15	15	15	842	732	726
11.Mustang	7	7	7	35	36	34
12.Gorkha	18	18	18	853	775	713
13.Tanahu	25	27	25	936	947	838
14.Syangja	28	27	27	977	941	871
15.Nawalparas	24	23	23	986	1009	1117
16.Kalikot	15	15	15	184	194	192
17.Pyuthan	11	11	11	498	580	527
18.Jajarkot	13	13	13	496	437	423
19.Banke	16	16	16	810	720	637
20.Bajhang	10	10	9	271	296	259
21.Doti	17	17	17	359	365	454
22.Kanchanpur	15	15	15	698	584	622
23.Lalitpur	12	12	12	272	336	400
24.Bhaktapur	11	9	9	171	245	335
25.Kathmandu	43	43	41	1375	1441	1531
Altogether[1]	421	419	413	16,033	16,350	16,299

1) Note that in several schools, two or even three subjects were tested. Finally, the number of different schools was 1201.

2.1.2 Sample size

At the second phase of the sampling, the number of the schools in each of 25 districts was fixed proportionally on the basis of the number of schools in each stratum. In the sample-based national level student assessments, the conventional maximum sample size is less than 5% of the population (see Cochran 1977; Bartlett, Kortlik & Higgins, 2001).[5] On the basis of the latest official list of schools, the number of schools giving teaching to the 8th graders exceeded 7000. Thus the number of schools per subject was fixed to be 400, that is, coverage of 1200 schools in the final testing. Note that the schools were not selected at this phase but only the *number* of schools in each stratum was fixed.

For the convenience of the selected schools, all the students in the selected schools were originally included to take part in the final test, so that the schools did not have to arrange an alternative programme for those students excluded in the sample. The number of the students was therefore much larger than needed but it was easier to organize the test in this way. Additionally, because the number of the students was not cut to a fixed number (as is a convention in the international comparisons), there is no need for using complex sampling procedures. Initially 400 schools per subject were chosen on the basis of a rough estimation of the number of the students in the schools, as discussed above. After fixing the number of schools, it was estimated that 48,000 students would attend and on average 40 students per class would come.

2.1.3 Selection of the schools

At the third phase of the sampling, the heads of the District Educational Offices (DEO) of the selected 25 districts provided the NASA unit with a list of the functioning schools in their districts. This list included information such as the number of the students, school type (community - institutional), and school location (rural - urban) which were then used as the basis for the random selection of the schools in the district.

The sample schools for two districts in the Himalaya area were selected in advance in October. In those two districts, Mustang and Kalikot, the schools closed three months before those in the other areas because of the difficult and dangerous

---

[5]   Actually, there is no rule for 5% of population. However, the classical formula of Cochran (1977) for estimating the sample size reduces the sample size when the suggestion at the first round exceeds 5% of the population. Conventionally, oversampling is suggested when it is expected to see loss in response rates (see Salkind, 1997, 107; Fink, 1995, 36). In the national level testing, this is not expected and hence oversampling was not planned.

climate conditions – in particular due to the incidences of snowfall which often made it impossible to reach the schools after mid-November. The rest of the schools were selected in December 2011.

2.2 Item writing, Pretesting, Final Tests and Their Reliability and Validity

2.2.1 Item writing

During 13.4.–25.4.2011 and 26.5.–10.6.2011, three groups (representing school teachers teaching in grade 8, experts and university teachers) of item writers formed three "factories" to accomplish the task of creating a sufficient amount of items for four pretest versions of the pretest – and ultimately for three versions of the final test. The Mathematics item writing group was led by Prof. Hari Upadhyaya, the Nepali language group by Dr. Madhav Paudel, and the Social Studies group by Prof. Sri Ram Upadhyaya. All the item writers had a meeting at the end of each day with the statisticians; the day's work was inventoried and the progress was checked.

The item writers were very effective: within two weeks, altogether 600 (Mathematics) + 550 (Nepali) + 450 (Social Studies) = 1600 items were created. This amount of items was far too many for effective testing and thus there was a need for the selection of the best ones. During May and the beginning of June, all the items were translated into English (also Nepali language items) and they were screened, edited and rewritten when needed. Pre-test papers were printed in the same secured manner as the examination papers in the governmental printing house (Janak Shikksya Samagree Kendra) in Sanothimi.

2.2.2 Pre-testing

During June 2011, items for three subjects with 12 versions of test papers (3 subjects x 4 versions) were pre-tested in 13 district and 90 schools by the ERO personnel. 250 to 270 items for each subject were tested for final use. Administration of the pre-tests was done by the personnel in the District Educational Offices (DEO); DEOs organized an orientation session, monitored the processes in the schools, collected the papers and sent them to the main office in Kathmandu for marking and data entry. The District officers themselves were oriented to the process in a two-day seminar where an elementary understanding about objective testing was shared; they were also motivated by a visit from the Minister of Education himself as well as by the Secretary and other high-level officers from MoE.

ERO personnel monitored the pretest process on site. 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 DEO offices.

2.2.2.1 PISA and TIMSS pretests and notions from the field

On the top of the main pretest in June 2011, four other smaller scale pretests were administered during August and September to acquire the item information about the released items from the Trends in International Mathematics and Science Study (TIMSS) in Mathematics, TIMSS items on Geography (part of the Social Studies in Nepal), and of the Programme for International Student Achievement (PISA) reading items. The item parameters based on the Item Response theory (IRT, see Section 2.4.1) of the TIMSS and PISA items were released but it was important to collect other relevant information, such as classical item parameters and the time on task of the items for the final item selection.

During the pretest for the PISA reading items five alarming observations were made. First, there was a very low efficiency in the technique in the reading test. The evidence collected during the testing (time spent in reading each page) showed that many pupils spent only one to two minutes to read two pages of informative text related to polices' work and the latest advances in it. Hence, it is evident that the pupils did not even read the texts but started to guess the answers straight away. As a consequence, the reliability of the test stayed quite low – even the good readers seemed to guess the answers. Because the PISA reading items are constructed cleverly, one cannot guess the correct answers easily. Second, students used very open "social work" in the testing: pupils – especially somewhat poorer ones – were very willing to help each other. This became easy because the students usually sat very close to each other. Third, the general atmosphere was somewhat "do it as fast as possible" as it would be a heroic work to finalize the test as soon as possible. Fourth, the pupils (9th graders) seem to panic in the reading test situation. Especially when there were two longer texts (two pages) to read, it seemed to present a kind of task that the students were unfamiliar with. Finally, there seemed to be some item types which were not familiar to the pupils. Especially difficult were such items where the students were asked to justify why they held a particular view after reading two texts which disagreed with each other in a particular area. None of the students were able to produce their own opinion without copying from the text itself. Because of the observations, another testing session was organized with a different administration.

22

2.2.2.2 Experimental pretest with reading items

To find an optimal environment for the final reading tests, another testing setting was created. In the other pre-test settings, the students were selected purposefully in the test so that the 15 highest performing students of the 9th grade and the 15 lowest performing students of 8th grade (note one year difference and the extremes) completed the same reading test as in the first pretest setting. The students were put into a classroom so that there was enough space between the students – maximum two in a row – and the students were encouraged to read the texts first (this was done also in the first settings) because there is no hurry in the testing, and to start to work with the items after reading the texts. The time spent on the test was the same as in the first settings with the students in a "panic". The result was that the average performance of the poorest students of the grade 8 in the latter setting was better than the average students' in the first settings. The best of the latter setting performed much better than the best in the first setting. In both settings, the schools were mediocre type of community schools. The only difference between the settings was the administration of the tests.

On the basis of these experiences – and because of the willingness to see the best performance of the students in the final test – two suggestions were given to the final testing. First, the reading test should be administered so that at first the students are given only the reading material – not the test papers – to make sure that they spend some time just on reading the texts. After 30–45 minutes, they will be given the final test papers with reading, writing, grammar and vocabulary items. Second, it is vital that elementary level students would have their own space in the testing to avoid the "social work" or cheating, during the test. In the best option, the final test setting would follow the practices of examinations. The idea was that the NASA tests would be administered one week before the final examination as a "preliminary" test for the examination – the students are at their highest level of knowledge because of the repeated studying of the study material for the examination.

2.2.3 Principles of the item selection for the final test

Six basic principles were used when selecting items for the final tests: (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 the results with 2008 results and with the international results (TIMSS and PISA). These matters are discussed here from the theoretical point of view and in the next section from the practical viewpoint; in Section 2.2.4 it is shown, what kind of characteristics the final tests had.

2.2.3.1 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. Those documents were used as the main criteria for the validity of the tests. Three specification grids (or Table of Specifications) – one for each subject – were prepared on the basis of the curricula. Actually, the work already completed by the Examination board in this regard was utilized. In their grids, the time spent on the tasks in the curricula was operationalized as percentages for each topic and subtopic. This information was used as a basis in item writing and item selection: the number of marks on the tests is proportional to those percentages in the grid. The independent element in the national student assessment comes from the fact that sometimes the assessors need to think outside of the learning outcomes rather than stay strictly in what is written in the curriculum. For example, in the national curriculum of the Nepali language the 8th grade includes only 9 % on reading and 65% on writing tasks. It would be, however, a poor reading test when out of 100 points only 9 would come from the reading. There was thus a need to alter the structure of the final test from this respect.

Another issue is that some of the subtopics or even main topics can be omitted from the final testing when it was deemed necessary. In the Mathematics and Social Studies there was not this kind of need. However, in the Nepali Language, the listening and speaking comprehension were omitted because of the practical requirements of setting up comparable testing situations in the field.

2.2.3.2 Content coverage to be as wide as possible

From the content validity viewpoint it is wise to select the items to cover as broad a range of topics as possible. All the main topics are covered on the specification grid. However, the tests are not that long to make it possible to cover all the subtopics as well. A sub-test length of 3−4 items may be taken as a technical minimum length to discriminate the test takers from each other sufficiently. However, in the Nepali test for instance, out of all items there are only eight vocabulary items and nine grammar items. In this case, the selection of the subtopics is done on the basis of wide coverage rather than condensing the areas to allow better sub-tests. Thus an attempt was made to include as many subtopics as possible in the tests. The selection was, however, proportional: when there were subtopics of wider coverage in the curriculum than the others, more items were selected from those areas in the test constructions.

The content coverage was widened by using three different versions in the final testing in each subject. One version was used in Himalaya four months before

the final testing in the other areas. Because there has to be linking items between the different versions and because there was a possible threat of leaking the items, the Himalaya test was made shorter than the other versions

2.2.3.3 Proper structure of cognitive levels

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

It was noticed that in the international tests the classifications of Comprehension and Application seemed to have been difficult to differentiate from each other and thus they seem to be connected as one group in the TIMSS and PISA tests. The same logic was used also when combining the NASA versions. Approximately 65% of the items fell into category 'Comprehension and Application', 10 to 'Knowledge' and 25 to 'Higher skills'.

2.2.3.4 High test discrimination (Reliability)

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

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

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

It is noteworthy that there are only two sources of information needed for estimating the reliability of the test: the item discrimination (r_{it}) and item variance (s_i^2). It is also noteworthy that the Alpha reliability is maximized when the sum of the elements r_{it} and s is the highest. 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 1) as high item discrimination as possible and 2) mediocre difficulty level in the test.

Generally, the values for reliability lower than $\alpha = 0.70$ 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 closer Nunnally, 1978; DeVellis, 1991; Hair *et al.*, 1998). The boundaries are not strict as, for example, Knapp & Brown (1995) have discussed.

2.2.3.5 Proper difficulty level

In the pretest phase it was noticed that many items were too difficult to measure the average achievement level in students' population. Generally, the pass rate for these items was less than 50%. Especially in the Mathematics test, quite a number of the students did not even start to do the open-ended questions which lowered the proportion of correct answers. This means that the pretest 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 MCQ-type items with five options,
- 74 % in the MCQ with four options,
- 77 % in the MCQ with three options,
- 85 % in the Wright/Wrong and True/False type items.

Other boundaries have also been introduced (see Mehrens & Lehman, 1991). The most balanced test for the national student assessment is achieved when the items are selected from the whole range of ability. Thus, there should be easy items, mediocre items and demanding items on the test. These kinds of items can discriminate the best students and the poorest students as well as the mediocre students. One possible solution is to select the items so that

- 10 % of the items should be very easy,

- 20 % of the items should be quite easy,

- 40 % of the items should be of medium difficulty,

- 20 % of the items should be quite demanding,

- 10 % of the items should be very demanding.

Combining the principles, by selecting the easiest items from the pretests the aim was to raise the average difficulty level in the tests near $\overline{p} = 0.60$ (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 on the basis of the classical item difficulty parameter p though the IRT parameter B would have been more accurate.

2.2.3.6 Comparability with the 2008 results and with the international results

The fifth principle in item selection was that the results should be comparable with 2008 results and with international test results. The 2008 tests were very short and they were not planned for use as a basis for IRT modeling. Thus the test markers were also allowed to give half points. In order to transform the data to make it suitable for IRT equation, no half points were allowed and they were systematically transformed to the lower whole number. Thus the score 0.5 was transformed to 0, and 1.5 to 1. This logic is the same as in the IRT modeling: if there is not enough ability to give the correct answer, it is more probable that they will fail the item. The obvious coding mistakes in the 2008 datasets were also corrected. The 2008 items were reanalyzed with the transformed and checked values. Such items were selected as linking items where the marking would be as unambiguous as possible in comparison to the 2008 marking. When possible, the linking items were the multiple choice type of questions – or otherwise the short answer type.

One interesting possibility is to compare the Nepalese results with the international test results. TIMSS- and PISA communities have released some items and their parameters and it makes sense to use those international items as the baseline in the comparison (TIMSS 2007; 2009a; 2009b; Adams & Wu, 2000; PISA, 2006c; 2009). The idea is that when one knows the difficulty parameter of the international items, those values can be fixed in the datasets in Nepal and thus the local items are calibrated onto the same scale as the international items are. Pretested TIMSS Mathematics and Science items came from the year 2007 and PISA reading comprehension texts and items from the years 2000 and 2006. All the item parameters were not published in PISA reports. However, psychometrician Greg Macaskill at PISA research office ACER (Australia) kindly estimated the missing parameters for the NASA unit.

2.2.4 Final tests

2.2.4.1 Mathematics test

On the basis of the pre-test, the average difficulty levels of the Mathematics tests were expected to be 55−56% of maximum score in versions 1 and 2 and 60% in the Himalaya version. Between the versions 1 and 2 there were 11 linking items and the Himalaya version was a compilation of versions 1 and 2. Version 1 was used for linking the results to the 2008 dataset with 9 items and version 2 was used for linking the results to the TIMSS dataset with 13 items. The number of items is sufficient to equate the 2011 results with the other datasets because even one item would have been technically sufficient.

Versions 1 and 2 were parallel when it comes to their item-wise characteristics (Tables 2.2 and 2.3). The construct validity is quite high from the curriculum viewpoint: the subtest-wise number of items match almost perfectly the proportions in the curriculum. Different versions include a wider variety of subtopics under each topic.

Table 2.2 Topic-wise Characteristics of Mathematics Tests

Topics	VER1	VER2	VER3	VER1%	VER2%	VER3%	Curriculum%	Reliability[2]
Total	74[1]	76[1]	63[1]	100	100	100		0.92
Algebra	22	22	17	29.7	28.9	27.0	30	0.78
Arithmetics	16	15	15	21.6	19.7	23.8	20	0.70
Geometrics	21	23	18	28.4	30.3	28.6	30	0.80
Sets	7	8	6	9.5	10.5	11.1	10	0.59
Statistics & Data	8	8	7	10.8	10.5	11.1	10	0.42

1) marks
2) weighted mean over the different versions

Reliabilities of the total score of the different versions were over $\alpha = 0.90$, that is, the test scores can 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 reliabilities of the scores of Statistics & Data and Sets remain low.

Most of the items and marks come from the comprehension and application type of items (Table 2.3). Higher skills are measured mainly by the open-ended,

productive type of questions. In Mathematics – contrary to the Social Studies and Nepali language – the students were not able to even start the most difficult open-ended questions in the pretest. Thus, in the Himalaya version the number of these items was kept lower than in the other two versions. Hence, also the proportion of marks from the items measuring the higher skills is lower in the Himalaya version than in the other versions.

Table 2.3 Hierarchical-wise Characteristics of Mathematics Tests

Hierarchical level	VER1	VER2	VER3	VER1%	VER2%	VER3%
Knowledge	11[1]	11[1]	12[1]	14.7	14.5	19.0
Comprehension & Application	42	42	36	56.0	55.3	57.1
Higher skills	21	22	15	28.0	28.9	23.8

1) marks

2.2.4.2 Nepali language tests

The Nepali language tests took one hour more to complete than the other tests, that is, almost 2.5 hours. The reason for this is, as stated above, it was decided to include 30–45 minutes just in the reading the texts in the reading comprehension test, that is, prior to when the question papers were distributed. Average difficulty levels were 55% of the maximum score and in the Himalaya version 57% but the final values were somewhat lower. Between versions 1 and 2 there are 12 linking items and the Himalaya version was a compilation of versions 1 and 2.

Contrary to the tests of Mathematics and Social Studies, the linking items for the year 2008 and PISA were used as linking items practically in all the test versions. The reason is that there were really few usable reading items available in the 2008 test and all the PISA items are reading items. Though the number of items was smaller than in the Mathematics and Social Studies tests, it is still sufficient to equate the 2011 results with the other datasets. Another difference is that six of the items were planned to be used two times in the process. Namely, there were not enough short writing items on the pretest versions. Therefore six reading items with short-answer type of questions were selected to be scored twice, first for reading (the content of the writing) and second, for writing (the grammar of the writing).

One additional challenge came from the fact that in the curriculum of Nepali Language 65% of time is allocated for writing and only 10% for reading with the remaining 25% for Grammar and Vocabulary. Because of the international practice

of assigning reading to more than 10% of the contents, it was proposed that some of the content area of writing should be changed into reading to obtain enough information on the reading comprehension in Nepal. It was decided that 50% of the marks would come from writing and the reading would gain more space in the tests (Table 2.4). There were not enough varying texts in the pretests and thus some international texts (released PISA texts) and related items were selected to cover different text types: short opinion type of texts, short letters, medium length informative texts, numerical tables and graphs, long informative texts, and long narrative texts. These text types were pretested with a small scale pretest after the actual pretest (see Section 2.2.2). However, the opinion types of texts were omitted from the final tests because, without few exceptions, the pupils were not able to answer the questions which required the pupils' own opinions of the texts. It is worth noting that the number of marks in the vocabulary score is higher than they should have been because one set of the linking items from the 2008 test with five items was originally scored as an entity and hence it was not possible to take only one or two items out of the score. Hence, it increased the number of marks and biased the percentage of the content coverage somewhat.

Table 2.4 Topic-wise Characteristics of Nepali Language Tests

Topic	VER1	VER2	VER3	VER1%	VER2%	VER3%	Curriculum%	Reliability[2]
Total	66[1]	65[1]	59[1]	100	100	100		0.90
Writing	37	37	34	51.4	51.4	51.5	65	0.70
Reading	16	15	14	22.2	20.8	21.2	10	0.81
Grammar	11	12	10	15.3	16.7	15.2	15	0.63
Vocabulary	8	8	7	11.1	11.1	10.6	10	0.54

1)   marks
2)   weighted mean

Reliabilities of the versions were round $\alpha = 0.90$, that is, the test scores can 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.

As in Mathematics and Social Studies, in the Nepali Language test most of the items and marks come from the comprehension and application type of items (Table 2.5). Contrary to Mathematics and Social Studies there were very few items reflecting the knowledge level of achievement. Instead, there are more items measuring the higher skills because most marks in writing tasks came from the open-ended, productive type of questions.

Table 2.5 Hierarchical-wise Characteristics of Nepali Language Tests

Hierarchical level	VER1	VER2	VER3	VER1%	VER2%	VER3%
Knowledge	3[1]	3[1]	3[1]	4.2	4.2	4.5
Comprehension & Application	39	34	31	54.2	47.2	47.0
Higher skills	25	28	24	34.7	38.9	36.4

1) marks

2.2.4.3 Social studies test

The average difficulty level was 49% of maximum values and 45% in the Himalaya version. Between versions 1 and 2 there were 13 linking items and the Himalaya version was a compilation of versions 1 and 2. Version 1 was used for linking the results to the 2008 dataset with eight items and version 2 was used for linking the results to TIMSS Geography dataset with eight items. The number of items is sufficient to equate the 2011 results with the other datasets. However, it is worth noticing that the TIMSS linking items concern only a part of the Social Studies test, namely Geography. Thus, the number of items on the sub-test of Geography is higher than it should be from the curriculum view point. All the selected TIMSS linking items came from the general "our Earth" sub-topic and thus the Nepali geography and international geography items were collected mainly into version 1.

Versions 1 and 2 are quite parallel when it comes to their item-wise characteristics (Table 2.6). The number of Geography items is higher in version 2 because of the international linking. The construct validity is still very high from the curriculum viewpoint: the subtest-wise number of items match almost perfectly the proportions in the curriculum. Different versions include a wider variety of subtopics under each topic. Reliabilities of the subtopics in the Social Studies tests range $\alpha = 0.86$–0.88, that is, the total test score can 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. It is, however, worth noting that the reliabilities of the sub-scores of History, Politics and Geography are very low and cannot be entirely relied upon to give accurate information; the obvious reason for this being the small number of items on the test.

Table 2.6 Topic-wise Characteristics of Social Studies Tests

Topic	VER1	VER2	VER3	VER1%	VER2%	VER3%	Curriculum%	Reliability[2]
Total	86[1]	86[1]	82[1]	100	100	100		0.87
Civics	39	38	36	45.3	44.2	44.4	45	0.73
Economics	18	17	17	20.9	19.8	21.0	20	0.55
Geography	11	13	11	12.8	15.1	13.6	13	0.49
History	9	9	10	10.5	10.5	12.3	12	0.34
Politics	9	9	8	10.5	10.5	9.9	10	0.46

1) marks   2) weighted mean

As in the Mathematics and Nepali Language tests, most of the marks in the Social Studies test come from the comprehension and application type of items (Table 2.7). Higher skills are measured mainly by the open-ended, productive type of questions. In Social Studies – contrary to the Mathematics – the students were willing and able to start even the most difficult open-ended questions. Thus, their proportion is higher than in Mathematics. The somewhat higher proportion of knowledge type of questions in the version 2 is caused by the TIMSS linking items. The same linking items were taken into the Himalaya version.

Table 2.7 Hierarchical-wise Characteristics of Social Studies Tests

Hierarchical level	VER1	VER2	VER3	VER1%	VER2%	VER3%
Knowledge	8[1]	15[1]	12[1]	9.3	17.4	14.6
Comprehension & Application	41	41	43	47.7	47.7	52.4
Higher skills	37	30	27	43.0	34.9	32.9

1) marks

2.2.5 Marking schemes

The data collection, marking, and data entry were outsourced to a private company. Because the company has not been involved strictly with the item writing, pretesting, nor marking of the pretest papers[6], it was important to prepare a marking scheme for each subjective type of item[7] so that the marking would be as reliable as possible. A marking scheme was prepared by adding stricter examples of what kind of varieties of correct answers were allowed for the marks and which kinds of answers were not credited by marks.

Later, in the analysis phase, it was noted that in the Social Studies bank some items needed to be discarded as linking items because it was obvious that the marking schemes of 2008 and 2011 did not correspond. The 2008 marking scheme has obviously been much lighter than 2011. One indication of this is a linking item where the students were asked to draw a volcano with relevant characteristics. In 2008, more than 80% of the students gave a response with the maximum point (one mark) but in 2011 only 7% of the student got the one mark; declining of the achievement level cannot explain the decrease in the percentage of the correct answer. This led to an unfortunate fact that the test scores of 2008 and 2011 were not possible to equate in a credible way.

---

[6] Some of the workers in the company were involved the processes but most were not.
[7] Most of the objective type of items – True/False-, Multiple choice-, and Matching type of items – are not marked.

2.2.6 Final test administration, marking, and data entry

2.2.6.1 Test administration

The final test was administered in two phases: at the end of December in Mustang and Kalikot (Himalayan region) and at the end of February in all the other 23 districts. The administration was outsourced to a private company, and the officers from the ERO monitored the process. Actual information collection was made by the District officers.

The second part of the testing was administered at the end of February. Unfortunately, the teachers were rioting against government policy at the time of the final test and they refused to do anything which involved governmental offices and their duties. Hence in many districts, the teachers were not participating in the briefing meetings or testing. However, within two to three weeks all the districts administered the tests. It was later noticed that, even though it was not intended, in ten schools two subjects were tested at the same school. Hence the number of schools in Table 2.1 exceeded 1200.

2.2.6.2 Marking and data entry

The marking of the papers as well as the data entry was outsourced. Altogether 48.682 students' papers, around 1,200 teachers' questionnaires and 1,200 head teachers' questionnaires were collected into an "examination center" in Kathmandu. Within three months of April, May, and June 2012, the outsourced company marked and entered the material.

One nationally remarkable phenomenon in marking may be worth noticing – this had an effect on the subjective items in the Nepali Language and Social Studies tests. The teachers and hence also the markers seem to be extremely tight when marking open-ended writings such as essay type of items. It is more or less a tradition that the highest marks are not given but to a few, ultimately the best, students even though there would be objective criteria for gaining the highest marks. This phenomenon was seen, for example, in an essay type of item with a maximum eight points in the Nepali Language test: out of 8023 students none was given a maximum score of 8 and only six students (0.07%) were given seven points. It is, however, difficult to believe that out of 8000 students there would not be more of those who would have earned 7 and 8 points rather than only six.

The student datasets were screened and, except for some minor mistakes in language groupings, no mistakes were found. 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. The teachers' and head teachers' datasets were entered twice because of numerous errors in the data.

2.3   Background Questionnaires

In student assessment, as important as *describing* the achievement results is to try to find factors which *explain* the differences between the students and schools. Hence, one needs background questionnaires including the relevant contextual information on the students, students' family as well as the school and teaching. The real fact is that there are hundreds or maybe thousands of important factors explaining the results, some of which may be important in Finland, for example, and some specifically in Nepal. Some of the indicators may be interesting to know though one cannot *do* much of the matter. This kind of variable is, for example, the socioeconomic status (SES) of the pupil's family. It is known that this characteristic of the pupil seems to explain the achievement level quite well (see Section 2.4.3) but it cannot be changed easily; 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 increase the adult literacy rate in the country.

There are many indicators which may be valuable to measure during the student assessment process as noted above. Hence, how to make the selection?

2.3.1 Conceptual model for the background questionnaires

A sketchy modeling of this complex phenomenon of learning is in use in the Finnish National Board of Education (FNBE, Metsämuuronen, 2009). The same model, with contextual modification, was used as the basis for compiling the background questionnaires (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 near the pupils are the **peer group factors**: social support to studies, bullying, atmosphere in the classroom, and so on. All these matters were asked straight from the pupils. **Teacher factors** evidently play some role in student achievement, however, usually much less than are expected. Such factors as classroom activities, teaching skills, use of teaching materials, for example, may affect the student achievement. These are asked from 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, safeness, absence, and so on, are asked from the head teacher or principal. **Economic factors** and **demographic factors** can be found in national statistics – in the NASA activities, the demographic factors are part of the sampling scheme.
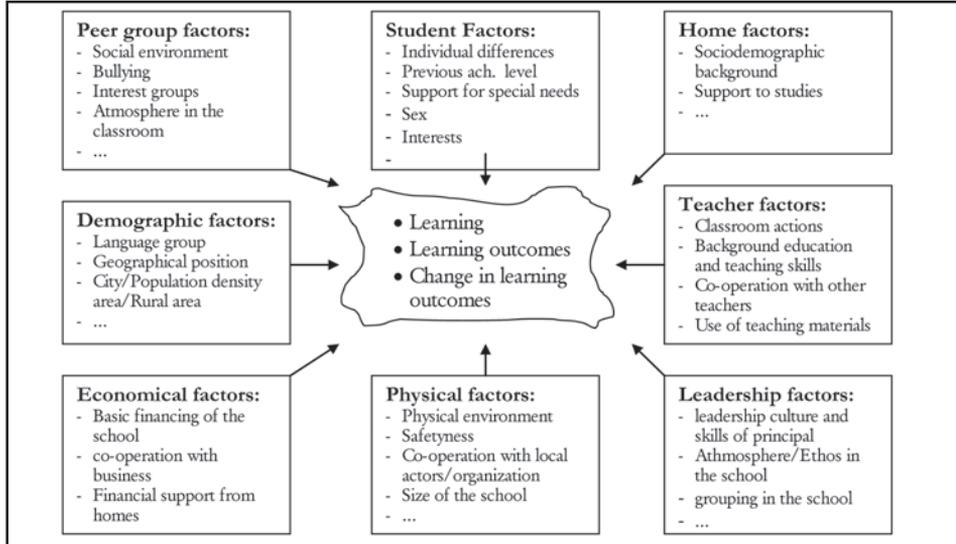
Figure 2.2 **Conceptual framework for the background information (adapted and contextualized from Metsämuuronen, 2009)**

2.3.2 Selection of the questions

The variables for the background questionnaire were selected in several phases. At the first phase, the statisticians familiarize themselves with the 2008 background questionnaire and both TIMSS- and PISA questionnaires (PISA 2003a; 2003b; 2006a; 2006b; TIMSS, 2003; 2006). In the second phase, the regression models prepared in 2008 were studied. Two things were decided: 1) the variables which showed statistically significant relevance with the student achievement in 2008 should, at least, be included in the questionnaire and 2) there will be three questionnaires: questionnaires for students, teachers and head-teachers. The family questionnaire was omitted and, instead, the questions related with the family were asked from 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, more relevant variables instead. At the final phase, the subject committees took the final stand for the questionnaires. Altogether three background questionnaires were prepared: 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 resources related questions.

2.4 Specific Variables Used in Analysis

Within the analysis, three sets of variables may be worth handling more carefully in order to fully understand the results: the concept of equated scores, Fennema-Sherman attitude scale, and the indicator for socioeconomic status (SES). Because three different versions were used in student achievement, the scores are not automatically comparable. Hence, the scores have to be equated before the analysis; the logic and procedure of these equated scores are handled in Section 2.4.1. The Fennema-Sherman test is a widely used test (e.g. in PISA- and TIMSS questionnaires) to measure attitudes toward school subjects; it is handled in Section 2.4.2. The socioeconomic background variables are used to explain the achievement results – they seem, in many cases, to explain the differences between the students quite well. The motivation and construction of the SES variables are handled in Section 2.4.3.

2.4.1 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 (Rasch, 1960; Lord & Novick, 1968; Lord, 1980; Hambleton, 1982; 1993; 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 modeling has in practice superseded 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 2011 is typically this kind of complex endeavor which benefits from IRT modeling. Practically speaking, the IRT modelling is the only credible way to assess how the achievement level is changed from 2008 to 2011 and to compare the results in Nepal with the international standards (such as PISA 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 subtopics – in the testing process, it was natural to use several versions in the testing. In NASA 2011 testing, three versions were used. As a comparison, in PISA and TIMSS testing procedures more than 15 versions are administered in each school. Second, because two Himalayan districts (Mustang and Kalikot) were tested two months earlier than the other areas, it was important to produce a separate version in those schools to avoid the possible leaking of the test items. Third, in order to compare the results of 2008 to 2011, IRT modelling is, by far, the most accurate method in order to perform the comparison when equal (or parallel) tests are not used. Fourth, 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. Additionally, using IRT modelling made it possible to free one of the equal lengths of the test papers; the Himalayan version was planned to be shorter than the other two versions to avoid the leaking of the items. When the tests were of unequal length and deliberately of somewhat different difficulty level, IRT modeling 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 equation process is as follows (see more exhaustively in Béguin, 2000, 17–36):

1. Define the structure of the test so that the linking items are connecting the tests into each other. Because the values of difficulty β 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.

2. Use *Conditional Maximum Likelihood* (CML) procedure to estimate the difficulty level (β parameter) for each item.

3. Use *Marginal Maximum Likelihood* (MML) procedure to estimate the distribution of each student's latent ability (θ parameter) in each version.

4. Estimate the θ parameter of the scores of each version using means and deviations of distributions of β and θ. This results in a unique latent value, however measured in a common scale, for each observed value of the scores in all versions.

The success of the equating depends on three things. First, the linking items should represent a sufficient range of ability level; too easy and too difficult items should, however, be avoided. Second, the linking items should represent a short test inside the test; the items should cover the different content areas as widely as possible. Third, the stable parameters in the 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 in 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 is the higher

is his/her θ above the zero line and parallel, the weaker the student the lower is θ below the zero line. 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 will be calibrated into that international scale and hence the mean of the "average Nepalese student" does not get the value of zero but either above or below. This makes it possible to assess what is the achievement level of Nepalese students compared with the international standard.

At the final phase, the Mathematics scores were calibrated into the TIMSS Mathematics scale, Social Studies scores were calibrated into the PISA Geography scale, and Nepali scores were calibrated into the PISA reading scale. In Section 3, the results produced by the IRT modeling are hidden; 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 all the tasks (of the total score or sub-scores) by the maximum marks.

In deeper analysis in Section 4, the original Theta values are used in analysis. However, another type of equating was used to calibrate the different Theta values from three subject-wise datasets. Namely, after merging all three student datasets (Mathematics, Nepali, and Social Studies) the average score in the Mathematics test was lower than in the other tests.[8] Hence, without any further transformations, the lowest students would be automatically those with low achievement in the Mathematics test. This is why the equated scores were further transformed by using so called 'mean equating' which means that the means of the different subject populations were shifted so that the mean of all the distributions were equal.[9] Technically speaking, the mean of the means (46.91) was taken as the baseline for the transformation. All the student scores were shifted depending on the subject tested in the school; students in the Mathematics test were shifted up and students in the Nepali Language- and Social Studies test were shifted down. After the shift, a very small number of the students appear to have lower value than 0 or higher than 100. These are out-of-range values but they were not altered.

---

[8] The mean in Mathematics was 42.8, in Nepali 48.6 and in Social Studies 49.3.

[9] In the 'mean equating' one assumes that the distributions are equal but their locations are different. Now it is known that the distribution of Mathematics differed from the other subjects (see Section 3.1.3). It was taken as a minor thing compared with the radically lower mean in the test. Other option would have been so called 'linear equating' which is used when there is a difference between the means as well as between the variance.

2.4.2 Fennema-sherman attitude scale

A shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) are used in several international comparisons, like in Trends in International Mathematics and Science Study 2007 (TIMSS, Mullis, Martin, & Foy, 2008) and its predecessors 1995, 1999, and 2003 as well as in Programme for International Student Assessment (PISA). Original scales include nine dimensions but in these international comparisons only three dimensions with four items on each (see Table 2.8) and two negative items on 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). Metsämuuronen (2012a; 2012b) noted that this kind of "Expected factor structure" can be found in all Western countries including European countries (except in Bulgaria and Romania), Australia, Canada, Israel, the United States and Russia. However, after performing exploratory factor analysis (EFA) with Principal Axis Factoring, 3 factors, and Promax rotation with Kaiser Normalization separately in all countries, it is notable that in several countries, this structure cannot be found. Instead, an unexpected factor structure ("Totally unstructured", see Table 2) characterized by one factor of pure negative items can be found in almost all countries in the Middle East and several countries in East Asia – and in Nepal.

Table 2.8 Expected Factor Structure of FSAS in North America in TIMSS 2007

Pattern Matrix[a]

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

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

Table 2.9 Totally Unstructured Factor structure in Syria in TIMSS 2007

Pattern Matrix[a]

	Factor[b]		
	1	2	3
I enjoy learning Mathematics	.846b		
I hate Math	-.783		
I would like to take more Math in school	.588		
I learn things quickly in Math	.442		
Usually I do well in Math			
I need Mathematics to get into the <university> of my choice		.659	
I need to do well in Math to get the job I need		.609	
I need Math to learn other school subjects		.494	
I think learning Nepali language will help me in my daily life		.396	
*Math is not one of my strengths			.696
*Math is more difficult for me than many of my classmates			.662
*Math is boring	-.312		.364

a) Extraction Method: Principal Axis Factoring.
   Rotation Method: Promax with Kaiser Normalization.
   Rotation converged in 5 iterations.

b) Loadings over .30 are seen

Metsämuuronen (2012a) showed that some of the items in FSAS are too difficult for the international population. Additionally, he shows (Metsämuuronen, 2012b) that the test carries culturally sensitive elements. Especially challenging are the long negative items where the lowest level students most probably did not understand the questions and hence their responses are illogical. In several Asian countries, the negative items were not answered in an appropriate way and hence it was assumed – though not analyzed – that there has to be some cultural elements which lower the reliability of the test.

Table 2.10 Unstructured Factor Structure in Nepal in NASA 2011 (n = 48,682)

Pattern Matrix[a]

	Factor		
	1	2	3
[Subject] is not one of my strengths[c]	.777[b]		
Sometimes, when I do not initially understand a new topic in [Subject], I know that I will never really understand it[c]	.674		
I learn things quickly in [Subject]	.412		
[Subject] is more difficult for me than many of my classmates[c]	.390		
I need to do well in [Subject] to get into the <university> of my choice		.625	
I need to do well in [Subject] to get the job I need		.620	
I need [Subject] to learn other school subjects		.406	
I would like a job that involved using [Subject]		.316	
I think learning [Subject] will help me in my daily life			
I enjoy learning [Subject]			.631
I would like to take more [Subject] in school			.410
I usually do well in [Subject]			.359

a) Extraction Method: Principal Axis Factoring.
   Rotation Method: Promax with Kaiser Normalization.
   Rotation converged in 4 iterations.
b) Loadings over .30 are seen
c) Negative items are reversed

In NASA 2011, a somewhat modified set of items of FSAS, borrowed from PISA 2009, was in use. Two original items, *I hate* [Subject] and [Subject] *is boring* were replaced by a long negative item *Sometimes, when I do not initially understand a new topic in* [Subject], *I know that I will never really understand it.* Another item in PISA 2009 was added to the dimension of "Experiencing utility in Subject", namely *I would like a job that involved using* [Subject]. Table 2.10 shows the factor structure in the whole dataset including the students from Math-, Nepali Language- and Social Studies tests. Note that the last item on the table, *I usually do well in* [Subject] should load with the other variables measuring the "Self-Efficacy in Math" and not with the variables of "Liking Math". Note also that the loading of *I think learning* [Subject] *will help me in my daily life* is very low – it is not seen on the table because the values less than 0.30 are hidden.

Because the structure of FSAS in Nepal is somewhat deviating from the intentioned structure, within the texts, mainly, only the total score is reported. The reliabilities of the scores are condensed in Table 2.11. The reliabilities are low – as in several Asian countries in TIMSS 2007 dataset.

Table 2.11 Reliabilities of the FSAS Scales in Nepal in NASA 2011 (n = 48,682)

Scale	number of items	reliability
Total score	12	0.57
Liking Subject	2	0.43
Self-Efficacy in Subject	5	0.57
Experiencing utility in Subject	5	0.61

2.4.3 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 a belief that high SES families provide for their children an array of services, goods, parental actions, and social connections that potentially rebound to the benefit of children and a concern that many low SES children lack access to those 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, 2011).

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, 5). The challenge in measuring SES is that there has not been a complete consensus on precisely what it represents (Liberatos, Link & Kelsey, 1988; McLoyd 1997) – 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).

Hence, because there is no consensus of the measures for SES it freed the NASA team to create – on the basis of the variables measured in NASA 2011 – a SES indicator which fits the Nepalese reality. Seven indicators of SES were selected into the final SES indicator:

- Father's education
- Mother's education
- Father's occupation
- Mother's occupation
- Home possessions (e.g. "*do you have* a dictionary in your home")
- Home accessories (e.g. "*how many* mobiles do you have in your home")
- Attending to Private school (yes – no)

Because the variables were of different scales (from nominal to ordinal scales) and because of incomparable scores (from 0–1 to 0–12), all the variables were re-scaled first to fit each other. At the first phase, the variables were analyzed with the respect of educational outcomes. Decision Tree Analysis (DTA, see Section 2.6), the data mining tool in SPSS software, and ANOVA, the basic tool for analyzing the differences between the group means, were used to find the best classification of each variable in regard to the statistical differences in learning outcomes. At the second phase, five variables comprising the home accessories were dichotomized of the basis of DTA and ANOVA and summed up. At the third phase, all seven variables for SES were dichotomized on the basis of DTA and ANOVA. Hence, all the variables – regardless of their original scale – were scaled as 0 or 1 where 1 indicates the higher SES (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 indicator. The indicators and their cut-offs are condensed in Table 2.12.

It is worth noting that the final SES indicator (1) is strictly geared toward educational outcomes (and not health, for example), (2) is balanced with education and occupation though is somewhat over-representing the economic dimension (3 indicators), (3) has 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 2011.

Table 2.12 Indicators of SES in NASA 2012

Variable	cut-off[1]	effect on total score[2]
Father's education	less than SLC-passed = 0, other = 1	+6% points, $\eta^2 = 0.025$
Mother's education	less than SLC-passed = 0, other = 1	+7% points, $\eta^2 = 0.028$
Father's occupation	Agriculture = 0, other =1	+6% points, $\eta^2 = 0.023$
Mother's occupation	Agriculture = 0, other =1	+6% points, $\eta^2 = 0.017$
Home possessions	5 or less out of 12 possessions = 0, 6 or more = 1	+5% points, $\eta^2 = 0.016$
Home accessories		
Mobile phone	2, 3 = 1, other = 0	+6% points, $\eta^2 = 0.025$
Television	1–3 = 1, other = 0	+7% points, $\eta^2 = 0.032$
Computer	1–3 = 1, other = 0	+8% points, $\eta^2 = 0.028$
Car	1–3 = 1, other = 0	+3% points, $\eta^2 = 0.003$
Bathroom	2, 3 = 1, other = 0	+3% points, $\eta^2 = 0.006$
all together	0 out of 5 = 0, other = 1	+6% points, $\eta^2 = 0.019$
Attending to Private school	no = 0, yes = 1	+19% points, $\eta^2 = 0.142$
Total SES		+25% points, $\eta^2 = 0.106$

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

2.5 Criterion-based Assessment in Nepali Language and CEFR levels

An additional possibility in the Nepali Language test was to use the criterion-based assessment. In the testing of Mathematics and Social Studies 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 and Social Studies produces a *norm* with which the different groups (such as Ecological zones, Developmental regions, or sexes) 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 and Social Studies, the only external norm able to be used is the international norm coming from the large international population (TIMSS and PISA datasets).

Contrary to the situation in Mathematics and Social Studies, in Nepali Language it is possible to apply criterion-based assessment, because in the language

testing settings there are several common frameworks for setting the standards. One of those, the Common European Framework of Reference for Languages (CEFR)[10] 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. CEFR levels and the national results are handled in detail in Section 4.6.

2.6 Used Statistical Methods

2.6.1 Analytical tools used in the statistical analysis

Basic statistical methods are used throughout the report. 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. These methods are described in all standard text books of statistical description and –inference (e.g. Metsämuuronen, 2013). Analyses are done in SPSS environment. The Analysis of Variance (ANOVA) and – Covariance (ANCOVA) are used in the General Linear Modeling (GLM) way when several means are compared. All the p-values are corrected by using Multilevel modeling (Goldstein, 1986; or Hierarchical Linear modeling, Bryk & Raudenbush, 1987) by using SPSS Linear Mixed models module. In some cases (for example, when explaining the schools' average student achievement level with SES variables), traditional Linear regression analysis and Logistic regression analysis are also used

Somewhat more exotic methods – 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 – it is very effective when it comes to finding statistically the best groupings of the independent variables. The selections related to DTA are explained with the results in Section 4. In any case DTA produces a chart such as the one below:

---

[10] It could be called also as the Common *International* Framework for Reference of Language (CIFR) because it is in use in several other countries – not only in Europe.
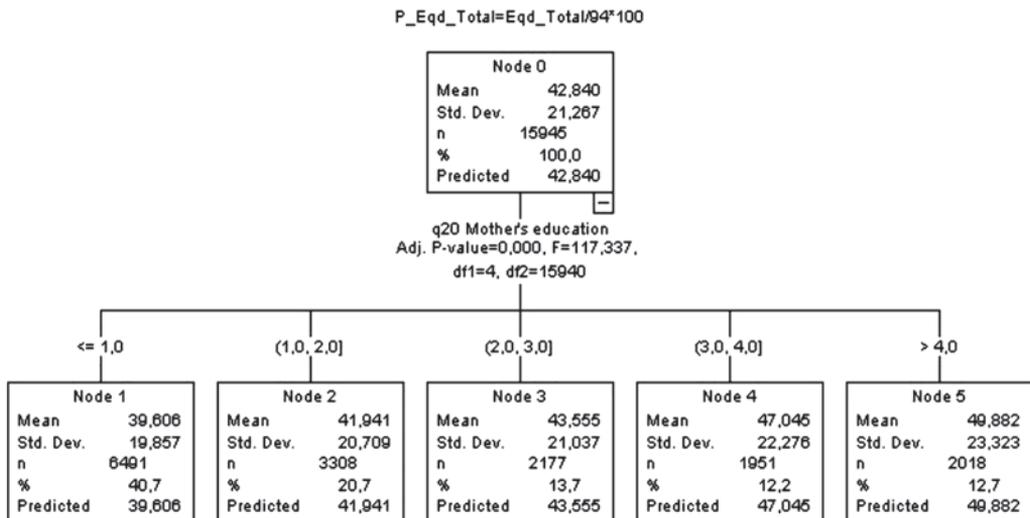
P_Eqd_Total=Eqd_Total/94*100

Node 0	
Mean	42,840
Std. Dev.	21,267
n	15945
%	100,0
Predicted	42,840

q20 Mother's education
Adj. P-value=0,000, F=117,337,
df1=4, df2=15940

<= 1,0	(1,0, 2,0]	(2,0, 3,0]	(3,0, 4,0]	> 4,0

Node 1		Node 2		Node 3		Node 4		Node 5	
Mean	39,606	Mean	41,941	Mean	43,555	Mean	47,045	Mean	49,882
Std. Dev.	19,857	Std. Dev.	20,709	Std. Dev.	21,037	Std. Dev.	22,276	Std. Dev.	23,323
n	6491	n	3308	n	2177	n	1951	n	2018
%	40,7	%	20,7	%	13,7	%	12,2	%	12,7
Predicted	39,606	Predicted	41,941	Predicted	43,555	Predicted	47,045	Predicted	49,882

Figure 2.3 An example of DTA modeling

The chart shows that the original "mother node", node 0 where the student average achievement is 42.8% of the maximum score, can be divided into five categories which all differ from each other in a statistically significantly manner. In node 1 (here "<=1", that is, the lowest class of mother's education, that is, "illiterate") the student achievement level is 39.6 and in the highest group (node 5, here ">4", that is, "SLC passed mothers") the average student achievement level is 49.9. In the boxes there are also information on Standard Deviation (Std. Dev.), sample size (n), the percentage of cases that are in this node (%), and what would be the predicted mean in the node (Predicted) which appear to be the same as the mean. 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 4 (which is the number of groups minus 1) and 15940 (which is N minus the number of groups).

2.6.2 Some statistical concepts used in the text

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

When the sample size is huge – as the sample of 16,000 students is – the p-value very easily gives a signal that the difference between the groups is real in population. P-value does not, however, tell whether the difference is small or big. For this purpose there is another statistic, Effect Size (ES). Effect size indicates how far the lowest and highest groups are from each other. The especially used indicators of ES are Cohen's d used for two means and Cohen's f for several means (Cohen, 1988). Cohen has given boundaries for small, medium and large 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 high effect sizes are collected in Table 3.1.3.

Table 2.13 Rough boundaries of Effect Sizes

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

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

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

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

Hence, if $\eta^2 = 0.30$, then the effect size $f = \sqrt{(0.3/0.7)} = \sqrt{0.43} = 0.65$ showing high effect size (see Cohen, 1988, 284).

2.7 Conclusions

As a conclusion of the sampling, sample sizes, item writing and the final tests, questionnaires, procedures used, as well as used methods, it is fair to say that NASA 2011 data meets the requirements of international standards for the serious national assessment of student achievement. It is, though, hubris to say, that anything in the processes and tests could not be better. In any case, the reader can be assured that the results in Sections 3 and 4 are based on valid and reliable original observations.

References for Section 2

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

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

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

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

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

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.

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.

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.

FNBE (2004). *National Core Curriculum for Basic Education 2004.* The Finnish National Board of Education. Helsinki: Yliopistopaino. Retrieved from http://www.oph.fi/english/sources_of_information/core_curricula_and_qualification_requirements/basic_education.

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,* (5th Edition). Upper Saddle River, NJ: Prentice Hall.

Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on Mathematics attitude. *The Teaching of Mathematics*, **9**(2), 41–51. Retrieved from http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf.

Kadijevich, Dj. (2008). TIMSS 2003: Relating Dimensions of Mathematics Attitude to Mathematics Achievement. Retrieved from http://www.ipisr.org.rs/Dokumenta/zbornici%20PDF/Zbornik%202008%2040-2/Zbornik-08-2-006.pdf .

Kaftandjieva, F. (2009). Standard Setting. Reference Supplement, Section B. In **S. Takala** (Ed), *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual.* Language Policy Division, Strasbourg.

Knapp, T. R. & Brown, J. K. (1995). Ten Measurement Commandments That Often Should Be Broken. *Research in Nursing & Health,* **18**, 465–469.

Krieger, N., Williams, D. R & Moss, H. W. (1997). Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu. Rev. Public Health,* **18**, 341–78.

Liberatos, P., Link, B. G., Kelsey, J. I. (1988). The measurement of social class in epidemiology. *Epidemiol. Rev.,* **10**, 87–121.

Lord, F.M. (1952). The relationship of the reliability of multiple-choice test to the distribution of item difficulties. *Psychometrika,* **18**, 181–194.

McLoyd, V. C. (1997). The impact of poverty and low socioeconomic status on the socioemotional functioning of African-American children and adolescents: mediating effects. In R. D. Taylor & M. Wang (Eds), *Social and Emotional Adjustment and Family Relations in Ethnic Minority Families*. Mahwah, NJ: Erlbaum. pp. 7–34.

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology.* 4th edition. Harcourt Brace College Publishers, Fort Worth.

Metfessel, N., Michael, W. B., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the writing of behavioral Objectives. *Psychology in the Schools*, **6**, 227–231.

Metsämuuronen, J. 2009. *Methods assisting the assessment*. Oppimistulosten arviointi 1/2009. The Finnish National Board of Education. Helsinki: Yliopistopaino. [In Finnish].

Metsämuuronen, J (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) December, 1–17.

Metsämuuronen, J (2012b). Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies,* 4(3) September.

Metsämuuronen, J. (2013). *Handbook of Research Methods in Human Sciences.* International Methelp Oy. Available as E-book.

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*. Mahwah, NJ: Lawrence Erlbaum. 249–281.

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.) McGraw-Hill, New York.

Oaks M. (2011). Measuring Socioeconomic status. In e-Source. Behavioral & Social Scoence Research. Retrieved from http://www.esourceresearch.org/Portals/0/Uploads/ Documents/Public/Oakes_FullChapter.pdf.

PISA (2003a). PISA Student Questionnaire. Retrieved from http://pisa2003.acer.edu.au/downloads.php.

PISA (2003b). PISA School Questionnaire. Retrieved from http://pisa2003.acer.edu.au/downloads.php.

PISA (2006a). PISA Student Questionnaire. Retrieved from http://pisa2006.acer.edu.au/downloads.php.

PISA (2006b). PISA School Questionnaire. Retrieved from http://pisa2006.acer.edu.au/downloads.php.

PISA (2006c). PISA Released Items – Reading. OECD

PISA (2009). Take the Test. Sample Questions from OECD's PISA Assessments. OECD.

Salkind, N. J. (1997). *Exploring research* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Takala, S. (Ed.) (2009). *Relating Language Examinations to the CommonEuropean Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual.* Language Policy Division, Strasbourg.

TIMSS (2003). TIMSS Contextual Background Questionnaires. Retrieved from http://timss.bc.edu/timss2003i/context.html.

TIMSS (2006). Questionnaires. Retrieved from http://nces.ed.gov/timss/questionnaire.asp.

TIMSS (2007). *TIMSS 2003 Science Items. Released Set, Eight grade.* IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2009a). *TIMSS 2007 User Guide for the international Database. Released Items, Mathematics – Eight grade.* IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2009b). *TIMSS 2007 User Guide for the international Database. Released Items, Science – Eight grade.* IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

van der Schoot, F. (2009). Cito variation of the Bookmark Method. Reference Supplement, Section I. In **S. Takala** (ed.), *Relating Language Examinations to the CommonEuropean Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual.* Language Policy Division, Strasbourg.

Verhelst ND, Glas CAW, Verstralen HHFM (1995). *One-Parameter Logistic Model OPLM.* Cito, Arnhem.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychol. Bull.,* **91**, 461–81.

Wright, B. D. (1968). Sample-free test calibration and person measurement. *Proceedings of the 1967 Invitational Conference on testing Problems.* Educational Testing Service, Princeton, NJ.

Basic Results of Student Achievement in Mathematics, Nepali, and Social Studies

3.1 Mathematics Achievement in NASA 2011

Mr. Shyam Prasad Acharya[1],
Dr. Jari Metsämuuronen[2] &
Prof. Dr. Siddhi Koirala[3]

1. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal
2. Helsinki University, Faculty of Behavioral Sciences, Finland
3. Tribhuwan University, Kathmandu, Nepal

Abstract

Achievement in Mathematics of the grade 8 students was assessed by utilizing student achievement tests and the related background questionnaires of 16,033 students from 421 randomly selected schools from 25 randomly selected districts. The schools represented all Ecological zones and Developmental regions, rural- and urban areas as well as community- and institutional schools.

The achievement in Mathematics is not normally distributed; there seems to be three populations: low-performing, medium-performing and high-performing students. The learning outcomes are weaker in Geometry and Sets than in other content areas. The students' ability to solve complex problems in Mathematics is low. Students are good in very fundamental operations, the basic manipulation of data and numbers, and calculations with few steps. They are much weaker in reasoning, problem solving, plotting, proving the theory or formula, and constructing the shape and figures. In all content areas, the results are slightly lower in the year 2011 than 2008. The average achievement of Mathematics seems to be slightly better in Nepal than the international average (+0.5).

There are some positive signs of educational equity. Though the boys are slightly out-performing girls in all content areas of Mathematics the differences are not remarkable. Also, though the results are somewhat better in cities than in the rural area, the difference is not remarkable. However, there are some signals which are alarming: The results in Bhojpur (18), Pyuthan (29), Illam (30), Sankhuwasabha (32), Kanchanpur (32), Morang (34), and Jajarkot (35) are very low in comparison with the districts in the Valley (on average, 55). Of the Developmental region, the difference between the lowest performing area (Eastern region, 33) and the highest performing region (Kathmandu Valley, 55) is remarkable – more than 20 percent points. The students in the institutional schools perform very well in Mathematics, whereas the variety in performance between the community schools is remarkable. There are wide differences between the language groups: the small groups of Sherpa and Newari speakers perform very well in Mathematics (54 and

52 respectively). However, such equally small groups as Rai-, Limbu-, and Magar speakers perform very low (20, 28, and 37 respectively).

As a whole, the socioeconomic status (SES) plays a strong role in the educational processes in Mathematics. The educational level of the parents highly predicts the children's future achievement level in Mathematics. Especially challenging is the situation that both parents are illiterate. If the father or mother or both are coming from an agricultural or related occupation, the students' achievement in Mathematics is significantly lower than with the other occupational groups. The difference in mathematical achievement between the lowest and highest SES groups is remarkable (22–25 percent points). When the children need to work more than 2 hours per day either paid professionally or unpaid in the household work, the achievement level is remarkably low. 2.5% of the students lacked the proper textbook in Mathematics. An alarmingly high number of the students (46%) have encountered bullying in the school within last one month.

3.1.1 Introduction

Mathematics as a school subject – with the Nepali language – is, and will be, the one assessed systematically and frequently in the national assessments of student achievement (NASA) in Nepal. Unfortunately, the results of the previous national assessments (see BPEP, 1997; EDSC, 1997; BPEP, 1998; PEDP, 1998; EDSC, 1999; CERID, 1999; EDSC, 2001; 2003; CERSOD, 2001; EDSC, 2008; Fullbright, 2008) are not comparable with each other because of the missing linking procedure between the tests. Of the previous datasets, only the year 2008 (EDSC, 2008) could be used as a basis for comparison because it is the one where an item-wise data was found. Hence, that dataset is used as a basis for the comparison of 2011 results with the previous one. Several items from the 2008 test were used as the linking items between the years 2008 and 2011. With these linking items it is possible to assess what kinds of changes have happened between the two measurement points.

Another procedure was administered to link the results in Nepal to the international Trends in Mathematics and Science Study (TIMSS) data. Several published items from the international bank were borrowed in order to compare the results in Nepal with an international standard (see in detail Section 2.4). By using the Item Response Theory (IRT) modeling, the 2008 dataset, TIMSS datasets, and the 2011 NASA dataset were linked together to give unique new information on the achievement level in Mathematics.

In what follows, Section 3.1.2 describes the specific methodological solutions in the Mathematics testing – more detailed information on the process is found in Section 2. Section 3.1.3 reveals the basic results in the different content areas of Mathematics in general (Section 3.1.3.1), this deepens to the effects of different

equity matters from the diversity view point (Section 3.1.3.2) and gives some hints on the deeper factors explaining the differences in the Mathematics achievement (Section 3.2.3.3). Finally, the results are condensed in Section 3.1.4.

3.1.2 Methodological solutions

In the process of testing the Mathematics achievement of 8th graders, 16,033 students, 421 schools, 407 head teachers and 407 teachers from 25 districts participated in the survey and testing (for more details of procedures and sampling, see Chapter 2.1). Three test versions (in what follows, M1, M2, and M3) were administered. M1 and M2 were administered in 23 districts simultaneously in the same classroom. M3 was administered in two Himalayan districts, Mustang and Kalikot, two months earlier in a way that one version of two other subjects (Nepali and Social Studies) were administered simultaneously in the same classroom. Hence, more schools were encountered from the Himalaya region though the number of the students was proportionally fewer than in the Hill and Tarai regions.

3.1.2.1 Item analysis and characteristics of the test versions

NASA 2011 is the methodological shift in the National Assessment of Student Achievement. The modern test theory, that is, IRT modeling was used from the beginning of the item construction and preparing the marking scheme for the analysis of the data. In practice, this means, first, certain restrictions on the marking of the papers: no decimal number scores are allowed in the assessment during the answer sheet marking. If the students are not qualified to secure full score, 0.5 score is not provided in any case i.e., students' responses are marked in whole numbers. Obviously, it was found, the marking scheme in 2008 differed from this logic. Hence the 2008 data was adjusted to meet the requirement of IRT modeling: half points were lowered to the *lower* full mark because the students did not show enough achievement to gain the upper full mark. Secondly, the marking scheme has to be more rigorously prepared because of the need to make exact the same judgments in the years to come with the linking items. Thirdly, IRT modeling requires that all the possible marks have to be observed in the dataset; this made some difficulties when analyzing some of the productive items because of the markers' tendency *not* to give full marks to the students. This was not a severe problem in Mathematics compared with the Nepali subject (see Section 3.2). Finally, IRT modeling requires a linking procedure between the different versions in the test. Hence, the common items for each test version, the linking items, were carefully selected from the 2008 test booklet, TIMSS dataset, as well as in the pretest items banked for the item selection. Classical item- and test analysis methods were

used in the pretest phase for finding the percentage of correct answers, item discrimination power, and the test reliability whereas IRT was used for item calibration, finding the latent ability, comparing and equating the versions M1–M2, the dataset from the year 2008, TIMSS database. SPSS software was used for classical Analysis and One Parametric Logistic Model (OPLM) software was used for the IRT modeling.

Versions M1 and M2 were parallel whereas M3 was shorter and easier than the other two. M3 was administered earlier than the other two; because of the willingness to avoid any leaking of the items. All the versions were linked with each other by the use of the identical linking items. The longer versions M1 and M2 were scored out of a maximum of 73 marks and M3 has a maximum score of 62 marks. There were 18 TIMSS items as the linking items all over three versions. The parameters of the international items were fixed during the item calibration so that all the test items of the years 2011 and 2008 were calibrated in the international TIMSS scale (see Sections 2.4 and 4.7 for more details). After the calibration of the items, all the scores in the versions M1–M3 and the version of the year 2008 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 different difficulty levels of the tests have been modeled by using IRT modeling. The original output is the latent ability (Theta, θ) which is a standardized normal score ranging usually from –4 to +4. These values in each test versions were later transformed to equated scores and further, the equated scores were converted in to the percentage of maximum score so that the score 100 means that the student made a perfect score and 0 means that no items were successfully answered. Now onwards, marks or average or mean score refers to the percentage of the maximum marks ranging from 0 to 100.

Table 3.1.1 shows the average marks (mean) calculated for three versions, M1, M2, and M3, separately.

Table 3.1.1 Comparison of the characteristics of mathematics test versions

Version	N	Mean[1]	SD	Remarks
M1	7954	43	21.9	Administered in 23 districts, in same classroom
M2	7860	43	20.6	
M3	219	52	18.5	Administered in Mustang and Kalikot districts
Total	16033	43	21.3	Mathematics as a whole (M1, M2 & M3)

1) percentage of the equated maximum score rounded in whole number

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 items were classified into five categories: *algebra, arithmetic, geometry, sets,* and *statistics*. The number of items asked and the weighting of the items is based on the recommendations of the curriculum for grade 8. Overall internal consistencies (given by reliability) of the whole tests on each version were very high ($\alpha > 0.90$), however, some of the categories (Sets and Statistics & Data) contain a few items and hence the reliability is low. The overall summary of the content wise test analysis is given below. The reliability of the score in the total sample cannot be given in a classical way because it can be estimated only version-wise.

Table 3.1.2 Construct validities and reliabilities of the scores

Topics	Marks			Percentages			Percentage in	Reliability		
	M1	M2	M3	M1	M2	M3	Curriculum	V1	V2	V3
Total	74	76	63	100	100	100		0.92	0.92	0.88
Algebra	22	22	17	29.7	28.9	27.0	**30**	0.79	0.79	0.65
Arithmetic	16	15	15	21.6	19.7	23.8	**20**	0.69	0.71	0.64
Geometry	21	23	18	28.4	30.3	28.6	**30**	0.82	0.78	0.75
Sets	7	8	6	9.5	10.5	11.1	**10**	0.61	0.57	0.50
Statistics & Data	8	8	7	10.8	10.5	11.1	**10**	0.47	0.38	0.29

3.1.2.2 Analytical tools used in the analysis

Basic statistical methods are used throughout the report. The basic tools of statistical description (means, Standard Deviations, percentages, and frequencies), correlations (Pearson's product moment correlation coefficient and Spearman's rank order correlation), and comparison of two means (t-test) as well as statistical inference (*p*-values, effect sizes) are used when appropriate. These methods are described in all basic books of statistical description and -inference. The Analysis of Variance (ANOVA) is used in the General Linear Modeling (GLM) way when several means are compared. All the p-values are corrected by using Multilevel modeling (or Hierarchical Linear modeling) by using SPSS Linear Mixed models module. Decision Tree Analysis (DTA) is used in finding the cut-offs of the continuous variables. All analyses are done in SPSS20 environment.

3.1.2.3 Some statistical concepts used in the text

Within the text, three important concepts should be understood – these are repeated here from Section 2. *Statistical significance*, that is, p-value refers to the possibilities to generalize the results to the population. Practically speaking, behind the p-value (from "probability") is the fact that when measuring human mental processes – as the learning outcomes or attitudes are – there will always be measurement error. This means that the results of each individual student as well as each mean score carry error. Especially when the population is examined by using a sample, all the means carry both measurement error as well as sampling error. For example, there can be a small difference between the boys and girls. In this case, the p-value tells us how probable the same difference would be in the population at large. If the probability is $p < 0.05$, this means that the difference would be found at risk of 5% – only in five samples out of 100 the results would differ from that obtained. Parallel, if the p-value is $p = 0.002$ the risk for a faulty decision (of difference) is only 0.2%.

When the sample size is huge – as the sample of 16,000 students is – the *p*-value very easily gives a signal that the difference between the groups is real in population. *P*-value does not, however, tell whether the difference is small or big. For this purpose there is another statistic, Effect Size (ES). *Effect size* tells how far the lowest and highest groups are from each other. Commonly used indicators of ES are Cohen's d used for two means and Cohen's f for several means (Cohen 1988). Cohen has given boundaries for small, medium and large effect sizes. During the text, these boundaries are used as a "measurement stick" to indicate whether the difference is small, medium/moderate or high. The rough boundaries of the small, medium, and high effect sizes are collected in Table 3.1.3.

Table 3.1.3 Rough boundaries of effect sizes

ES	Cohen's d	Cohen's f
small	< 0.2	< 0.1
medium	round 0.4	round 0.2
high	> 0.8	> 0.4

Hence, in the text one may read that "the difference between boys and girls is statistically significant ($p < 0.001$) but the effect size is small". This means that first, the difference between the boys and girls is real, but second, the difference is, in reality, very small or not notable.

59

Third related concept is the *explanatory power* of the variable. Especially when using the Analysis of Variance (ANOVA) as an analytic tool, the output also gives possibilities to tell how well the grouping factor explains the variation in the data. An A usual indicator for this is *Eta* squared (η^2) which actually is a Correlation between a grouping variable and continuous variable. When the Eta squared is $\eta^2 = 0.30$, this means that the grouping factor (such as the socioeconomic status) explains 30% of the variation in the dataset. Cohen's *f* uses this information strictly:

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

Hence, if $\eta^2 = 0.30$, then $f = \sqrt{(0.3/0.7)} = \sqrt{0.43} = 0.65$ showing very high effect size.

3.1.3 Results

3.1.3.1 Basic results

3.1.3.1.1 Population Analysis

When assessing the student achievement, the population is usually normally distributed in a large sample. Then, 50% of the students lie below average and 50% of the students lie above the average. This kind of distribution can be found in, for example, the Nepali language (Figure 3.1.1).
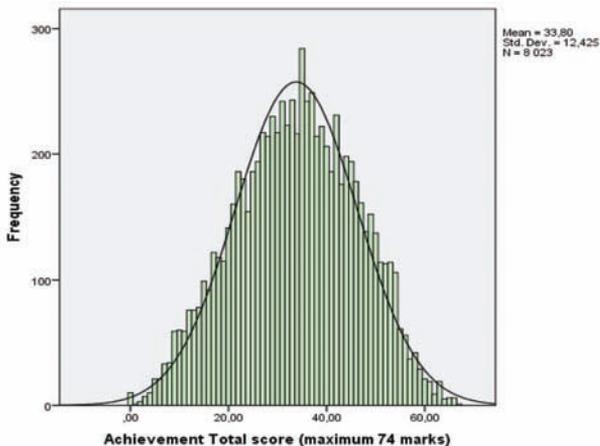

Figure 3.1.1 Normally distributed population of Nepali test

The Mathematics sample was big enough to form a normal distribution (16,033 students). However, Figure 3.1.2 shows that the population is obviously *not* normally

populated. There seems to be three populations: low-performing students, medium-performing students, and high-performing students. The low-performing students seem to achieve 20–25% of the maximum score, the medium-performing students 40–50%, and the high-performing students as high as 70–80% of the maximum score. The majority of the students seem to lie in the low-performing group.
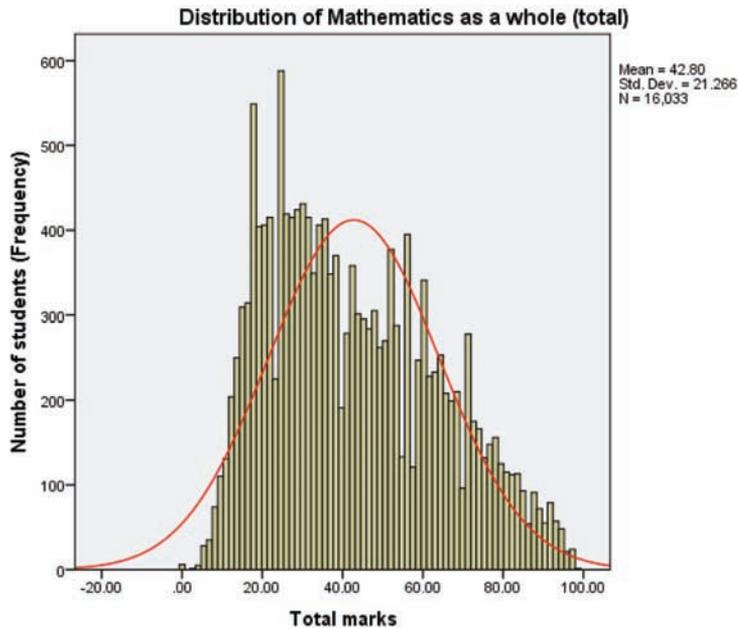

Figure 3.1.3 Non-normally distributed population in mathematics

The dataset shows strong evidence for the fact the 8th graders population is not normally distributed and that there are too many low-performing students.

After separating the schools in to community schools and institutional schools and plotting the population chart, a chart is given in Figure 3.4. The left hand side distribution shows that the community school students are distributed distinctly at least in three populations (groups) in the same way as the total student population. There is a remarkably high number of the students below the 30% of the maximum marks. The right hand side distribution shows the population of institutional school students where the whole system is shifted towards better performing levels however there are quite a number of the students getting very low marks. This indicates that the students in the institutional schools are varying from the low-performer to the highest performer though most of the students are the higher performers.
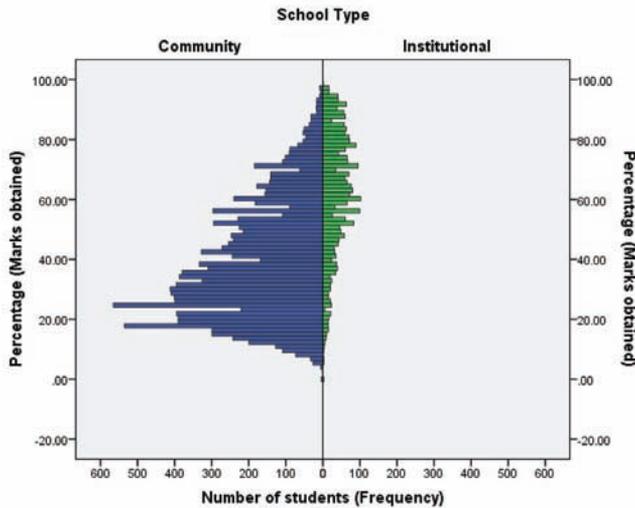
Figure 3.1.4 Distribution of the students' means in community schools and institutional schools

Another related fact is that the schools are clearly divided into two "populations": the highly performing schools and the lowly performing schools. Figure 3.1.5 shows that, on the basis of the school mean of the student performance, there are two categories of schools. One population seen on the left-hand side shows the average of 30% of the maximum score and the other on the right-hand side with the mean of round 70%. The difference between the populations is remarkable. The difference between the school populations can be partly explained by the division between private schools and community schools.
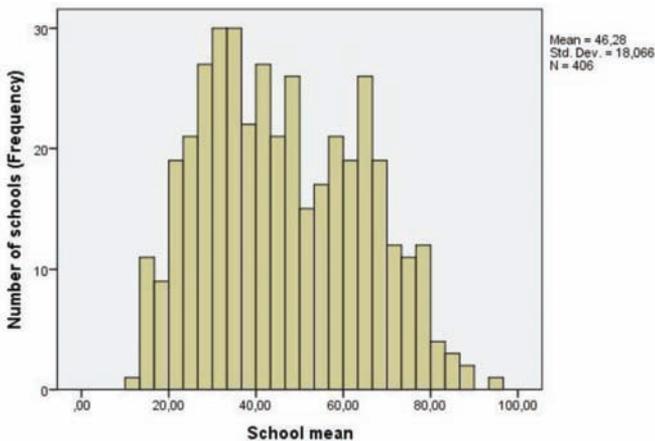

Figure 3.1.5 School means

By analyzing the matter further with the scatter plot, Figure 3.1.6 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 schools to very low-performing schools.
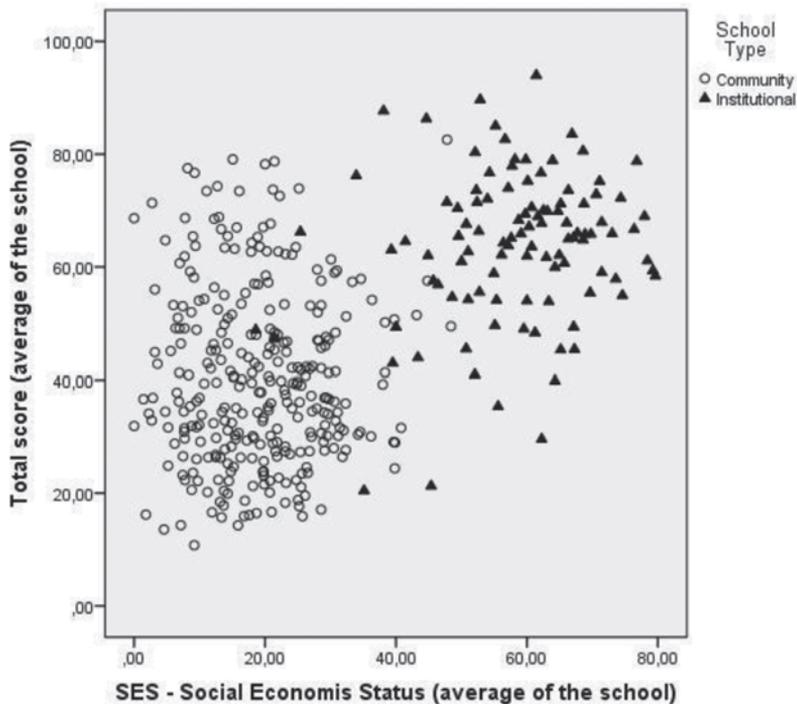
Figure 3.1.6 Achievement and social economic status and type of schools

The dataset gives a strong signal that the students in the institutional schools perform very well and the students in the community schools form two kinds groups of schools: high-performing schools and low-performing schools. The variety between the community schools is remarkable.

3.1.3.1.2 Different content areas and achievement

The whole Mathematics test was the combination of five content areas 1) *algebra*, 2) *geometry*, 3) *arithmetic*, 4) *sets*, and 5) *statistics & data*. The maximum marks of the content areas were proportionally equal to the weighting given by the curriculum. To compare the achievement in all the topics, these sub-scores are

converted into a percentage of the maximum score of the content area. Table 3.1.4 shows the students achievement in Mathematics as a whole (total) and the achievement level in five content areas and Figure 3.1.7 illustrates the differences.

Table 3.1.4 Comparison of the total score average with different topics

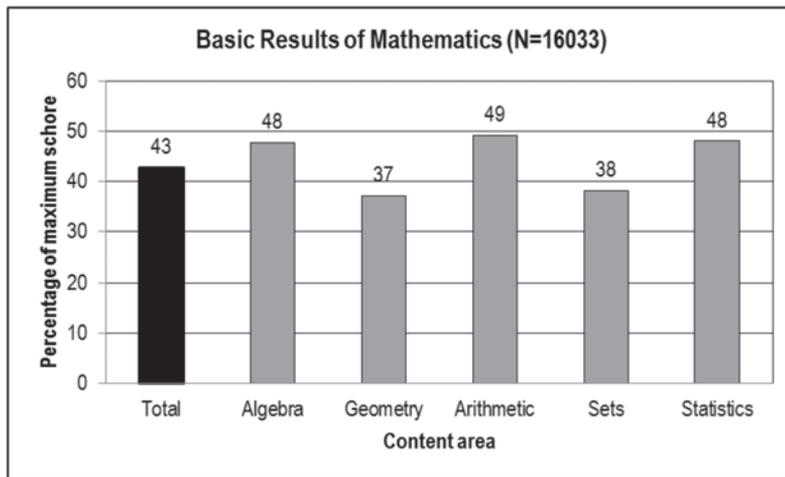| Content area	Mean	SD	Min	Max
Algebra	48	26.1	0	100
Geometry	37	21.8	0	100
Arithmetic	49	22.3	0	93
Sets	38	23.6	0	100
Statistics & Data	48	28.1	0	100
Mathematics as a whole (total)	43	21.3	0	99

Figure 3.1.7 Basic results in the content areas of Mathematics

The percentage of achievement score shows that the national average of Mathematics is 43. Of the different content areas, students are the weakest in geometry (37) and sets (38). They perform better than average in arithmetic (49), algebra (48) and statistics (48).

Dataset gives a signal that the learning outcomes are weaker in Geometry and Sets than in other content areas.

3.1.3.1.3 Hierarchical level of cognitive domain and achievement

The Mathematics test as 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 is shown in Table 3.1.5 and illustrated in Figure 3.1.8.

Table 3.1.5 Achievement level based of hierarchical level of the items

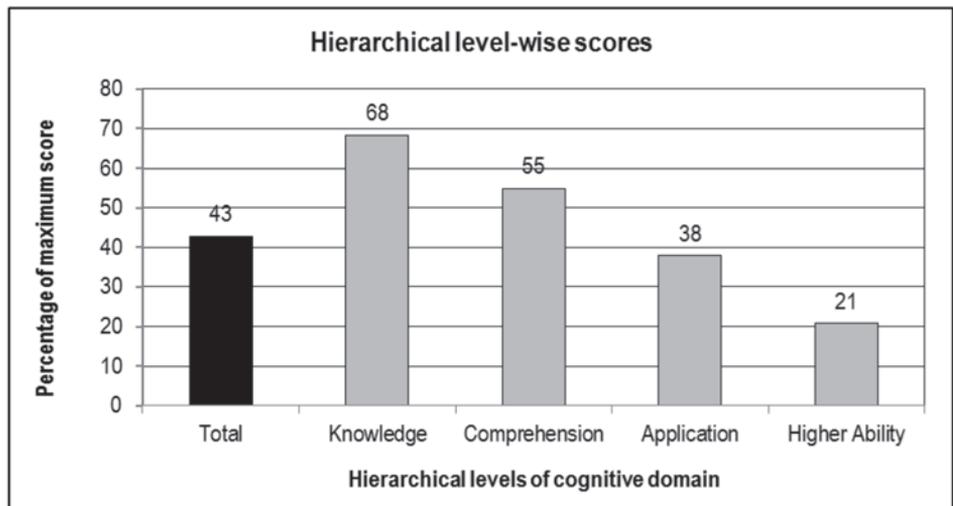| Hierarchical level of item	Mean	SD	SE
Knowledge	68	22.4	0.18
Comprehension	55	24.4	0.19
Application	38	21.9	0.17
Higher Ability	21	22.3	0.18

Figure 3.1.8 Achievement in different hierarchical levels of cognitive domain

Remarkably high a number of students were able to solve only 15% or less of maximum score of the practical problems, that is, the application type of items (17% of the students). In Mathematics, the situation is remarkably more serious compared with the other subjects with the tasks which needed higher cognitive abilities: as many as 53% of the student in Mathematics could solve just less than

15% of the tasks requiring the higher skills (compare with 14% of the students in Nepali, and 12% in Social Studies).

The dataset gives a signal that, in general, the students' ability to solve complex problems is low; only 21% of the maximum scores were reached. Students are much better in the recalling type of questions (68%), which may be explained by the educational system which seems to be geared to remembering the things rather than solving novel problems.

3.1.3.1.4 Type of item and achievement

There were basically two types of questions in the test: objectively marked and subjectively marked items. Objective items covered a wide range of content areas and were very specific to judge because there were only one correct answer or one explicit piece of information was needed to give the correct answer. There were some subjective items on each test version which require a longer procedure to get the full marks. Both the objective and subjective types of items were written 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 level. Table 3.1.6 comprises the basics statistics of the item type-wise achievement levels.

Table 3.1.6 Item type-wise means

Type of items	N	Mean	SD	SE
Objective	16012	59.5	21.42	0.17
Subjective	16033	24.5	23.64	0.19

It is obvious that the subjectively scored tasks – usually those with more demanding requirements for the correct answer – are solved much poorer (25%) than the objective items (60%). Most of the objective items were knowledge-, comprehension- and application type whereas subjective items were application- and higher ability type.

Dataset gives a strong signal that the students are good in recognizing the correct answer and in very fundamental operations, such as the basic manipulation of data and numbers, and calculations with few steps. They are much weaker in reasoning, problem solving, plotting, proving the theory or formula, and constructing the shape and figures. In many cases, the students did not even start to do the open ended questions – hence, the low score.

3.1.3.1.5 Comparison of 2011 with 2008 achievement result

The national assessment carried out in various years aims to assess the achievement and the progress over the period of the years. The National Assessment of grade 8 students carried out by the Research Centre for Education Innovation and Development (CERID, 1999) shows that the national average of the students is 28.87 (≈ 29). Later, the National Assessment of grade 8 students carried out by the Educational Development Service Centre in 2008 (EDSC, 2008) showed the national average of the students to be 31.7 (≈ 32). The National Assessment of 2011 carried out by the NASA Unit of ERO (MOE) shows that the national average of Mathematics in grade 8 is 43. These figures are coming from *Classical Test Theory (CTT) 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 different difficulty level of the tests.*

The Ministry of Education has made the decision of using IRT modeling in NASA to make the results of the national assessment comparable with the previous and to-come datasets. In NASA 2011, the tests were linked with the test of 2008 so that the comparison was possible. Tests of 2008 and 2011 were equated to the same scale by using international parameters (fixed difficulty parameters) from the TIMSS dataset. Certain selected items from the 2008 test were used to make the linkage between the tests. The results of the year 1999 were not used in comparison because the whole database was missing.

By using the IRT modeling, the total scores of the year 2008 were equated with the total score of the year 2011datasets. After equating, the results are as condensed in Table 3.1.7; the scores are somewhat 4 percent points weaker in the year 2011 than in 2008.

Table 3.1.7 IRT modeled comparable means of 2008 and 2011

Year	N	Mean	SD	SE
2008	2640	46.9	16.8	0.32
2011	16033	42.8	21.3	0.17
Total	18673	43.4	20.7	0.15

The reduced level of skills can be illustrated by comparing the percentages of correct answers in the identical linking items (Table 3.1.8). In all items except the most demanding, the students in 2011 performed lower than the students in 2008. In some cases, the difference is remarkable.

Table 3.1.8 Comparison of the linking items of the years 2008 and 2011

	Percentage of correct answers		
	2008(N = 2640)	2011(N = 7954)	Change
2008_Q1	0.80	0.67	−0.13
2008_Q2a	0.71	0.45	−0.26
2008_Q2b	0.87	0.70	−0.17
2008_Q2c	0.54	0.39	−0.15
2008_Q3a	0.29	0.22	−0.07
2008_Q3b	0.43	0.41	−0.02
2008_Q4a	0.60	0.43	−0.17
2008_Q4b	0.54	0.37	−0.17
2008_Q18	0.31	0.71	+0.40

This result is interesting and it raises the question of what could explain the decreased achievement? One possible explanation is that the marking schemes have changed somewhat – though not remarkably – between the years. In that case the markers in 2008 could have been looser in their scoring compared with the stricter markers in 2011. This phenomenon was evident in the dataset of Social Studies (see Section 3.3.1). Another explanation for the decreasing trend could be that, according to the latest household survey, during the last ten years, more low-status students have been involved in the educational system. If this has happened during the last four years, it may have caused a temporary decrease in the achievement level.

The Mathematics as a whole (total) and the content area-wise comparison between the years is seen in Figure 3.1.9.
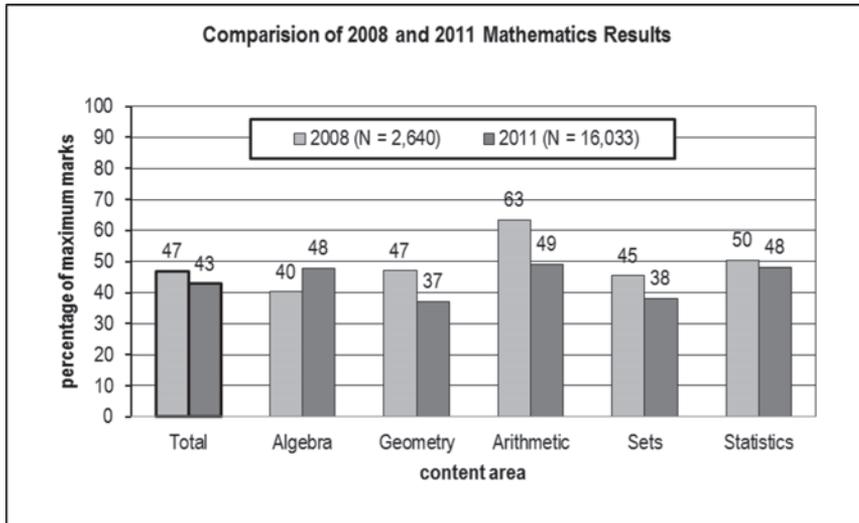
Figure 3.1.9 Content-wise differences between years 2008 and 2011

In all content areas except Algebra, the achievement level in 2011 is lower than that in 2008. As a total, the reduction is 4 percent points. The possible reasons are discussed above.

The dataset gives a signal that the mathematical skills have not been developed as is expected. In all content areas, the results are slightly lower in the year 2011.

3.1.3.1.6 Comparison with the international standard

The NASA 2011 was made comparable with the international TIMSS assessment. A good number of items (18) of the released TIMSS items were used as linking items. Their known difficulty parameters were fixed in the calibration of the local items. Hence the international average of $\theta = 0$ was fixed in the Nepalese datasets; when a student's ability level in NASA 2011 would be zero it corresponds to the average level of the international students.

Figure 3.1.10 shows the comparison of the 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 indicates the international average. As the ability is below the average, the bars are going down whereas when the ability is above the international average, the bars are going upwards.
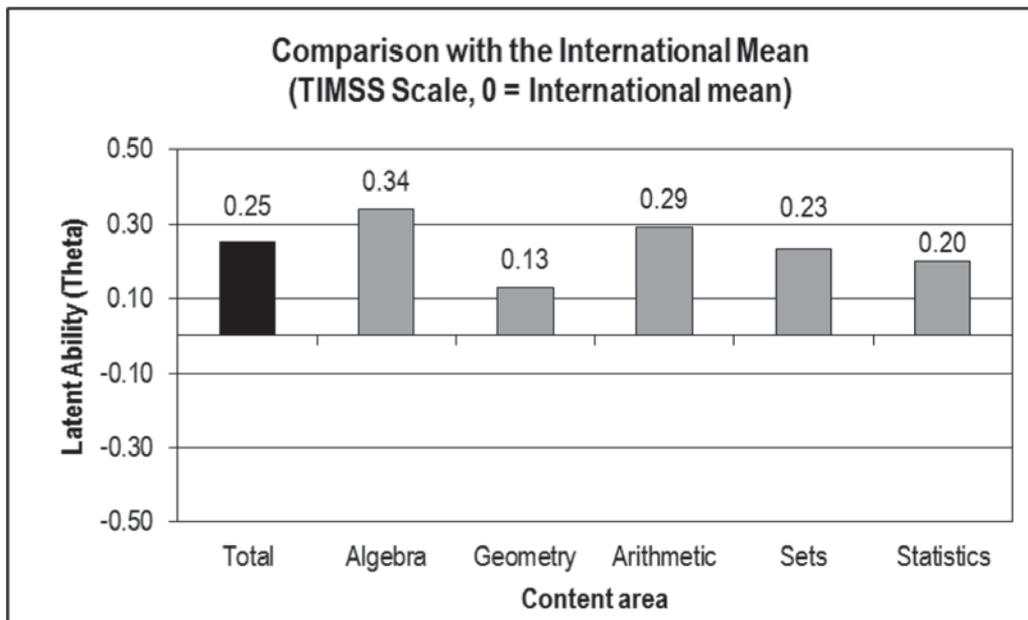
Figure 3.1.10 Student achievement in the international TIMSS scale

Figure 3.1.10 shows that the average ability shown by Nepali students in Mathematics as a whole (*Total*) is above the average line.[11] This indicates that the students in Nepal are somewhat better in Mathematics subject when compared with the international average. Actually, in all content areas the results are higher than the international mean. On the basis of the raw transformation of the TIMSS

---

[11] This comparison is based on item parameters published by TIMSS in http://timss.bc.edu/timss2007/ mathreport.html. Although the comparison is based on equating the tests with a good number of selected linking items, the high result leads into two deviating ideas. On one hand, the mathematical skills are – in general – appreciated in Nepal and the curricula are very demanding in Mathematics. Hence, the result of average (or higher) achievement could be acceptable. On the other hand, the result of the Indian students in 2012 PISA inquiry casts a shadow over the high (or average) result in mathematics in Nepal. Assuming the results in Nepal are not much higher than in India – and India was clearly below the average in Mathematics and Reading. Now, in Reading the level in Nepal is the same kind as in India, that is, much lower than the average (see Section 3.2.3.1.6). The method of equating used in Mathematics was exact the same as in Reading (see Section 2.4.1). Hence, what may explain the difference in Indian and Nepalese results? One good explanation is that the selection of the items was done on the different basis in Nepali than in Mathematics and Geography. In the latter subjects, only such items were selected which fits the local curriculum where in Nepali, the Reading items were selected more or less randomly. It is advisable to use the same idea also in Mathematics in the assessments to come.

scale (of mean 500 and Standard Deviation 100) back to Theta values, this ability level of $\theta = 0.25$ of Nepal seems to be in the range of some developed countries like Hong Kong, Japan, Hungary, England. The latter comparison is not accurate even though the results are otherwise quite robust because of the good number of test items. More comparable research may be needed to confirm the result.

Dataset gives a signal that, from the international comparison viewpoint, the average achievement of Mathematics in Nepal seems to be better than the international average (+0.25). In all content areas the results are higher than the international mean. It is, however, good to be cautious of the possible error in the estimation because of the process of the item selection.

3.1.3.2 Results based on diversity factors

Diversity is a relative and contextual term. In the context of Nepal, experts have defined six to eight diversities in Nepal (see Section 4.1): geographical/ecological-, language-, gender/sex-, religious-, ethnic-, cultural-, disability- and economic diversity. NASA 2011 background information questionnaire included six of the above diversities; two diversities, 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) diversity. These factors can be taken as equity factors; all children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goal.

3.1.3.2.1 Districts and student achievement

Out of 75 districts, 25 were randomly selected to represent the Ecological zones and Developmental regions and ultimately 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 condensed in Table 3.1.9 and Figure 3.1.11. The table shows the achievement in ascending order according to the achievement. The *mean* represents the average achievement percentage of the particular district.

Table 3.1.9 Average achievement score in the selected districts

| Districts	Mean	N	SD	Districts	Mean	N	SD
Bhojpur	18	484	9.7	Banke	45	810	20.4
Pyuthan	29	498	14.1	Nawalparasi	45	986	22.7
Ilam	30	896	15.1	Mustang	45	35	17.0
Sankhuwasabha	32	241	15.9	Bajhang	46	271	21.1
Kanchanpur	32	698	17.1	Tanahu	47	936	21.0
Morang	34	1041	17.3	Gorkha	51	853	17.9
Jajrkot	35	496	14.6	Bara	51	842	20.1
Sindhupalchowk	38	370	17.7	Lalitpur	52	272	18.6
Syanjha	39	977	18.7	Kalikot	53	184	18.5
Doti	39	359	17.2	Kathmandu	55	1375	19.6
Jhapa	40	1100	20.3	Bhaktapur	61	171	22.8
Ramechhap	42	581	22.7	Rautahat	66	775	15.5
Dhading	43	782	20.0				
				Total	43	16033	21.3
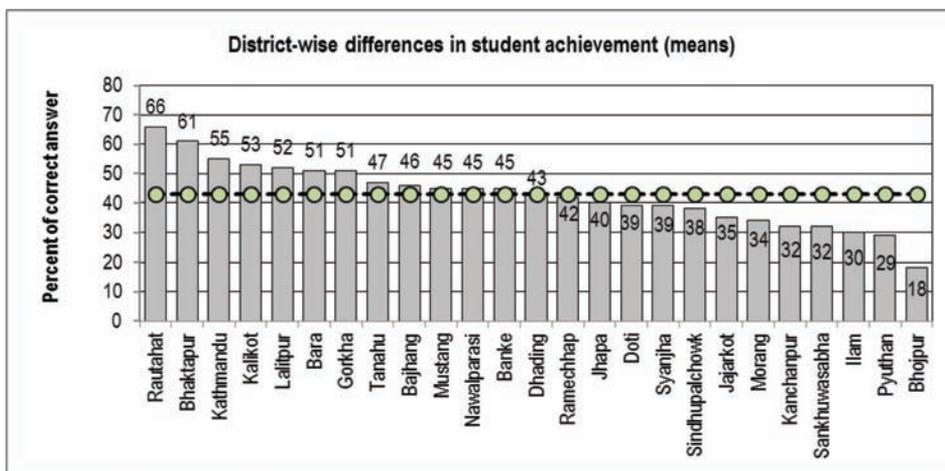
Figure 3.1.11 Achievement in the districts in the sample

72

Of the randomly selected districts in the sample, the student performance was very low in Bhojpur (18), Illam (30) and Sankhuwasabha (32) from the Eastern region, in Pyuthan (29) from the Mid-Western, and in Kanchanpur (32) from the Far-Western region. Except for Kalikot district (53), the best performing five schools come from the Central region and specifically in the Valley area: Rautahat (66)[12], Bhaktabut (61), Kathmandu (55), and Lalitpur (52). The data mining module of the SPSS software – Decision tree analysis – points out that lowest performing schools can be found in the Eastern Hill area.

The difference in achievement due to the different district is statistically significant ($p < 0.001$). The variation explained in achievement due to district is $\eta^2 = 0.23$; the district explains 23% of the variation in the data. Effect size is $f = 0.54$ indicating that the difference between the lowest performing district (18) and highest performing district (66) is remarkably high.

The dataset gives a strong signal that there is a wide difference between the districts when it comes to the equal opportunities of children to reach the preset goals in Mathematics. The results are bound to the 25 districts selected randomly in the sample; even lower-performing districts could be found if other districts would have been selected. In any case, the results in Bhojpur (18), Pyuthan (29), Illam (30), Sankhuwasabha (32), Kanchanpur (32), Morang (34), and Jajarkot (35) is so low that raising the standard in these districts would arise the standard in the whole country.

3.1.3.2.2 Ecological zone and student achievement

Access to educational facilities can play a vital role in students' achievement. The Mountain, Hill and Tarai are three geographical features in Nepal though the Valley can be taken as a special geographical feature because of it 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, weather conditions, economic activities, aggressive development, as well as the dense human capacity make the Valley a unique fourth geographical area in the analysis. The variation in the Ecological zones in NASA 2011 is condensed in Table 3.1.10 and Figure 3.1.12.

---

[12] Several readers of the draft of the report noted that, based on SLC results, Rautahat seems to be too high in ranking. The reason for the discrepancy was tried to found out by contacting the DEO but no explanation was found.

Table 3.1.10 Achievement in the Ecological zone

Region	N	Mean	SD	SE	Minimum	Maximum
Mountain	1101	41	19.8	0.60	5.5	95.9
Hill	6862	38	19.9	0.24	0.0	97.3
Valley	1818	55	19.9	0.47	0.0	98.6
Tarai	6252	44	21.8	0.28	0.0	97.3
Total	16033	43	21.3	0.17	0.0	98.6
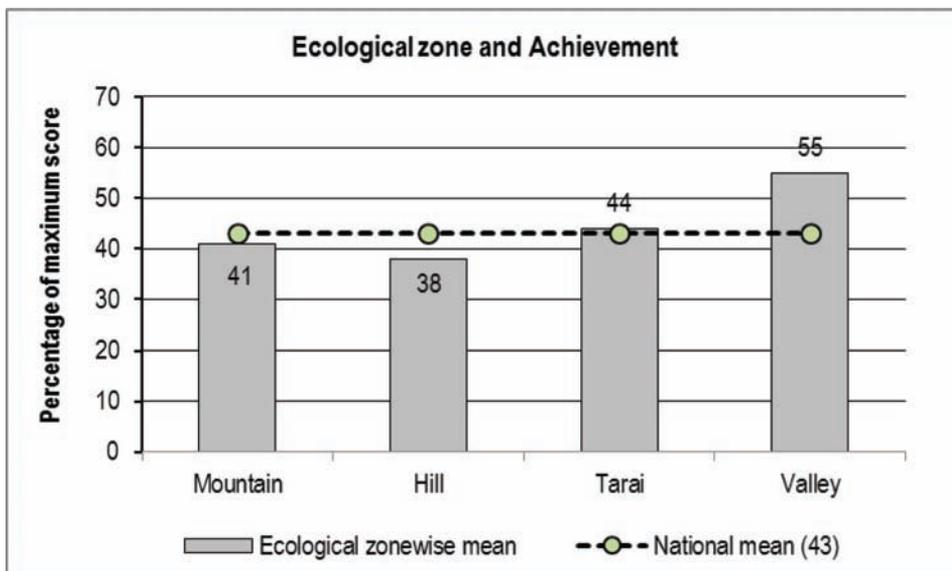
Figure 3.1.12 Differences in achievement in different Ecological zones

The data shows that, on average, the students from the Valley (55) outperform the students from all the other Ecological zones. The students from the Hill areas perform the lowest (38). When knowing that the students' performance in Kalikot (55) and Mustang (45) was higher than the average, there seems to be a wide difference between the districts of the Mountain areas.

The achievement in the regions differs significantly ($p < 0.001$). Tukey's *post hoc* test tells that all the zones deviate from each other in a statistically significant manner of at least $p = 0.05$ level. The effect size is $f = 0.25$ showing moderate difference between the highest and lowest performing Ecological zones; Ecological zone explains 5.7% of the variance in the data. As a comparison, the district explains more than 20% of the variation.

Dataset gives a signal that there is a moderate difference between the student performances in four Ecological zones. Students in the Kathmandu Valley outperform the other students. The achievement is the lowest in the Hill area.

3.1.3.2.3 Developmental region and student achievement

The student achievement varies according to the Developmental regions which are divided into 1) Eastern-, 2) Central-, 3) Western-, 4) Mid-Western-, and 5) Far-Western regions. Additionally, the Kathmandu Valley is taken as the 6th Developmental region though technically it falls under the Central Developmental region. The mean achievements in the Developmental regions are given in Table 3.1.11 and illustrated Figure 3.1.13.

Table 3.1.11 Achievement in the Developmental regions

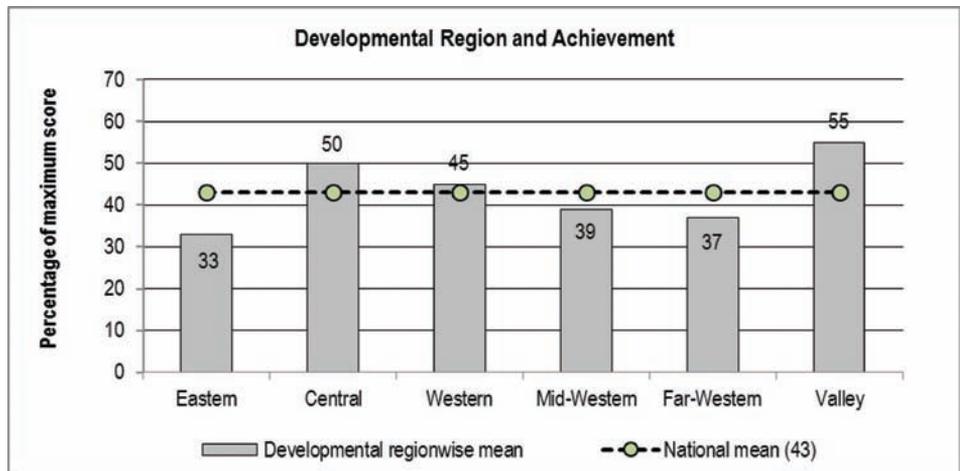| Region	N	Mean	SD	SE	Minimum	Maximum
Eastern	3762	33	18.2	0.30	0.0	97.3
Central	3350	50	21.7	0.38	0.0	97.3
Western	3787	45	20.7	0.34	4.1	97.3
Mid-Western	1988	39	19.2	0.43	0.0	95.9
Far-Western	1328	37	18.8	0.52	5.5	91.8
Valley	1818	55	19.9	0.47	0.0	98.6
Total	16033	43	21.3	0.47	0.0	98.6

Figure 3.1.13 Comparison of student achievement in the Developmental regions

75

The best performance can be found in the Valley (55) and in the Central region (50). The performance is the lowest especially in the Eastern region (33) though the Far-West (37) and Mid-Western (39) region are also regions were student performance is lower than the average. The difference between the regions is statistically significant ($p < 0.001$) and Tukey's *post hoc* test shows that all the Developmental regions differ from each other significantly with at least at 5% significant level. Effect size is moderate or high ($f = 0.37$); Developmental region explains 12% of the student variation.

The dataset gives a strong signal that there is waste inequity in the Developmental regions of children's opportunities to reach an adequate level of Mathematics. The difference between the lowest performing area (Eastern region, 33) and the highest performing region (Kathmandu Valley, 55) is remarkable – more than 20 percent points.

3.1.3.2.4 School type and student achievement

All the schools are categorized into community-managed- and institutional-managed schools. In what follows, these are referred to as community schools and institutional schools; the latter could also be called private schools. The difference in the achievement of the two types of schools is condensed in Table 3.1.12.

Table 3.1.12 Type of school and the average achievement

Type of school	N	Mean	SD	SE
Community	13,278	39	19.1	0.17
Institutional	2,755	63	20.0	0.38
Total	16,033	43	21.2	0.17

The achievement levels in the community schools and institutional schools differ from each other remarkably. The average performance in the private schools is 63 whereas in the community schools it is 39; 24 percent point's difference is remarkable. The difference is statistically significant ($p < 0.001$) and the effect size is high ($f = 0.47$) showing a wide difference between the community- and institutional schools. Division of the children to the community- and institutional schools explains 18% of the student variation ($\eta^2 = 0.182$). From Figure 3.1.6 above it is known that the deviance within the community schools is remarkable

ranging from 20% to 80%; contrarily, most private schools in the sample show very high performance. Natural reason for this is the student selection. The selection of the students evidently takes the best students from the community schools and hence the average performance of the community schools may be lower when there are possibilities to select the school. It may be worth noting that the best community schools are not in Kathmandu Valley. It is also worth noting that the number of the students in the private schools is not very high; the main educational work is done in the community schools.

The dataset gives a strong signal that the students in the institutional schools outperform the students in the community schools. Most probably this deviance can be explained by the student selection, however, this cannot be known on the basis of the dataset.

3.1.3.2.5 School location and student achievement

One of the strata of sampling in NASA 2011 was the school location. The schools were divided into two: rural schools and urban schools. The achievements of the students in rural and urban schools are condensed in Table 3.1.13.

Table 3.1.13 Student achievement on the basis of location of school

Type of school	N	Mean	SD	SE
Rural	11899	41	20.5	0.19
Urban	4134	48	22.5	0.35
Total	16033	43	21.3	0.17

The achievement level of the students in the urban schools (48) is somewhat higher than that of rural schools (41). The difference in average score is significantly different ($p < 0.001$) though the effect size is not high ($f = 0.16$); division into rural and urban schools explain only 2.4% of the student variation ($\eta^2 = 0.024$).

Data gives a positive signal that though the results are somewhat better in cities than in the rural area, the difference is not remarkably high. From the equity viewpoint, this is a positive thing.

3.1.3.2.6 Language at home and student achievement

In the context of Nepal, the student achievement may depend on the language spoken in their homes i.e., the mother tongue. In many cases, the mother tongue reflects the ethnical background and hence any difference may be taken as a possible source for inequity in society. In Nepali language assessment (see Sections 3.2 and 4.6) the results are obvious. However, it is not known what the effect may be in the subject of Mathematics.

On the basis of the total data, 30.4% of the 8th 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 Magars (3.2%), Tamangs (3.1%), and Tharus (2.2%). After dividing the languages into ten groups excluding Nepali, there were still 18.9% of the students classified into the group "else". 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 were grouped into "non-Nepali speakers". The results are condensed in Tables 3.1.14 and 3.1.15 and Figure 3.1.14.

**Table 3.1.14 Student achievement in the different language groups**

Language group	N	Mean	SD	SE
Nepali	11356	43	21.2	0.20
Non-Nepali	4677	42	21.5	0.31

When combining all the minor language groups as "Non-Nepali", there is no difference between the language groups. However, on the basis of Table 3.1.15 with the original categorization of the minor languages, the issue looks very different: It is evident that the small groups of Sherpa and Newari speakers perform very well in Mathematics (54 and 52 respectively). It is also noteworthy that the equally small language groups, such as Rai- , Limbu-, and Magar speakers perform very low (20, 28, and 37 respectively). These communities would need remarkable support to raise their level.

Table 3.1.15 Achievement in the different language/ethnic groups

Language/Etnicity[1]	N	Mean	SD	SE	Maximum	Minimum
Sherpa	13	54	26.1	11.0	7.23	84.9
Newar	167	52	20.0	13.7	1.55	98.6
Gurung	137	45	20.1	8.2	1.72	83.6
Tharu	446	45	21.9	5.5	1.04	95.9
Nepali	11,356	43	21.2	0.0	0.20	97.3
Tamang	461	41	20.4	5.5	0.95	94.5
Magar	397	37	18.4	6.9	0.92	93.2
Limbu	51	28	14.4	12.3	2.02	95.9
Rai	153	20	12.1	5.5	0.98	82.2
Other language	2,726	43	21.8	0.0	0.42	97.3

1) those language groups are omitted with less than 13 students
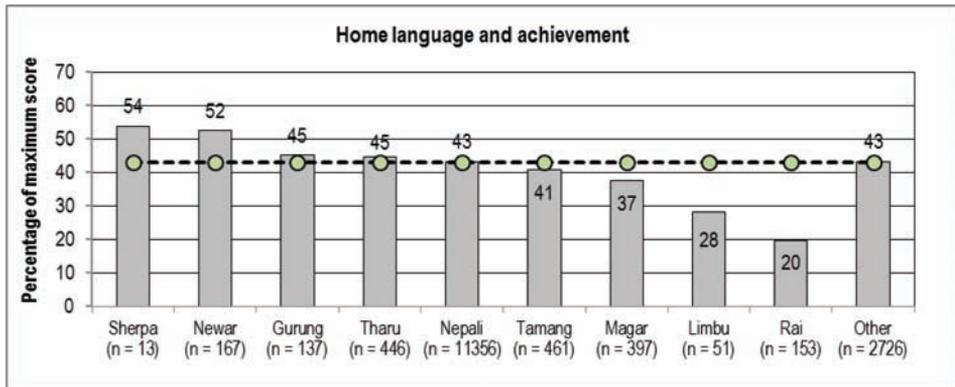
Figure 3.1.14 Relation between language at home and achievement

The difference between the language groups is statistically significant ($p < 0.001$). However, the effect size is small ($f = 0.14$) because the minority groups are small; division into smaller language group cannot explain the variation in the data ($\eta^2 = 0.018$). When analyzing only the minority languages and hence, excluding the Nepali speakers and the group "else", the effect size is high ($f = 0.41$) indicating really remarkable difference between the highest performing group (Sherpa, 54) and the lowest performing group (Rai, 20).

Language and developmental region

When combining the results from the Developmental region and mother tongue, the Eastern region seems to carry a couple of interesting characteristics. First, within *all* the language groups with more than 13 students, those students coming from the Eastern region perform lower than the average of the language group (Table 3.1.16). In some cases, the difference between the groups is remarkable. For example, in the Tamang population the average is 41 but those Tamang students coming from the Eastern region gain less than 20. The same is seen in Tharu population (45/27) and in Newar population (52/34). Second, with the exception of the Gurung population which was mainly reached from the Western region, the *poorest* results within the language group were found in the Eastern region.

Table 3.1.16 Achievement in the different language groups in different regions

Developmental region	Nepali[1]	Magar	Tharu	Tamang	Newar	Rai	Gurung	Sherpa	Limbu
1 Eastern	33.1[2]	30.4	27.4	19.9	34.2	18.7	43.8	45.0	27.8
2 Central	47.8	37.5	71.4	41.1	37.6	82.0	31.8	76.7	53.4
3 Western	46.1	36.9	49.2	51.0	41.9		47.0	58.9	
4 Mid-Western	38.8	42.0	49.6		56.5		18.1		
5 Far-Western	39.6	21.9	35.4		76.7			53.4	
6 Valley	56.2	62.5	65.3	41.3	54.5	49.3	60.8	60.7	
Total	43.1	37.5	44.6	40.7	52.4	19.5	45.1	53.8	28.3

1) The language groups of less than 14 students are not included in the table.
2) The main population is highlighted by the gray shade. In some un-highlighted cases, there may be only one student behind the mean.

Language and Ecological zone

As appear to happen in the Eastern region, also happen in the Hill region: in all language groups the Mathematics results in the Hilly region are lower than the National average and with the exception of the Gurung population, in all the language groups the lowest scores are found in Hill region (Table 3.1.17). The reason for the phenomenon stays open. However, it means that the home language alone does not determine the future of the children. As a single example, take the only Rai speaking student from the Central region whose score was 82.

Table 3.1.17 Achievement in the different language groups in different zones

Ecological zone	Nepali[1]	Magar	Tharu	Tamang	Newar	Rai	Gurung	Sherpa	Limbu
1 Mountain	41.9	44.3		37.0[2]	44.7		35.6	59.2	29.7
2 Hill	39.1	36.5	30.1	34.1	42.1	18.0	45.0	34.7	24.3
3 Valley	56.2	62.5	65.3	41.3	54.5	49.3	60.8	60.7	
4 Tarai	44.4	37.7	44.5	62.1	50.7	22.8	34.9	58.9	31.8
Total	43.1	37.5	44.6	40.7	52.4	19.5	45.1	53.8	28.3

1) The language groups of less than 14 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.

The dataset arises two difficult questions: What has been done differently – or not done at all – in the Eastern Hill area where Rai-, Limbu-, Tharu-, Tamang-, and even Sherpa students get very low scores? Why are the results in the Hill area poorer than in other regions? To answer these questions, one needs deeper knowledge of the local situation than what is available in the dataset.

3.1.3.2.7 Caste and student achievement

Modern education has been influenced in several ways by the legacy of the historical caste system. Though the caste system is officially abandoned it still lives in the mind-sets of most Nepali people. Historically, the Brahmans and Cheetris have been heavily involved in education, but Dalits, for example, have been practically outside of the educational system. Hence, modern society has made lots of efforts to make the education possible and accessible for all children. The latest household survey (CBS, 2012) shows that the number of Hill Dalits has been increased remarkably in the lower education system but their number in the secondary and higher education is still small. The results concerning the castes and achievement are condensed in Table 3.1.18 and illustrated in Figure 3.1.15.

Table 3.1.18 Achievement in the different castes

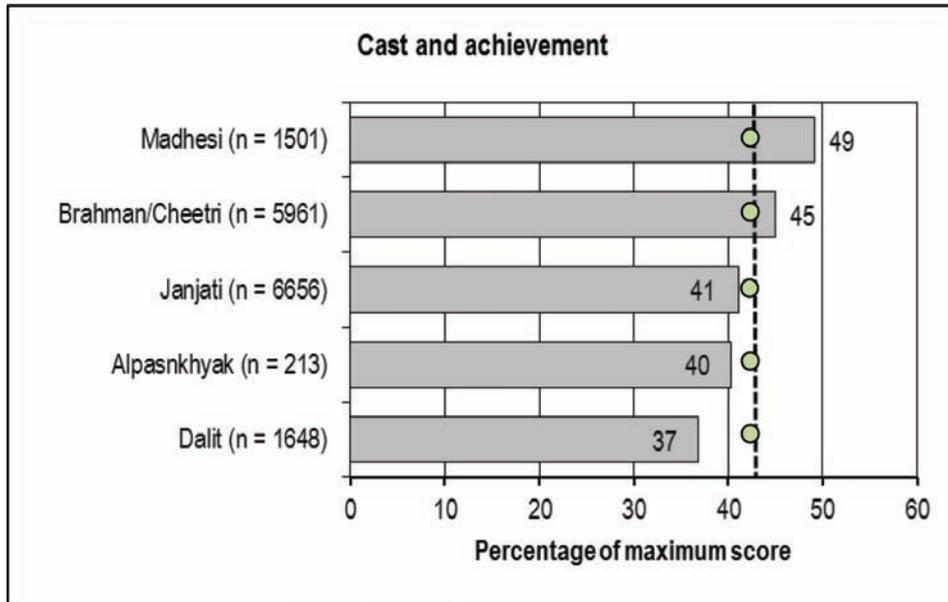| Caste	N	Mean	SD	SE	Minimum	Maximum
Brahman/Cheetri	5961	45	21.6	0.3	0.0	98.6
Janjati	6656	41	20.9	0.3	0.0	97.3
Dalit	1648	37	18.6	0.5	5.5	97.3
Madhesi	1501	49	22.2	0.6	0.0	97.3
Alpasankhyak	213	40	19.4	1.3	4.1	87.7
Total	15979	43	21.3			

Figure 3.1.15 Relation between language at home and achievement

On the basis of NASA 2011, it is obvious that the Dalits are still performing lower in Mathematics (37) than the other castes (40 or higher) and especially lower than Madhesis (49) and Brahmans/Cheetris (45). The overall difference between the groups is statistically significant ($p < 0.001$) but the effect size is small or moderate ($f = 0.15$); dividing students according to their caste background explains only 2.3% of the student variation ($\eta^2 = 0.023$). Tukey's *post hoc* test reveals that Alpasankhyak and Janjatis do not differ from each other. Otherwise all the means of the cast groups differ from each other statistically significantly with at least a 5% risk ($p < 0.05$).

A positive sign from the equity viewpoint is that Dalit students perform remarkably better than the national mean in the Mid-Western Mountain- (58) and Central Tarai areas (53) and the results are above average in the Mid-western Tarai area (44). Unfortunately, it though seems that the results are much lower than the average in the Eastern Hill- (21), Eastern Mountain- (26), Mid-Western Hill- (28), and Far-Western Tarai areas (28). The number of the students in certain strata is small and hence it may be wise not to make the too strong and farfetched implications of the results.

Table 3.1.19   Dalit students' achievement in Ecological zones and developmental   regions

	Eastern	Central	Western	Mid-Western	Far-Western	Total
Mountain	26.2	39.5	43.4[1]	57.9	36.0	40.4
Hill	20.9	42.6	38.5	27.7	32.8	34.2
Tarai	32.0	53.4	32.3	44.5	27.8	38.8
Total	27.4	47.8	37.2	36.6	30.7	

1) Highest means are highlighted

Dataset gives a signal that Dalit students' performance is lower than the results of the other castes. Their performance is especially low in the Eastern Developmental region. A positive signal is that the Dalits' performance is very high in the Mid-Western Mountain- (58) and Central Tarai areas (53).

3.1.3.2.8 Sex and student achievement

Lots of effort has been put globally into reducing the difference between boys' and girls' school achievement. In all previous assessments, there has been a significant difference between the sexes. Because the sex- or gender-wise equity seems to be important in the modern discourse, the matter is handled somewhat more extensively than the previous sections of equity. Basic results are condensed in Table 3.1.20 and Figure 3.1.16.

Table 3.1.20 Student achievement of boys and girls

Sex	N	Mean	SD	SE	Minimum	Maximum
Boys	7963	45	21.4	0.24	2.7	98.6
Girls	8043	41	21.0	0.23	0	97.3
Total	16006	43	21.3			
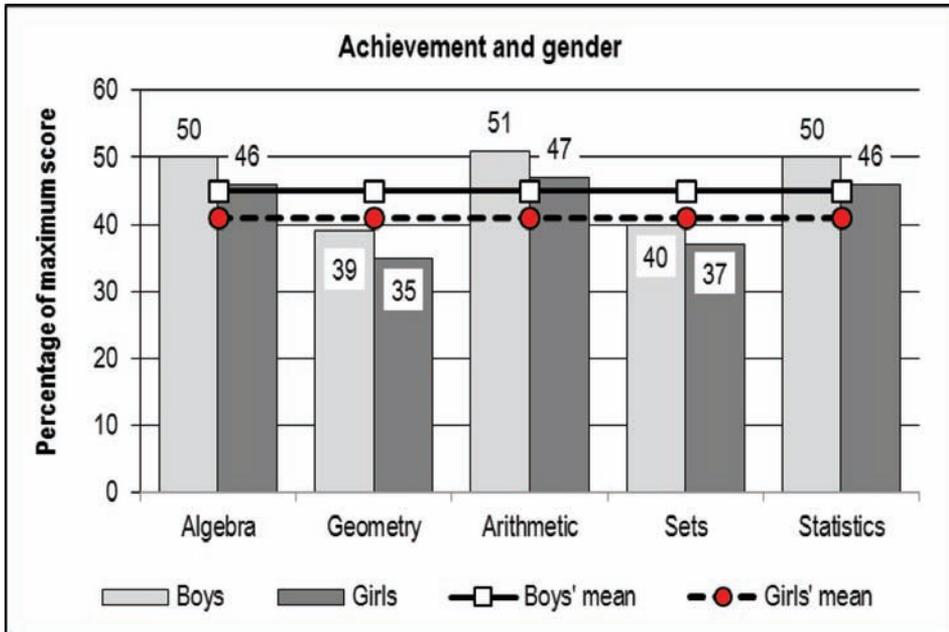
Figure 3.1.16 Comparison of boys and girls in different content areas

In NASA 2011, there still is a statistically significant difference between boys (45) and girls (41) when it comes to Mathematics achievement ($p < 0.001$). It is noteworthy that in all content areas boys outperform the girls. However, the effect size is small ($f = 0.09$) indicating that the difference is not remarkable. Sex explains only 0.8% of the student variation. From the equity point of view the signal is positive thought there is still work to be done to reduce the gap.

Sex and caste

On the basis of the NASA 2011 data, the difference between boys and girls is the highest within the Brahmans and Cheetris (difference is 5 percent points) and the Dalits (4 percent points). Tukey's *post hoc* test shows that the differences are statistically significant with at least a 5% risk level. There is practically and statistically no difference within the Alpasankhyak (1 percent points) and Madhesis (2 percent points) (Figure 3.1.17).

84
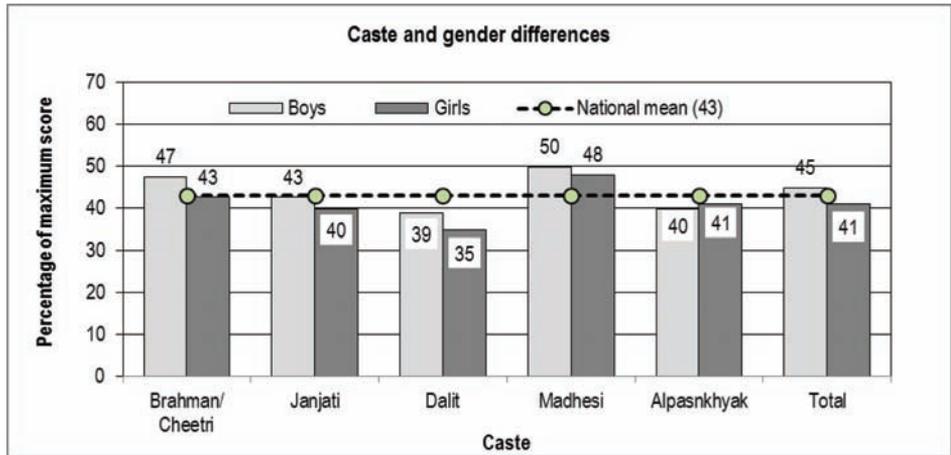
Figure 3.1.17 Caste/ethnicity-wise compression of achievement of boys and girls

Sex and Ecological zone

The achievement of girls and boys in Mathematics differs significantly among ecological belts ($p < 0.001$). The phenomenon is interestingly somewhat more complex, however. Namely, In Mountain-, Hill-, and Tarai regions boys outperform girls but in the Valley, there is no statistically significant difference between boys and girls though girls in the sample are somewhat better than the boys (Figure 3.1.18).
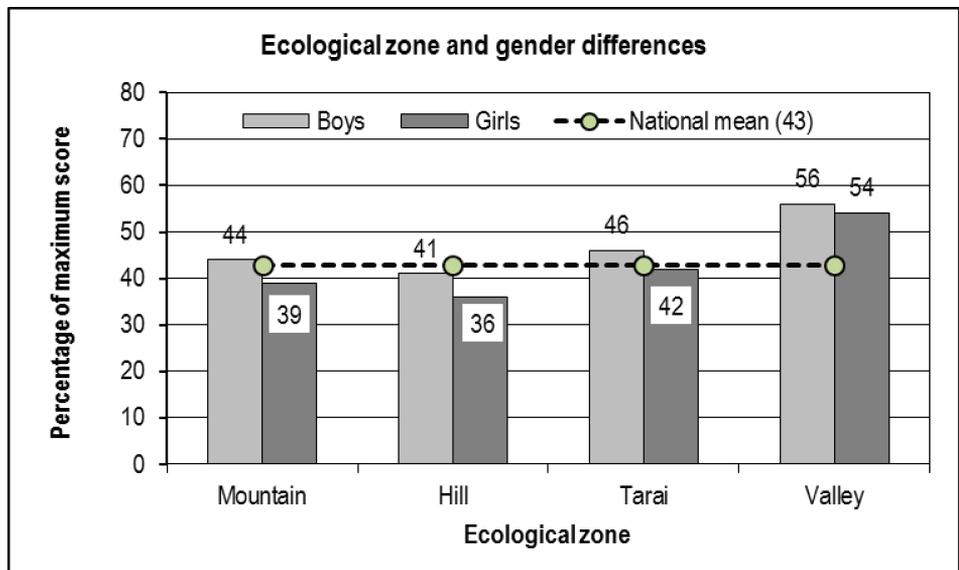

Figure 3.1.18 Ecological zone and gender-wise differences

85

Sex and developmental region

The difference between the boys and girls is the widest in the Far-Western region (7 percent points) and in the Western region (6%) and notably smaller in the Eastern- (2%), Central- (3%), and Mid-Western regions (2%). All the differences are statistically significant ($p < 0.001$) (Figure 3.1.19).
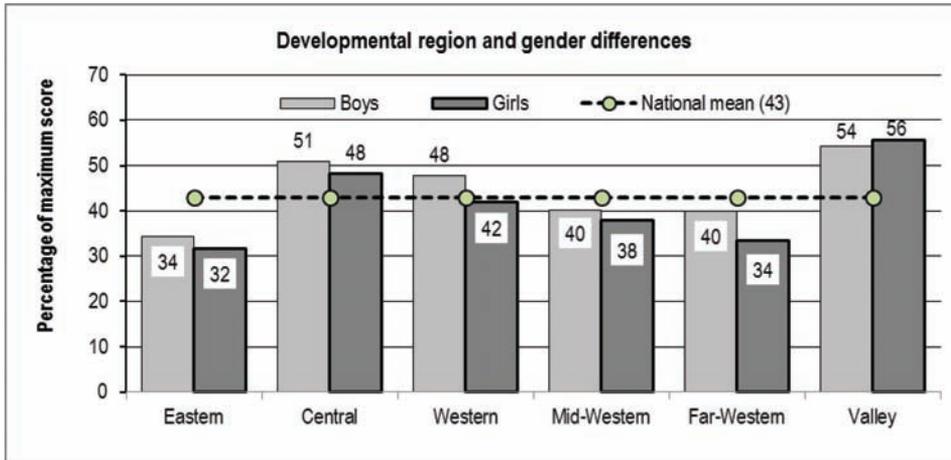
Figure 3.1.19 Developmental region and gender-wise differences

Dataset shows that the boys are outperforming girls in all content areas of Mathematics. Differences are not wide though they are significant. Sex explains only 0.8% of the student variation; from the equity point of view the signal is positive thought there is still work to be done to reduce the gap. Differences are the widest in Western and Far-Western regions; in Kathmandu Valley there is no difference between the genders.

3.1.3.3 Selected explanatory factors and achievement

The simplistic model in Section 2.3.1 represents several possible factors which may explain the differences in student achievement. Many of the factors have already been handled in Section 3.1.3.2: geographical factors, such as Districts, Ecological zone, and Developmental region, as well as School-related technical factors, such as school Type and school Location. Also some individual related factors were handled, such as Home language, Caste and Sex/Gender. In this section, several other factors are taken into consideration. The socioeconomic status (SES) of the students' families, paid work after school, students' attitude towards Mathematics as a school subject, age of the student, and help given to the

studies, are mainly family- and individual related factors. As a sample of deepening school- and teacher-related factors availability of schoolbooks, homework given by the teacher, and selected activities in the school are handled. Many other factors could be selected; the background questionnaire is rich.

3.1.3.3.1 Socioeconomic status (SES) and student achievement

There were many variables indicating the socioeconomic status. In NASA 2011 they were categorized into parents' education, parents' occupation, home possessions (whether or not the student has his own space to do homework, or a dictionary, for example), home accessories (how many mobile phones, cars or bathrooms there is 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 2.4.2). In this section, the education of the parents is further elaborated on, so that the illiteracy of the parents is analyzed in relation to the Mathematics achievement.

Several SES-related variables were analyzed by using a data mining tool of SPSS, Decision tree analysis (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. Some examples of this are handled with parents' educational background and its relation with students' achievement in Mathematics.

Parents' education

In NASA 2011 background questionnaire, parents' education is divided into nine categories: 1) illiterate, 2) literate, 3) grades 1–5 pass, 4) grades 6–10 pass, 5) SLC pass, 6) IA pass, 7) BA pass, 8) MA pass, and 9) Above MA pass. The question was asked from the students and hence there may be some impurities embedded in the data. However, with the huge dataset the results seem credible.

DTA classifies mother's education into five groups with statistically significant differences in students' achievement levels: illiterate mothers (students' average is 40% of the maximum score), just literate mothers (42), grade 1–5 passed mothers (44), grade 6–10 passed mothers (47), and SLC passed (or higher) mothers (50) (Figure 3.1.20, of the explanation of the elements in the figure, see Section 2.7.1). In each group, the number of mothers is high enough to make a credible prediction. The difference between each group is statistically significant ($p < 0.001$). In practical words, the results mean, for example, that when the mother has passed the SLC she can give 10 percent point advance in the national test compared with the illiterate mothers.
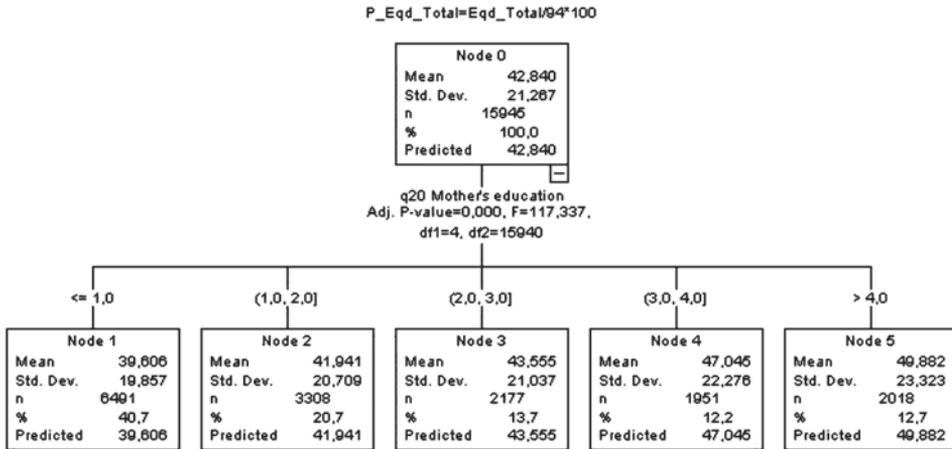
87

Figure 3.1.20 DTA of mother's education and students' achievement in Mathematics

In parallel, the DTA divides the father's education into six categories: illiterate or just literate fathers (39), grade 1–5 passed fathers (40), grade 6–10 passed fathers (44), and SLC passed fathers (48), IA passed fathers (53), and higher than IA-passed fathers (48) (Figure 3.1.21). Compared with the mother's education, note that, if the mother is literate, it raises the standard of the children more (42) than if the father is literate (39). The same can be seen in SLC-passed mothers (50) and fathers (48). However, when the father has passed AI, the results are remarkably higher (53).
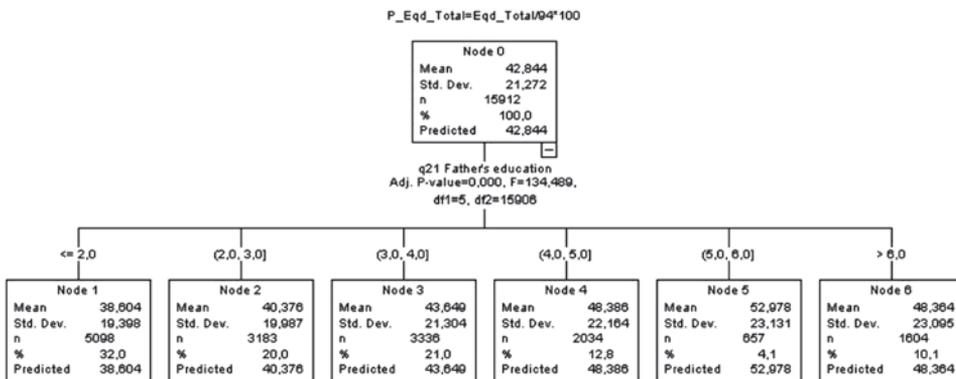
Figure 3.1.21 DTA of father's education and students' achievement in Mathematics

When combining the mother's and father's education, the lowest prediction for the children's future achievement in Mathematics comes when the father is either illiterate or just literate (39) or the father has passed grades 1–5- but the mother is illiterate (39). In the other extreme, the best results come when the father is IA-passed and the mother has passed grades 6–10- or higher (55), or the father is higher than IA passed and the mother has passed grades 6–10- (53), or both the father and mother have passed the SLC (53). It seems 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, the cut-off for parental education was set to "SLC-passed", that is, when being SLC-passed (or higher), the indicator for mother's (and father's) education for SES was set to 1, and the lower education than SLC-passed gave the value 0.

Dataset gives a strong signal that the educational level of the parents predicts the children's future achievement level in Mathematics. Especially harmful for the achievement level seems to be the situation that both parents are illiterate or the father is just literate but has not passed the lowest level of education (grades 1–5).

Parent's occupation

The occupation of parents was categorized into five categories: 1) agriculture, 2) teaching, 3) services, 4) business, and 5) others. In a similar manner as with the parents' education, the DTA was used to find the statistically the most deviating groups related to student achievement (Fig. 3.1.22). The student achievement is the lowest when the mother's occupational background comes from agriculture (40). It is statistically significantly higher when the mother comes from "other services than teaching" (46), "teaching" (51) or "business and other occupations" (53). It seems that either economic- or intellectual capacity or both at home helps the children to raise their standards.
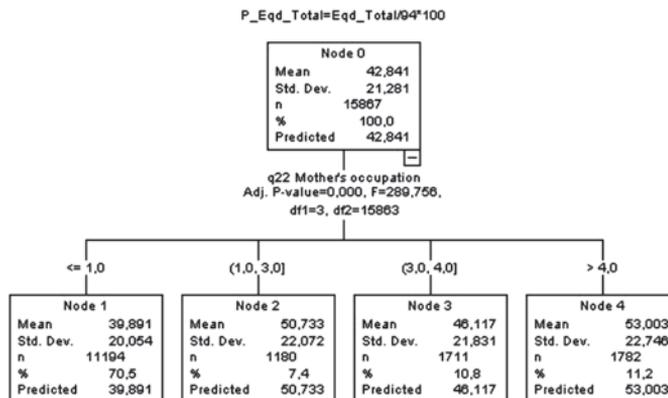

Figure 3.1.22 DTA of mother's occupation and students' achievement in Mathematics

89

When it comes to father's occupation, the main division is whether the father works in agriculture (40) or not (Fig. 3.1.23). If the father was a teacher, the students would gain 10 percent points more in the test (50) and if the father is either in other services, business or any other occupation the results are not that good (45) though better than of those coming from an agricultural background.
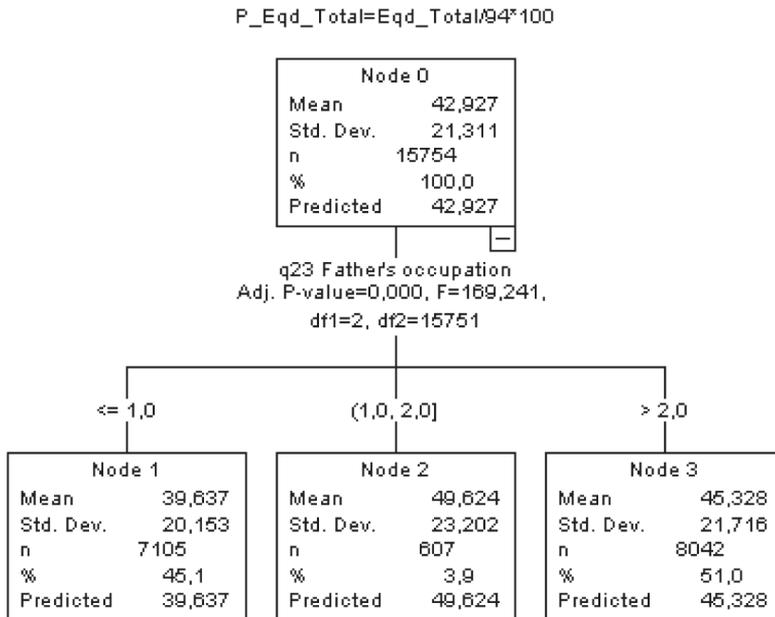
P_Eqd_Total=Eqd_Total/94*100

```
                    ┌─────────────────────────┐
                    │         Node 0          │
                    │ Mean           42,927   │
                    │ Std. Dev.      21,311   │
                    │ n              15754    │
                    │ %              100,0    │
                    │ Predicted      42,927   │
                    └─────────────────────────┘
                        q23 Father's occupation
                     Adj. P-value=0,000, F=169,241,
                           df1=2, df2=15751
```

<= 1,0	(1,0, 2,0]	> 2,0
Node 1	**Node 2**	**Node 3**
Mean 39,637	Mean 49,624	Mean 45,328
Std. Dev. 20,153	Std. Dev. 23,202	Std. Dev. 21,716
n 7105	n 607	n 8042
% 45,1	% 3,9	% 51,0
Predicted 39,637	Predicted 49,624	Predicted 45,328

Figure 3.1.23 DTA of father's occupation and students' achievement in Mathematics

When combining the mother's and father's occupation the lowest achievement is found in the families where both parents come from an agricultural background (39), or the mother works in agriculture and the father is in business (39) – here the business may refer to selling the agricultural goods in markets, or the father is in agriculture but the mother is in other service than teaching (39) – where, again, the students may have referred selling and serving the farming goods as the "services". Still another, somewhat low-performing group is the students from the families where the mother is in agriculture and the father is in "service" (41).

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

The dataset gives a strong signal that if the father or mother or both are coming from an agricultural or related occupation, the students' achievement in Mathematics is significantly lower than with the other occupational groups.

Parents illiteracy and agriculture

An additional question related to parents' education and occupation is whether the illiteracy of either parent is related to the agricultural occupation. On the basis of the previous analyses, the student performance is the lowest in the illiterate homes and in agriculture. When focusing only on the illiterate parents, the DTA shows that there are two pockets of very harmful combinations for the children's mathematical achievement. Generally speaking the results are better if the parents do not come from an agricultural background. However, *if both parents come from somewhere else than the agricultural background but both are illiterate*, the student's achievement is very low (36). Similarly low achievement is also in the group where *the mother comes from an agricultural background (and father does not) and both parents, or the father alone, is illiterate* (36). Of the analysis above, it is known that if both parents came from an agricultural background, the students' mathematical achievement is very poor regardless of the literacy (39). In the other extreme, when both parents come from a non-agricultural background and both are literate, the learning outcomes are very high (53).

Home possessions and accessories

Facilities and resources available at home may 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 the 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, poetry books, or artistic things like pictures, and books that help them for study, such as a 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, computers, cars, and bathrooms.

There are 12 questions in the student background questionnaire related to home possessions. Each would be scored 1 if the student had access to this possession (e.g. having a separate room or a table to study). Adding these items up, the maximum score was 12. Figure 3.1.24 shows the connection of home

possessions and achievement level: the achievement level of the students' raises logically the more there is access to home possessions. Pearson product moment correlation coefficient between the achievement level and the factor ($r = 0.20$) is statistically significant ($p < 0.001$) though it is not very high. Effect size is moderate – not high ($d = 0.40$).

For the later use in SES, the cut-off for the factors was set on 5 possessions: if there were 5 items or more mentioned in the background questionnaire, the student was rated 1, otherwise 0.
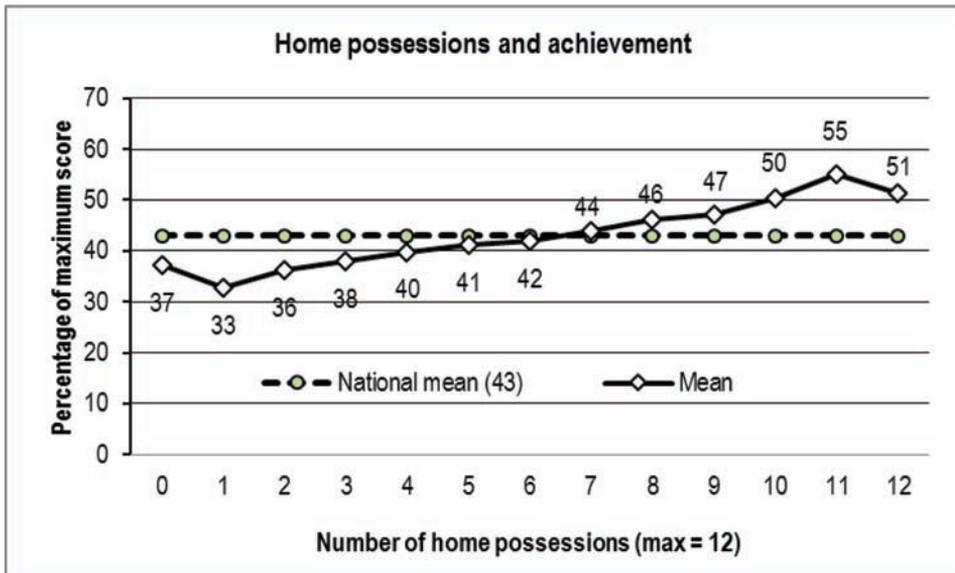
Figure 3.1.24 Connection of home possessions and achievement in Mathematics

The same pattern – the more possession, the better results – can be seen also with home accessories, as seen in Figure 3.1.25. 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). After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level, see Table 3.1.21), all the five indicators were summed. The maximum score was 5 indicating that the student's home possessed a set number of all the accessories.

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
Car	1–3	0, missing
Bathroom	2, 3	0.1, missing
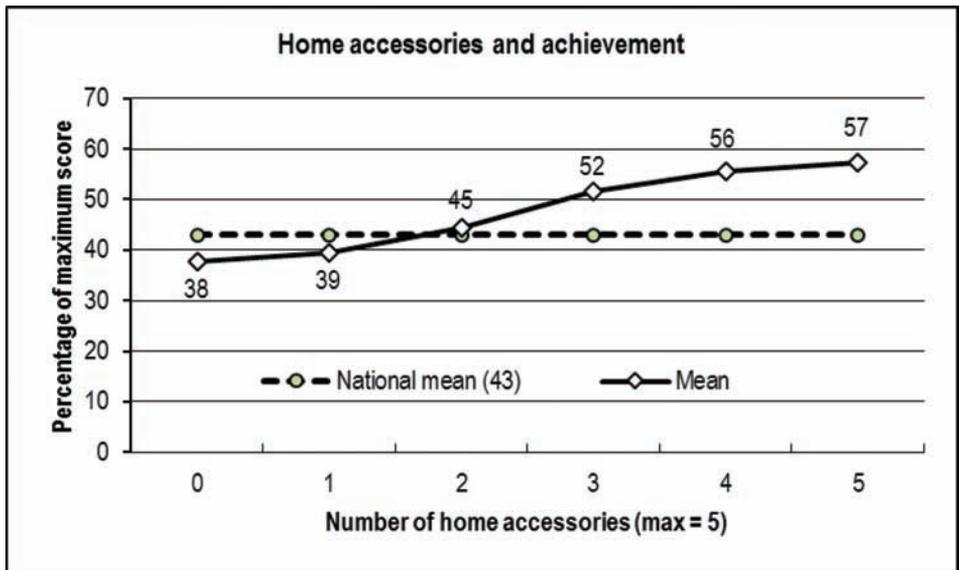
Figure 3.1.25 Connection of number of home accessories and achievement in Mathematics

Figure 3.1.25 clarifies how the number of home possession or accessories increases, students' achievement increases from 38 (if none of them are available) to 57 (if all of them are available). Availability of all the stated facilities indicates the high SES of the family. Correlation between home accessories and achievement is $r = 0.25$ ($p < 0.001$) certainly positive though not very high. The effect size of $d = 0.50$ indicates a medium size of difference between the groups. The difference between the lowest group with no mobiles, television, computer, cars, or bathrooms (38) and the highest group with all these accessories (57) is remarkable though; the parents with better economic possibilities can offer their children almost 20 percent points more marks than those without the economic possibilities. Note that 59% of those students, who have the access to all five accessories, live in Kathmandu Valley.

Data shows that when the children do have very few home possessions – 0 to 3 out of the 12 – the achievement level is statistically lower (37) than if there are more than four (> 40). With ten to twelve possessions the average score is very high (> 50) compared with the national average. The same pattern shows with home accessories: With no or only one accessory indicator met, the results are very poor (38–39) and when there are two or more met, the results are remarkably higher (> 45).With four or five indicators met the results are the best (56–57).

SES and achievement

The socioeconomic status was formed on the basis of seven indicators which were all first dichotomized (see the opening section of the Section 3.1.3.3.1). 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 (as SES) and changed into a percentage of the maximum score (P_SES). The deeper description of the transformations is seen in Section 2.7.2. The P_SES represents the percentage of SES the student possesses; 100 means that the student has the highest SES possible measured with these variables and with these transformations (that is, all the seven indicators of SES 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 3.1.26 presents the relationship between SES of the students and the achievement.
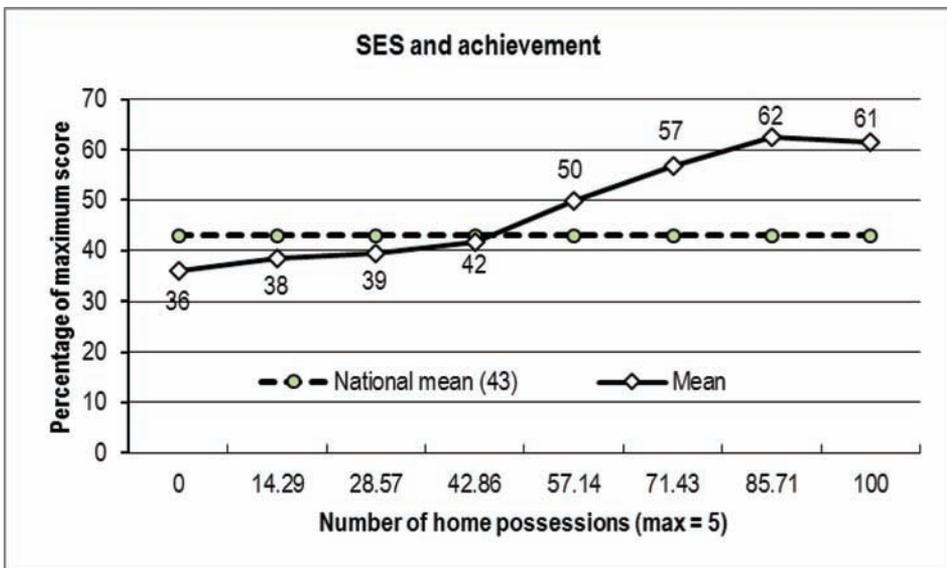
Figure 3.1.26 Connection of SES and achievement in Mathematics

Figure shows a strict positive relationship between SES and the achievement; the correlation between the variables is $r = 0.30$ which is a significant association ($p < 0.001$). The differences between the SES groups are statistically significant (ANOVA, $p < 0.001$), the effect size is moderate or high ($f = 0.34$); that is, the highest and lowest group differ from each other remarkably. SES explains somewhat 10% of the student variation ($\eta^2 = 0.102$) which is not, though, a very high percentage. The DTA suggests that three highest groups should be merged; there are not very many students in the highest SES groups. After combining three highest SES groups, the mean of the students in that merged group is 58.

The dataset gives a strong signal that the socioeconomic status plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (22–25 percent points). This means that if the social economic standard of the lowest performing students was raised into a decent level, that is, in practice, that the problems in four out of seven indicators would be solved, the results in these groups would raise remarkably. Especially challenging is the situation in the families where both parents are illiterate or they both work in agriculture.

3.1.3.3.2 Working after school and achievement

Several questions were set in the student background questionnaire of the students' activities outside the school. Two of them are briefly handled here: 'Working after the school in a paid capacity' and 'Participating in household work/chores'. The values of the variables are divided into five categories: 0 (no paid work 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). DTA shows that when the children are not engaged in paid work *at all*, the results are notably above the national average (46) (Fig. 3.1.27). *If the students were working – even less than one hour, the results were remarkably lower (35–37) than the average*. The differences are statistically significant ($p < 0.001$) though the effect size moderate – most of the children do not need to work in a paid capacity. Working after school indicates that the family is poor and the extra salaries are needed and when the student needs to work more than 4 hours per day there is no time or energy to handle school homework.
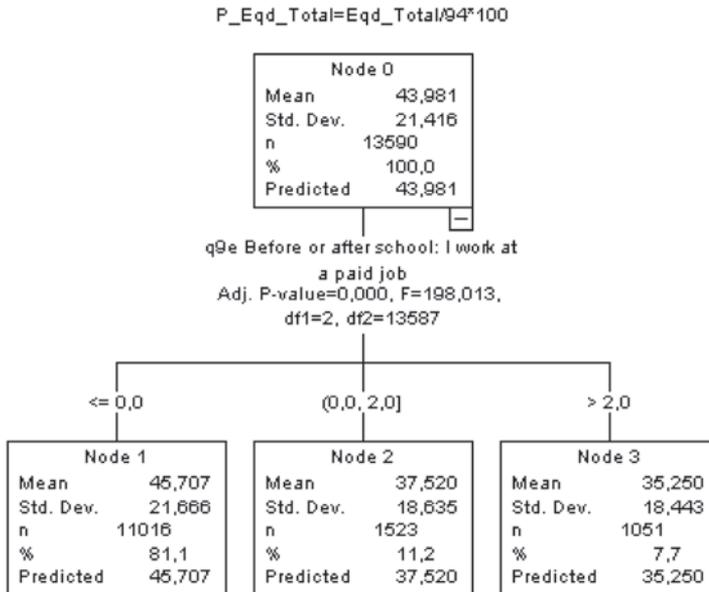
P_Eqd_Total=Eqd_Total/94*100

Node 0	
Mean	43,981
Std. Dev.	21,416
n	13590
%	100,0
Predicted	43,981

q9e Before or after school: I work at
a paid job
Adj. P-value=0,000, F=198,013,
df1=2, df2=13587

<= 0,0   (0,0, 2,0]   > 2,0

Node 1	
Mean	45,707
Std. Dev.	21,666
n	11016
%	81,1
Predicted	45,707

Node 2	
Mean	37,520
Std. Dev.	18,635
n	1523
%	11,2
Predicted	37,520

Node 3	
Mean	35,250
Std. Dev.	18,443
n	1051
%	7,7
Predicted	35,250

Figure 3.1.27 DTA of paid work and achievement in Mathematics

It is usual – and supported – practice in families that the children take part in the household work at home; this is part of the socializing process of the children. The DTA shows that when the child needs to spend more than 2 hours per day doing household work, the results are as poor (36) as if the student needs to work in a paid capacity (Fig. 3.1.28). However, when the amount of household work is decent – practically less than one hour per day – the achievement level is higher than the average (45–46). Even 1–2 hours household work reduces the achievement level to 41. Differences are significant ($p < 0.001$) though the effect size is small or moderate – few children are participating in household work for more than 2 hours per day.

P_Eqd_Total=Eqd_Total/94*100

```
                         ┌─────────────────────────┐
                         │        Node 0           │
                         │ Mean            43,352   │
                         │ Std. Dev.       21,393   │
                         │ n              14764     │
                         │ %              100,0     │
                         │ Predicted       43,352   │
                         └─────────────────────────┘
```

q9d Before or after school: I do jobs
at home
Adj. P-value=0,000, F=113,951,
df1=3, df2=14760

<= 0,0   (0,0, 1,0]   (1,0, 2,0]   > 2,0

Node 1		Node 2		Node 3		Node 4	
Mean	45,833	Mean	44,586	Mean	41,245	Mean	35,964
Std. Dev.	22,355	Std. Dev.	21,193	Std. Dev.	20,190	Std. Dev.	18,643
n	5408	n	4785	n	2735	n	1836
%	36,6	%	32,4	%	18,5	%	12,4
Predicted	45,833	Predicted	44,586	Predicted	41,245	Predicted	35,964

Figure 3.1.28 DTA of household work and achievement in Mathematics

The dataset gives a strong signal that working outside of school reduces the school achievement of the student. The phenomenon is most probably connected with the poor economic situation in the family. Especially when the children need to work more than 2 hours per day either paid or unpaid, the achievement level is remarkably lower.

3.1.3.3.3 Attitude and achievement

In the context of the Mathematics achievement assessment, the attitude tells us what the students think about Mathematics and its usefulness in their daily life and future utility. 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 Mathematics achievement and attitudes toward Mathematics 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 attitudes and achievement may be near zero, like in Macedonia (Kadijevich, 2008), the Philippines (Wilkins, 2004), 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, the same shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) as is in use in several international comparisons, like in TIMSS and PISA studies, was in use. The original scales included nine dimensions but in these international comparisons only three is used with four items on each and two negative items on each of the first two dimensions (see in detail in Section 2.4.2 and 4.5). The names of the factors can be "Liking Math", "Self-Efficacy in Math", and "Experiencing utility in Math" (compare naming in, e.g., Kadijevich, 2006; 2008). Factor analysis was used to identify the factors of the responses in FSAS and the negative items were reversed to make the whole test unidirectional. As in several countries in Asia, the expected factor structure cannot be found in Nepal (for a deconstruction of the test scales, see Metsämuuronen, 2012a; 2012b). Hence, only the total score is used to show the connection of attitudes and achievement. The relation between the attitude (divided into deciles of somewhat the equal number of the students) and achievement score is shown in Figure 3.1.29.
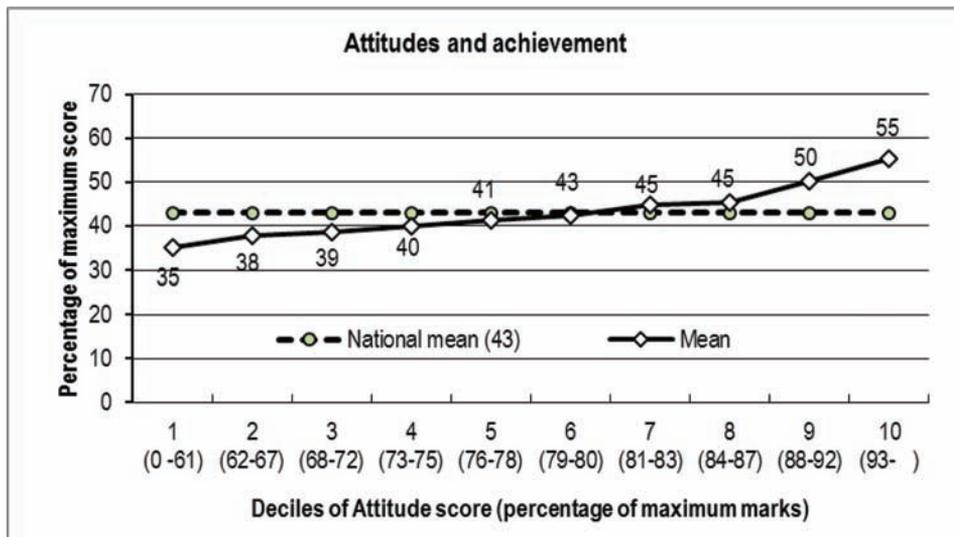
Figure 3.1.29 Connection of attitudes and achievement in Mathematics

Figure 3.1.29 shows that the difference between the lowest attitude group (35) and highest attitude group (55) is notable. However, the real difference in the increase in achievement occurs in the two highest attitude groups (50 and 55).

The correlation between the positive attitude towards Mathematics and achievement is $r = 0.25$ ($p < 0.001$); the effect size is moderate ($d = 0.50$).

A common challenge using attitudes to explain achievement levels is that one does not know which is the hen and which is the egg, that is, whether the attitudes are high because of high performance or the performance is high because of positive attitudes? By explaining the variables both ways with the univariate ANOVA, it is more probable that the achievement is caused by the attitude (attitude explains 9.5% of achievement) and thus it is less probable that the achievement caused the attitude (achievement explains 7.7% of attitude).

Data gives a strong signal that the more positive the attitude is towards Mathematics the higher is the achievement. The data also supports the fact that positive attitudes influence the positive achievement (and not the other way round).

3.1.3.3.4 Age and student achievement

In the Nepalese context, the age of the students attending to grade 8 studies varies remarkably. Some students have given their age as below 10 years and some above 20. All the ages of the students below 13 were encoded as 13, and all students above 19 as 19. The descriptive statistics of the mean in each year are given in Table 3.1.22 and visualized in Figure 3.1.30.

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

Age	N	Mean	SD
up to 13 years	3128	49	22.0
14 years	5108	47	21.8
15 years	4141	40	19.7
16 years	2132	36	18.6
17 years	966	32	16.7
18 years	337	33	19.0
19 years & above	182	35	19.2
Total	15994	43	21.3
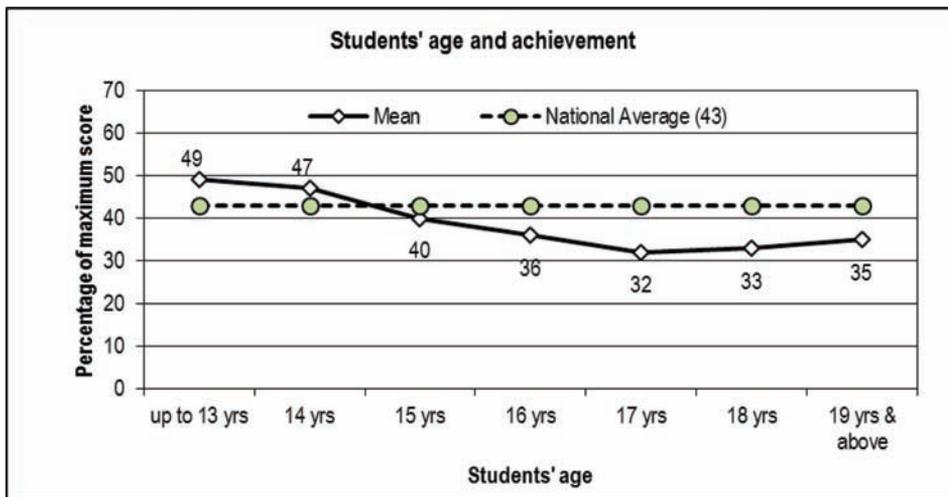
Figure 3.1.30 Connection of age and achievement in Mathematics

It seems evident that the best achievers are those students who are at the proper age for grade 8 studies (13 to 14 years old, scoring 49 and 47 respectively). The higher the age is – meaning that the students have either started much later than they should have, or they have doubled the classes – the weaker the results are. The achievement level is remarkable lower when the students are of age 16 or higher (32–36). Correlation between the variables is -0.17 ($p < 0.001$) indicating small or moderate effect size ($d = 0.35$). The ANOVA hints that the age (that is, the prolonged studies) explains the achievement level (7% of explaining) more probably than the achievement level the prolonged studies (4% of explaining). Another side of the matter is that it is good that these "over-aged" students are at school to learn; although they should have been identified at a much earlier age for extra tuition or support.

Dataset gives a signal that the highest performance is with those students studying with their normal age group, that is, at the age of 13 and 14 years. Otherwise the achievement decreases as the age increases.

3.1.3.3.5 Help in study and student achievement

The relation between the help in 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 helpers are given in Table 3.1.23.

Table 3.1.23 Descriptive statistics of helpers of the students

Who helps	N	Mean	SD
No one	312	42	20.5
Brother/Sister	2931	42	20.7
Father	814	42	22.7
Teacher	10448	43	21.3
Mother	227	45	22.0
Tuition	1019	48	20.9

It seems that help in Mathematics is necessary for the students to gain better than average marks on the test (compare Social Studies on Section 3.3.3.3.5). There is about 6 score difference between those who don't get any kind of help and those who receive (private) tuition. Mother's help at home raises the achievement level (45) over any other helper from the family members. It is, though, noteworthy that there were a few numbers of the students ($n = 227$) who reported the help of mothers. It is interesting to note that those who were helped by their fathers gained as much (42) as those who did not get any help (42). The extra (private) tuition brings the most advance for the student achievement in Mathematics (48). However, the advance is not at all at the same level as when comparing the illiterate and literate families, for example. It is also unknown whether the time spent on the tuition has something to do with the matter?

The dataset gives a slight signal that tuition given by the mother raises the achievement level more than any other family members' help. The highest achieving groups are those who receive private tuition. It is possible that this group also spent more time on their homework which may explain the high score.

3.1.3.3.6 Availability of textbook and student achievement

The data shows that there are still some students who don't have a Mathematics text book up to the last academic session. Table 3.1.24 shows the descriptive statistics of the viability of the textbook of Mathematics and the achievement (mean).

Table 3.1.24 Availability of textbook of Mathematics and the achievement (mean)

Do you have textbook of Mathematics?	N	Mean	SD
Yes	15317	43	21.3
No	371	36	17.8
Total	15688	43	21.3

Out of 15,688 who responded to the question, 2.5% of the students don't have a textbook available at school. The relation between the textbook and achievement is significant ($p < 0.001$) though the effect size is small due to the small group size in one group. The difference in achievement is anyhow quite big (8% of maximum marks).

Data shows that 2.5% of the students lack the proper text book in Mathematics. The achievement level of these students is significantly lower than those who have access to the text book.

3.1.3.3.7 Homework given/checked and achievement

Homework is one of the methods of enhancing teaching; it can be used as drill-, exercise-, and as an evaluation tool. When homework is systematically checked, it may boost achievement levels. Now, the results are based on students' reports; it may be possible that within the same classroom some students have given a slightly deviant response than the other students from the same class. However, in the wide scope, the results are credible. Statistics related to homework given and checked is condensed in Table 3.1.25 and visualized in Figure 3.1.31.

Table 3.1.25 Homework given and the achievement

Status of homework	N	Mean	SD
Given Some day - not checked	174	35	19.1
Not given	252	36	19.5
Given Every day - not checked	310	37	19.5
Given Some day - Checked some day	1510	41	20.5
Given Some day - Checked every day	1375	42	21.3
Given Every day - Checked some day	1964	44	21.5
Given Every day - Checked every day	10232	44	21.4
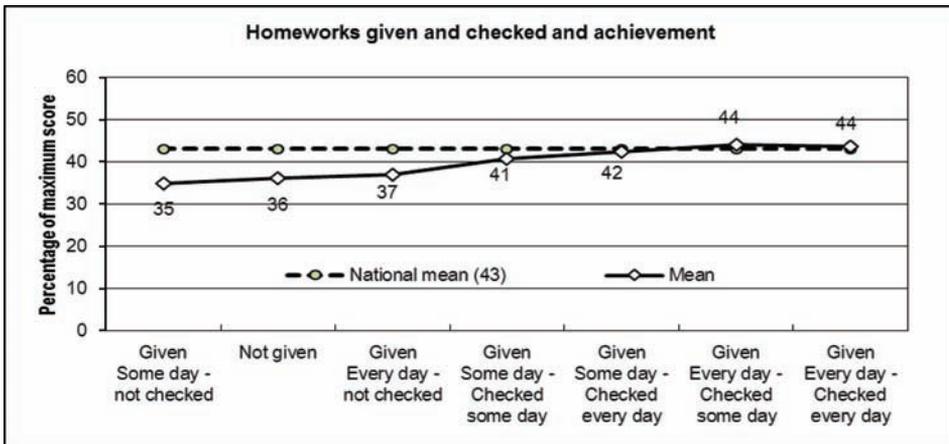
Figure 3.1.31 Connection of homework given and checked and achievement in Mathematics

102

It is evident on the basis of the dataset that if the teachers give homework every day to the students and it is checked – even only now and then, then student achievement is higher (44) than when homework is issued non-systematically (41–42) or homework is not given or they are not checked at all (35–37). The differences are statistically significant ($p < 0.001$). Those groups with no checking or homework given are, however, very small and hence, the effect size is small ($\eta^2 = 0.007$).

   Dataset gives a solid signal that if the teacher gives the homework daily and checks it, the achievement is higher than without checking or issuing of homework. By giving homework daily and by checking it, even not every day, the teacher may have raised the scores up to 9 percent point.

3.1.3.3.8 Activities in the school and student achievement

The activities of the students and 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 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' impression 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 learning environment for the students. International Studies like TIMSS and PISA give a specific emphasis to identify such indicators. In the NASA 2011 student questionnaire, five questions indicate the varieties of bullying that may 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 student response is condensed in Table 3.1.26 and visualized in Figure 3.1.32. 'No' (%) indicates the percentage of the students' response that describes there happened no such activity in the school and 'Yes' (%) indicates the percentage of the students who reported the particular type of bullying happened within last month. Alone, the fact that 22% of the students mention that, during the last month, something of their own was stolen is an alarming sign of the system.

Table 3.1.26 Bullying and the achievement

Type of Bullying	No (%)	Yes (%)
Something of mine was stolen	75.1	22.0
I was made fun of or called names	80.1	14.6
I was hit or hurt by other student(s)	81.1	14.2
I was made to do things I didn't want to do by other students	84.9	10.5
Fellow students kept outside without involving me in activities	86.2	9.1
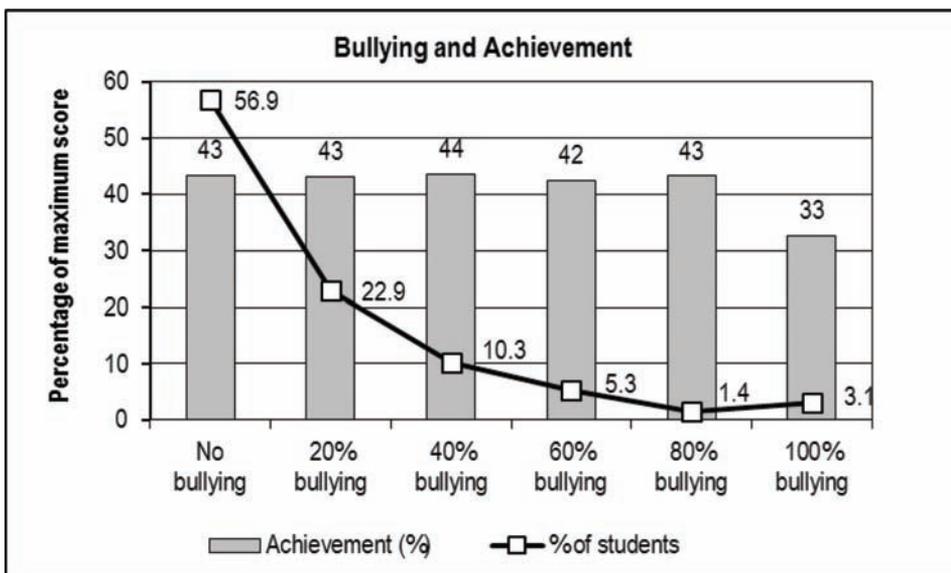
Figure 3.1.32 Relation of bullying and achievement in Mathematics

The sum of all five items is done as an indicator of bullying. Figure 3.1.32 shows the extent of bullying with the percentage of the students and achievement of the students in each category of bullying is shown in the line graph as bars. 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 56.9% of the students did not encounterany bullying during the last month, one can infer that the remaining 43.1% have encountered at least one type of bullying. This is a remarkable number of the students. About 2.5% of the students are experiencing a severe kind of bullying (the sum of 80% bullying and 100% bullying). It seems, however, that learning outcomes are remarkably lower only with those 0.8% of the students who have encountered extreme bullying including all five types of

harassments (36). Students who don't feel bullying and students who encountered extreme bullying of all five kinds have almost 7% achievement gap although a very small number of the students reported all kinds of bullying (n = 129). However, the difference is statistically significant ($p = 0.001$) though the effect size is small ($f = 0.03$). Though the extreme case of severe bullying is 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 the schools.

The dataset gives a signal that an alarmingly high number of the students (46%) have encountered bullying in schools within the last month. Though the phenomenon does not have a great affect except in the group of extremely bullied students, all possible efforts has to be done to root the phenomenon from the schools.

Positive activities in school

The activities that can boost the learning and achievement of the students are categorized as positive activities. Such positive activities about the school were asked from the students in two sets of questions collected in Table 3.1.27. The table shows the responses of the students in all four categories; the responses are in the 4-point rating scale anchored to fully agree and fully disagree.

Table 3.1.27 Students' response towards teacher- and school-related activities in the schools

Teacher and Students activities	Respondents in %			
	Fully agree	Partially agree	Partially disagree	Fully disagree
Students get along well with most teachers	75.3	21.3	2.2	1.2
Most teachers are interested in student's well-being	83.8	12.1	2.3	1.8
Most of the teachers really listen to what I have to say	53.9	36.1	7.3	2.6
If I need extra help. I will receive it from my teacher	76.4	19	2.8	1.7
Most of my teachers treat me fairly	55.9	24.6	8.6	10.8
I like come and stay in school	90.3	7.2	1.3	1.3
Students in my school try to do their best	73.2	22.9	2.7	1.3
Teacher in the school care about the students	78.4	17.4	2.8	1.4
Teacher wants the students to do their best	89.4	6.9	1.7	1.9
Average	75.2	18.6	3.5	2.7

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 with achievement. The overall result is that the feeling of the positive actions in the school relates positively with the student achievement. The correlation between the sum of nine positive activities and achievement is positive ($r = 0.11$, $p < 0.001$).

DTA finds 6 attitude groups in the indicator. These boundaries and descriptive statistics are seen in Table 3.1.28 and illustrated in Figure 3.1.33.

Table 3.1.28 Students' response towards teacher- and school-related activities in the schools

% or total score	N	Achievement	SD
0 - 73,3	1384	38	19.8
73.3 - 81.5	1258	39	20.2
81.5 - 85.2	1139	41	20.8
85.2 - 87.5	1632	42	20.4
87.5 - 96.3	4270	43	20.9
96.3 - 100	6251	46	22.0
Total	15934	43	21.3
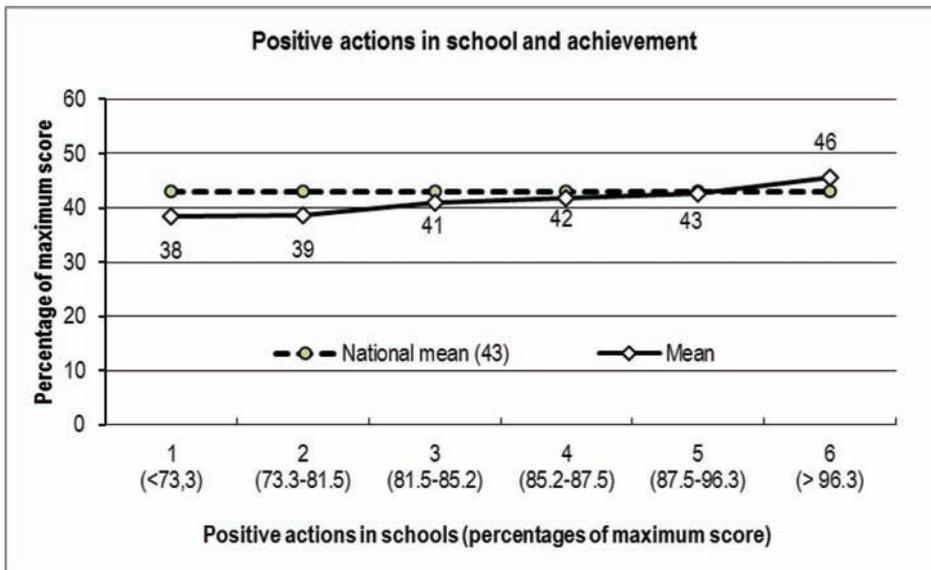
Figure 3.1.33 Relation of positive actions in school and achievement in Mathematics

The data shows that there is a positive connection between the student's feeling of the school activities and achievement. The increase in achievement is directly proportional to the increase in intensity of such activities. After dividing the indicator into six groups on the basis of DTA, the differences between the groups are statistically significant ($p < 0.001$), however, the effect size is small ($f = 0.12$). Only when the students are extremely positive towards school and teachers' behavior, is the learning achievement much higher than the average (46). However, the difference between the most positive group and the most negative group is notable (8%).

Dataset gives a signal that when students feeling that the actions of the teachers and the schools are ultimately good, the results are better than average (46). At the other extreme of feeling that such actions are negative, the results are far below the average (38).

3.1.4 Conclusions

The main findings of grade 8 Mathematics achievement are condensed as follows:

Basic results

- The students in the institutional schools perform very well in Mathematics but the students in the community schools form two kinds groups of schools: high-performing schools and low-performing schools. The variety between the community schools is remarkable.

- The learning outcomes are weaker in Geometry and Sets than in other content areas.

- The students' ability to solve complex problems in Mathematics is low; only 21% of the maximum scores were reached. Students are much better in the recalling type of questions (68%), which may be explained by the educational system which seems to be geared more to remembering the things than solving problems.

- In Mathematics, the students are good in recognizing the correct answer and in very fundamental operations, the basic manipulation of data and numbers, and in calculations with few steps. They are much weaker in reasoning, problem solving, plotting, proving the theory or formula, and constructing the shape and figures. In many cases, the students did not even start to do the open ended questions – hence, the low score.

- The mathematical skills have not been developed as is expected of the year 2008. In all content areas, the results are slightly lower in the year 2011.

- From the international comparison viewpoint, the average achievement of Mathematics in Nepal is better than the international average. In Algebra and Arithmetic the level seems to be higher and in Geometry, Sets, and Statistics it seems to be lower than the international mean.

Equity indicators

- There is a wide difference between the districts when it comes to the equal opportunities of children to reach the preset goals in Mathematics. The results are bound to the 25 districts selected randomly to the sample; even lower-performing districts could be found if other districts would have been selected. In any case, the results in Bhojpur (18), Pyuthan (29), Illam (30), Sankhuwasabha (32), Kanchanpur (32), Morang (34), and Jajarkot (35) are so low that raising the standard in these districts would raise the standard in the whole country.

- There is a moderate difference between the student performances in four Ecological zones. Students in the Kathmandu Valley outperform the other students. The achievement is the lowest in the Hill area.

- There is waste inequity in the Developmental regions of children's opportunities to reach an adequate level of Mathematics. The difference between the lowest performing area (Eastern region, 33) and the highest performing region (Kathmandu Valley, 55) is remarkable – more than 20 percent points.

- The students in the institutional schools outperform the students in the community schools. Most probably this deviance can be explained by the student selection, however, this cannot be known on the basis of the dataset.

- Though the results are somewhat better in cities than in the rural area, the difference is not remarkably high. From the equity point, this is a positive sign.

- The small groups of Sherpa and Newari speakers perform very well in Mathematics (54 and 52 respectively). However, such equally small groups as Rai-, Limbu-, and Magar speakers perform very low (20, 28, and 37 respectively). The dataset arises two difficult questions: What has been done differently – or not done at all – in Eastern Hill areas

where Rai-, Limbu-, Tharu-, Tamang-, and even Sherpa students get very low scores? Why are their results in Hill areas poorer than in other regions?

- Dalit students' performance is lower than the results of the other castes. In particular, their performance is low in the Eastern Developmental region. A positive signal is that Dalits' performance is very high in the Mid-Western Mountain- (58) and Central Tarai areas (53).

- The boys are out-performing girls in all content areas of Mathematics. Differences are not wide though they are significant. Sex explains only 0.8% of the student variation; from the equity point of view the signal is positive though there is still work to do to reduce the gap. Differences are the widest in Western and Far-Western regions; in Kathmandu Valley there is no difference between the genders.

Selected explanatory factors

- The educational level of the parents highly predicts the children's future achievement level in Mathematics. Especially harmful for the achievement level seems to be the situation that both parents are illiterate or the father is just literate but has not passed the lowest level of education (grades 1–5).

- If the father or mother or both are coming from an agricultural or related occupation, the students' achievement in Mathematics is significantly lower than with the other occupational groups.

- When the children have only a few home possessions (like their own table or a room or a dictionary) – 0 to 3 out of the 12 – the achievement level is statistically lower (37) than if there are more than four (> 40). With ten to twelve possessions, the average score is very high (> 50) compared with the national average. The result is the same with home accessories (like mobiles, cars, bathrooms): with no or only one accessory indicator met, the results are very poor (38–39) and when there are two or more met, the results are remarkably higher (> 45).With four or five indicators met the results are the best (56–57).

- As a whole, the socioeconomic status plays a strong role in the educational processes in Nepal. The difference in mathematical achievement between the lowest and highest SES groups is remarkable (22–25 percent points). This means that if the social economic standard of the lowest performing students was raised to a decent level, that is, in practice, that the problems in four out of seven indicators would be solved, the Mathematics results in these groups would raise remarkably. Especially challenging is the

situation in the families where both parents are illiterate or they both work in the agricultural field.

- Working after or out of school effectively drops the Mathematics achievement of the student. The phenomenon is most probably connected with the poor economic situation in the family. Especially when the children need to work more than 2 hours per day either paid professionally or unpaid in the household work, the achievement level is remarkably lower.

- The more positive the attitude is toward Mathematics the higher is the achievement. Data given also supports the fact that the positive attitudes influence the positive achievement (and not the other way round).

- The highest performance is with those students studying within their normal age group, that is, at the age of 13 and 14 years. Otherwise, the achievement decreases as the age increases.

- Tuition given by the mother raises the achievement level in Mathematics more than any other family member's help. The highest achieving group are those who receive private tuition. It is possible, however, that those with private tuition also spent more time on their homework, which may explain the high score.

- 2.5% of the students lack the proper text book in Mathematics. The achievement level of these students is significantly lower (36) than those who have access to the text book (43).

- If the teacher gives the homework daily and checks it, the achievement is higher than if this does not occur. By giving daily homework and by checking it, even not every day, the teacher may have raised the scores up to 9% per unit.

- An alarmingly high number of the students (46%) have encountered bullying in schools within the last month. Though the phenomenon does not much affect the result except in the group of extremely bullied students, all possible effort has to be done to root the phenomenon from the schools.

- When the students think that the teachers and schools actions are ultimately good, the results are better than average (46). At the other extreme if the feeling is ultimately negatively, the results are far below the average (38).

References for Section 3.1

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

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

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

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

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

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

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

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

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

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

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

Fullbright (2008). *A final report on national assessment of grade-V students*, Fullbright Consultancy & Cooperative Hands in restoration (CHIRAG). Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education.

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.

111

Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on Mathematics attitude. *The Teaching of Mathematics*, **9**(2), 41−51. Retrieved from http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf.

Kadijevich, Dj. (2008). TIMSS 2003: Relating Dimensions of Mathematics Attitude to Mathematics Achievement. Retrieved from http://www.ipisr.org.rs/Dokumenta/zbornici%20PDF/Zbornik%202008%2040-2/Zbornik-08-2-006.pdf.

Metfessel, N., Michael, W. B., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the Writing of behavioral Objectives. *Psychology in the Schools*, **6**, 227–231

Metsämuuronen, J. (2012b). Comparison of Mental Structures of Eighth-Graders in Different Countries on the basis of Fennema-Sherman test**.** *International Journal of Psychological Studies,* 4(4) December.

Metsämuuronen, J. (2012a). Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies,* 4(3) September.

Papanastasiou, C. (2000). Effects of Attitudes and Beliefs on Mathematics Achievement. *Studies in Educational Evaluation,* **26**(1), 27−42.

Papanastasiou, E. (2002). Factors That Differentiate Mathematics Students in Cyprus, Hong Kong, and the USA. *Educational Research and Evaluation: An International Journal on Theory and Practice,* **8**(1), 129−146.

PEDP (1998). Primary Education Development Project: An assessment. 1998. Kathmandu.

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.

TIMSS (2009). TIMSS 2007 User Guide for the international Database. Released Items, Mathematics – Eight grade. IEA, TIMSS & PIRLS International Study Center. Lynch School of Education, Boston College.

Wilkins, J. L. M. (2004). Mathematics and science Self-Efficacy: An international investigation. *The Journal of Experimental Education*, **72**(4), 331−346.

3.2 Nepali Achievement in NASA 2011

Mr. Devi Ram Acharya[1] &
Dr. Jari Metsämuuronen[2] &
Prof. Dr. Hemanga Adhikari[3]

1. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal
2. Helsinki University, Faculty of Behavioral Sciences, Finland
3. Tribhuwan University, Kathmandu, Nepal

Abstract

The Nepali proficiency of 8th graders was analyzed by utilizing student achievement tests and the related background questionnaires of 16,350 students from 418 randomly selected schools from 25 randomly selected districts. The schools represented all Ecological zones and Developmental regions, rural- and urban areas as well as community- and institutional schools.

The learning outcomes are the highest in the content area of Reading and weakest in Vocabulary. Students' ability to solve complex problems is quite low; students are much better in the recalling type of questions, which may be explained by the educational system, which seems to be geared more towards remembering the things rather than solving novel problems. The Nepali language skills have been developed in a positive way. The Nepali results rose 2 percent points within four years, especially, in Writing (+16) and Reading (+6) but it has dropped dramatically in Vocabulary (-22). From the international comparison viewpoint, the average reading proficiency in Nepal is much lower than the international average.

There is a wide difference between the districts when it comes to the equal opportunities of children to reach the preset goals in Nepali. There is also a moderate difference between the student performances in four Ecological zones, with students in the Kathmandu Valley out-performing the other regions. The achievement is the lowest in Tarai area. There is a waste difference between the lowest performing region (Far-western, 42) and the highest performing region (Kathmandu Valley, 64).

The students from Newar- (59) and Tamang (55) backgrounds performed higher than Nepali students (51). On the other hand, students from Limbu- (40), Tharu- (40), "other"- (41), and Rai (42) backgrounds performed much lower than the national average. The Madhesi students' performance is radically lower than the results of the other castes. A notable percentage of Madhesi students have not reached the required level of reading and writing Nepali in order to continue to study in higher education in the Nepali language. Differences between boys and girls are not wide though girls out-perform the boys.

The socioeconomic status (SES) plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (22 percent points). Especially challenging is the situation in the families where the father or both parents are illiterate or they both work in the agriculture sector. An alarmingly high number of the students (42%) have encountered bullying in schools within the last month.

3.2.1 Introduction

Nepali as a school subject – with Mathematics – is, and will be, the one subject that is assessed systematically and frequently in the national assessments of student achievement (NASA) in Nepal. Unfortunately, the results of the previous national assessments (see BPEP, 1997; EDSC, 1997; BPEP, 1998; PEDP, 1998; EDSC, 1999; CERID, 1999; EDSC, 2001; 2003; CERSOD, 2001; EDSC, 2008; Fulbright, 2008) are not comparable with each other because of the missing linking procedure between the tests. Of the previous datasets, only the year 2008 (EDSC, 2008) could be used as a basis for comparison because it is the one where an item-wise data was found. Hence, that dataset is used as a basis for the comparison of 2011 results with the previous one. Several items from the 2008 test were used as the linking items between the years 2008 and 2011. With these linking items it is possible to assess what kind of changes has happened between the two measurement points.

Another linking procedure was administered to link the results in Nepal to the Programme for International Student Assessment (PISA) data. Several published items from the international bank were borrowed in order to compare the results in Nepal with an international standard (see in detail Section 2.4 and 4.6). By using the Item Response Theory (IRT) modeling, the 2008 dataset, PISA datasets, and the 2011 NASA dataset were linked together to give unique new information on the achievement level in Nepali.

In what follows, Section 3.1.2 describes the specific methodological solutions in the Nepali testing – more detailed information on the process is found in Section 2. Section 3.1.3 reveals the basic results in the different content areas of Nepali in general (Section 3.1.3.1), this deepens to the effects of different equity matters from the diversity view point (Section 3.1.3.2) and gives some hints on the deeper factors explaining the differences in the Nepali achievement (Section 3.2.3.3). Finally, the results are condensed in Section 3.1.4.

3.2.2 Methodological solutions

In the process of testing the Nepali achievement of 8th graders, 16,350 students, 418 schools, 415 head teachers and 415 teachers from 25 districts participated in the survey and testing (for more details of procedures and sampling, see Chapter 2.1). Three test versions (in what follows, N1, N2, and N3) were administered. N1 and N2 were administered in twenty-three districts so that both versions were administered simultaneously in the same classroom. N3 was administered in two Himalayan districts, Mustang and Kalikot, two months earlier in a way that both other subjects (Mathematics and Social Studies) were also administered simultaneously in the same classroom. Hence, more schools were encountered from the Himalaya region though the number of the students was proportionally fewer than in the Hill and Tarai regions.

3.2.2.1 Item analysis and characteristics of the test versions

NASA 2011 is the methodological shift in the National Assessment of Student Achievement. The modern test theory, that is, IRT modeling was used from the beginning of the item construction and preparing the marking scheme for the analysis of the data. In practice, this means, first, certain restrictions on the marking of the papers: no decimal number scores are allowed in the assessment during the answer sheet marking. If the students are not qualified to secure full score, 0.5 score is not provided in any case i.e., the students' responses are marked in whole numbers. Obviously, the marking scheme in 2008 differed from this logic. Hence the 2008 data was adjusted to meet the requirement of IRT modeling: half points were lowered to the *lower* full mark because the students did not show enough achievement to gain the upper full mark. Secondly, the marking scheme has to be more rigorously prepared because of the need to make the exact the same judgments in the years to come with the linking items. Thirdly, IRT modeling requires that all the possible marks have to be observed in the dataset; this made some difficulties when analyzing some of the productive items because of the markers' tendency *not* to give full marks to the students. This was not a severe challenge in Mathematics compared with the Nepali subject (see Section 3.1). Finally, IRT modeling requires a linking procedure between the different versions in the test. Hence, the common items for each test version, the linking items, were carefully selected from the 2008 test booklet, PISA dataset, as well as in the pretest items banked for the item selection. Classical item- and test analysis methods were used in the pretest phase for finding the percentage of correct answers, item discrimination power; IRT was used for item calibration, finding the latent ability

115

(Theta, *θ*) as well as comparing and equating the versions N1–N3, the dataset from the year 2008, and PISA database. SPSS software was used for the classical analysis and One Parametric Logistic Model software (OPLM, Verhelst, Glass, Verstralen, 1995) was used for the IRT modeling. Table 3.2.1 shows the average marks (mean) calculated for three versions N1, N2 and N3 separately.

Table 3.2.1 Comparison of the characteristics of Nepali test versions

Version	N	Mean[1]	SD	Remarks
N1	8023	50	19.3	Administered in 23 districts, in same classroom
N2	8097	47	19.1	
N3	230	50	21.2	Administered in Mustang and Kalikot districts
Total	16350	49	19.3	Nepali as a whole (N1,N2 & N3)

1) percentage of the equated maximum score rounded in whole number

On the basis of the pretested items, the versions N1 and N2 were parallel whereas N3 was shorter and easier than the other two.[13] N3 was administered earlier than the other two; because of the willingness to avoid any leaking of the items. All the versions were linked with each other by the use of the identical linking items. The longer versions N1 and N2 were scored out of a maximum of 90 and 89 maximum marks and N3 has a maximum score of 87 maximum marks. There were 12 PISA items as the linking items all over three versions. The parameters of the international items were fixed during the item calibration so that all the test items of the year 2011 and 2008 were calibrated in the international PISA scale (see Section 2.4 for more details). After the calibration of the items, all the scores in the versions N1–N3 and the version of the year 2008 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 different difficulty levels of the tests have been modeled by using IRT modeling. The original output is the latent ability (Theta, θ) which is a standardized normal score ranging usually from -4 to +4. These values in each test versions were later transformed to equated scores and further, the equated scores were converted in to the percentage of maximum score so that the score 100 means that the student made a perfect score and 0 means that no items were successfully answered. Now onwards, marks or average or mean score refers to the percentage of the maximum marks ranging from 0 to 100.

---

13   Though the mean score in Table 3.2.1 shows that the average of N3 was the same as that of N1, the items actually were somewhat easier in N3.

Table 3.2.2 Construct validities and reliabilities of the scores

Topic	Marks			Percentages			Percentages in Curriculum	Reliability		
	VER1	VER2	VER3	VER1	VER2	VER3		V1	V2	V3
Total	66	65	59					0.90	0.89	0.90
Reading	16	15	14	22.2	20.8	21.2	10	0.72	0.69	0.66
Writing	37	37	34	51.4	51.4	51.5	65	0.81	0.81	0.84
Grammar	11	12	10	15.3	16.7	15.2	15	0.60	0.66	0.62
Vocabulary	8	8	7	11.1	11.1	10.6	10	0.55	0.52	0.64

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 items were classified into four categories: *reading, writing, grammar and vocabulary* (Table 3.2.2). The number of items asked and the weighting of the items is based on the recommendations of the curriculum for grade 8. Overall internal consistencies (given by reliability) of the whole tests on each version were very high ($\alpha = 0.898$), however, some of the categories (grammar and vocabulary) contains a few items and hence the reliability is somewhat 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 the validity viewpoint, it is notable that, in the Nepali curriculum, the weighting of reading items (10%) is much less than it is in the final test (21–22%); the weighting was changed because the reading comprehension is very appreciated in the developed countries and it was aimed to test this skill more comprehensively rather than just with a few questions.

3.2.2.2 Analytical tools used in the analysis

Basic statistical methods are used throughout the report. The basic tools of statistical description (means, Standard Deviations, percentages, and frequencies), correlations (Pearson's product moment correlation coefficient and Spearman's rank order correlation), and comparison of two means (t-test) as well as statistical inference (p-values, effect sizes) are used when appropriate. These methods are described in all basic books of statistical description and -inference. The Analysis of Variance (ANOVA) is used in the General Linear Modeling (GLM) way when

several means are compared. All the p-values are corrected by using Multilevel modeling (or Hierarchical Linear modeling) by using SPSS Linear Mixed models module. Decision Tree Analysis (DTA) is used in finding the cut-offs of the continuous variables. All analyses are done in SPSS20 environment.

3.2.2.3 Some statistical concepts used in the text

Within the text, three important concepts should be understood – they are repeated here from Section 2. *Statistical significance*, that is, p-value refers to the possibilities to generalize the result to the population. Practically speaking, behind the p-value (from "probability") is the fact that when measuring human mental processes – as the learning outcomes or attitudes are – there will always be measurement error. This means that the results of each individual student as well as each mean score carry error. Especially when the population is examined by using a sample, all the means carry both measurement error as well as sampling error. For example, there can be a small difference between the boys and girls. In this case, the p-value tells us how probable the same difference would be in the population at large. If the probability is $p < 0.05$, this means that the difference would be found at risk on 5% – only in five samples out of 100 the results would differ from that obtained. Parallel, if the p-value is $p = 0.002$ the risk for a faulty decision (of difference) is only 0.2%.

When the sample size is huge – as the sample of 16,350 students is – the *p*-value very easily gives a signal that the difference between the groups is real in population. *P*-value does not, however, tell whether the difference is small or big. For this purpose there is another statistic, Effect Size (ES). Effect size tells how far the lowest and highest groups came from each other. The commonly 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 high effect sizes. During the text, these boundaries are used as a "measurement stick" to indicate whether the difference is small, medium/moderate or high. The rough boundaries for these are collected in Table 3.2.3.

Table 3.2.3 Rough boundaries of effect sizes

ES	Cohen's d	Cohen's f
small	0.2 or below	0.1 or lower
medium	round 0.45	round 0.2
high	0.8 or higher	0.4 or higher

Hence, one may read "the difference between boys and girls is statistically significant ($p < 0.001$) but the effect size is small" in the text. This means that the difference between the boys and girls is real but it is, in reality, very small or not notable.

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

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

Hence, if $\eta^2 = 0.30$, then $f = \sqrt{(0.3/0.7)} = \sqrt{0.43} = 0.65$ showing very high effect size.

3.2.3 Results

3.2.3.1 Basic results

3.2.3.1.1 Population Analysis

When assessing the student achievement, the population is usually normally distributed in a large sample. Then, 50% of the students lie below average and 50% of the students lie above the average. This kind of distribution can be found in the Nepali language (Figure 3.2.1a).
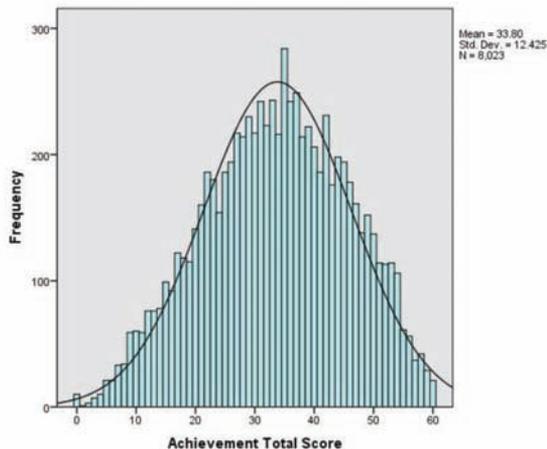
Figure 3.2.1a Normally distributed population in Nepali version 1

119

The Nepali sample was big enough to form a normal distribution (16,350 students). The original distribution of the latent ability (Theta) is normal. However, Figure 3.2.1b shows that though the population is normally distributed, in the final data, the total score seems to be slightly non-normally distributed. This is because of the equating process; some of the values were easier to gain than some others. In the final score of Nepali, the majority of the student seems to lie slightly in the high-performing part of the distribution. The distribution is normal enough for the statistical analyses.[14]
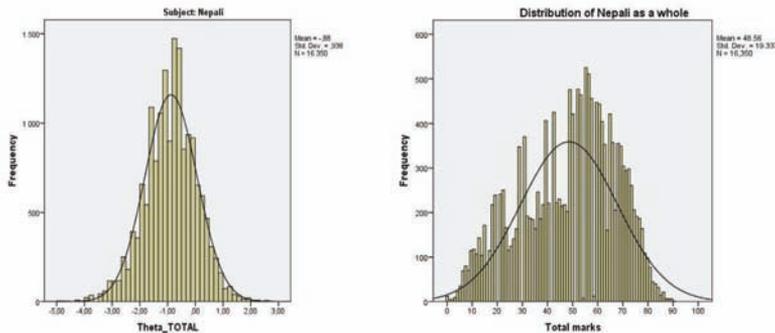
Figure 3.2.1b Final scores of Nepali as original latent ability (left) and as equated total (right)

The dataset gives evidence for the fact the grade 8 population in Nepali is normally distributed.

After separating the schools in to the community schools and institutional schools and plotting the population chart, a chart is given as in Figure 3.2.2:
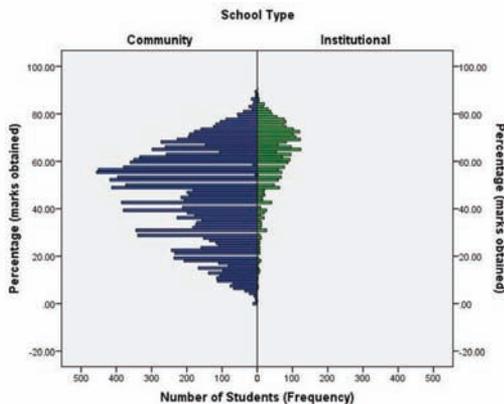
Figure 3.2.2 Distribution of the students' means in community schools and institutional schools

---

[14] Most of the classical methods for statistical analysis assume the Normal distribution of the population and in the sample.

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. Though the whole system is shifted to the better performing level, there are quite a number of the students in the community schools getting equally high marks as is gained in the institutional schools. Figure 3.2.2 strictly tells that the students in the community schools are varying from the low-performer to the highest performer whereas most of the students from the institutional schools are the higher performers.

Another related fact is that the schools are clearly divided into two "populations": the highly performing schools and the lowly performing schools. Figure 3.2.3 shows the distribution of the school means.
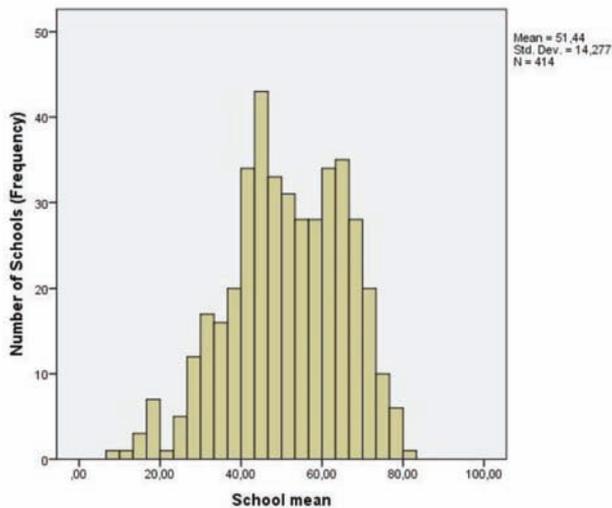
Figure 3.2.3 Distribution of school means

Figure 3.2.3 shows that, on the basis of school mean of the student performance, there are two categories of schools. One population seen on the left-hand side shows the average of round 40% of the maximum score and the other on the right-hand side with the mean of round 70%. The difference between the populations is remarkable. The difference between the school populations can be partly explained by the division between the private schools and community schools.

By analyzing the matter further with the scatter plot, Figure 3.2.4 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 schools to very low-performing schools.
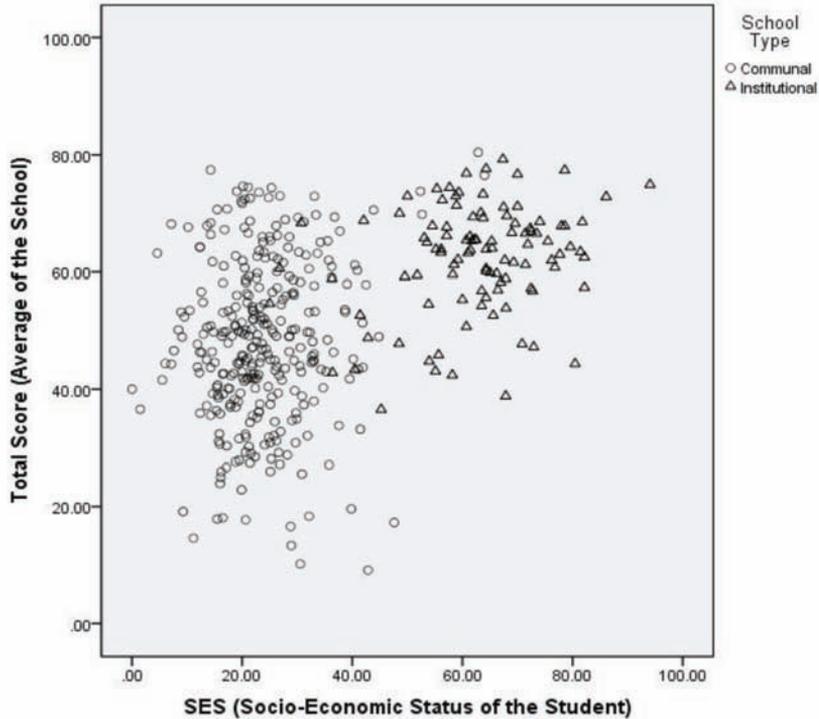
Figure 3.2.4 Achievement and Social economic status and type of schools

The dataset gives a strong signal that the students in the institutional schools perform very well and the students in the community schools form two kinds groups of schools: high-performing schools and low-performing schools. The variety between the community schools is remarkable.

3.2.3.1.2 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 four content areas were proportionally equal to the weighting given by the curriculum. To compare the achievement in all the topics, these sub-scores are converted into a percentage of the maximum score of the content area. Table 3.2.4 shows the student's achievement in Nepali as a whole (total) and the achievement level in four content areas; Figure 3.2.5 illustrates the differences.

122

Table 3.2.4.Comparison of the total score average with different topics

Content area	Mean	SD	Min	Max
Reading	55.6	19.7	0	100
Writing	47.7	21.2	0	93
Grammar	51.4	24.5	0	100
Vocabulary	43.0	13.1	0	69
Nepali as whole total	48.6	19.3	0	90
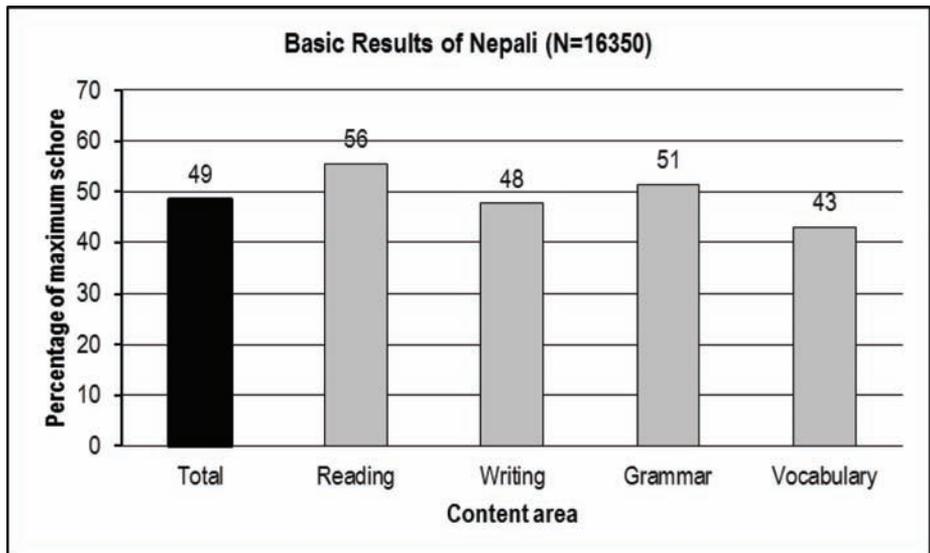
Figure 3.2.5 Basic results in the content areas of Nepali

The percentage of achievement score shows that the national average of Nepali is 49. Of the different content areas, students are the weakest in vocabulary (43) and writing (48). They perform better than average in reading (56), and grammar (51).

Dataset gives a signal that the learning outcomes are the highest in the content area of Reading and weakest in Vocabulary.

3.2.3.1.3 Hierarchical level of cognitive domain and achievement

The Nepali test as 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 is shown in Table 3.2.5 and illustrated in Figure 3.2.6.

Table 3.2.5 Achievement level based of hierarchical level of the items

Hierarchical level of item	Mean	SD	SE
Knowledge	74	25.3	0.20
Comprehension	54	16.4	0.13
Application	46	22.2	0.17
Higher Ability	42	20.6	0.16
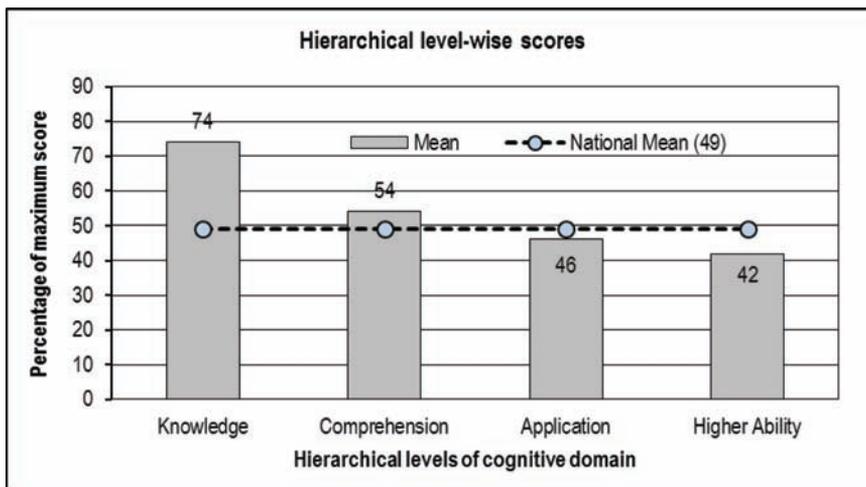
Figure 3.2.6 Achievement in different hierarchical levels of cognitive domain.

Remarkably high a number of students were able to solve only 15% or less of maximum score of the practical problems, that is, the application type of items (13% of the students). In Nepali, 14% of the students could solve just less than 15% of the tasks requiring the higher cognitive abilities.

The dataset shows that the students' ability to solve complex problems is quite low; only 42% of the maximum scores were reached. Students are much better in the recalling type of questions (74%), which may be explained by the educational system which seems to be geared to remembering the things rather than solving novel problems.

3.2.3.1.4 Type of item and achievement

There were basically two types of questions in the test: objectively marked- and subjectively marked items. Objective items covered a wide range of content areas

and were very specific to judge because there were 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 require 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 level. Table 3.2.6 comprises the basics statistics of the item type-wise achievement levels.

Table 3.2.6 Item type-wise means

Type of items	N	Mean	SD	SE
Objective	16350	58.0	18.03	0.14
Subjective	16350	39.9	21.99	0.17

It is obvious that the subjectively scored tasks – usually those with more demanding requirements for the correct answer – are solved much lower (40%) than the objective items (58%). Most of the objective items were knowledge-, comprehension- and application type whereas subjective items were application- and higher ability type.

Dataset gives a strong signal 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 started to do the open ended task (like free writing, problem solving and analysis) but the skills were not high enough for the highest marks.

3.2.3.1.5 Comparison of 2011 with 2008 achievement result

The national assessment carried out in various years aims to assess the achievement and the progress over a period of the years. The National Assessment of grade 8 student carried out by the Research Centre for Education Innovation and Development (CERID, 1999) shows that the national average of the students was round 69. Later, the National Assessment of grade 8 students carried out by the Educational Development Service Centre in 2008 (EDSC, 2008) shows the national average of the students is around 44. The National Assessment of 2011 carried

out by the NASA Unit of ERO (MOE) shows that the national average of Nepali in grade 8 is 49. These figures are coming from *Classical Test Theory (CTT) 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 different difficulty level of the tests.*

The Ministry of Education has made the decision of using IRT modeling in NASA to make the national assessment results comparable with the previous and to-come datasets. In NASA 2011, the tests were linked with the test of 2008 so that the comparison was possible. Tests of 2008 and 2011 were equated to the same scale by using international parameters (fixed difficulty parameters) from the PISA dataset. Certain selected items from the 2008 test were used to make the linkage between the tests. The results were not able use in comparison from the year 1999 because the whole database was missing.

By using the IRT modeling, the total score of the year 2008 was equated with the total score of the year 2011 datasets. After equating, the results are as condensed in Table 3.1; the scores are somewhat 2 percent points better in the year 2011 (49) than in 2008 (47). The difference is statistically significant ($p < 0.001$) though the effect size is small ($d = 0.23$).

Table 3.2.7 IRT generated means of 2008 and 2011

Year	N	Mean	SD	SE
2008	2640	46.8	15.1	0.29
2011	16350	48.6	19.3	0.15

The rising level of skills can be illustrated by comparing the percentages of correct answers in the identical linking items (Table 3.2.8). There were four items used as linking items from 2008 in the test of 2011. Item N08q3d and N08q3e are 1-point reading items. Variables N08q11 and N08q16 contain 4 and 5 sub-dichotomous items respectively. In Table 3.2.8, the percentage of maximum score is calculated. Only in one easy item, the average score has lowered. All the other, including the more difficult items, show that the students at the year 2011 performed better than the students at the year 2008. In most of the cases, however, the difference is very small.

Table 3.2.8 Comparison of the linking items of the years 2008 and 2011

	Percentage of correct answers						
	2008 (N = 2,640)			2011 (N = 16,350)			
Item no.	Reading	Vocabulary	Grammar	Reading	Vocabulary	Grammar	Change
N08q3d	0.87			0.76			+0.11
N08q3e	0.71			0.78			+0.07
N08q16		0.65			0.69		+0.04
N08q11			0.34			0.34	+0.01

The achievement in Nepali as a whole (total) and content area-wise comparison between the years is seen in Figure 3.2.7.
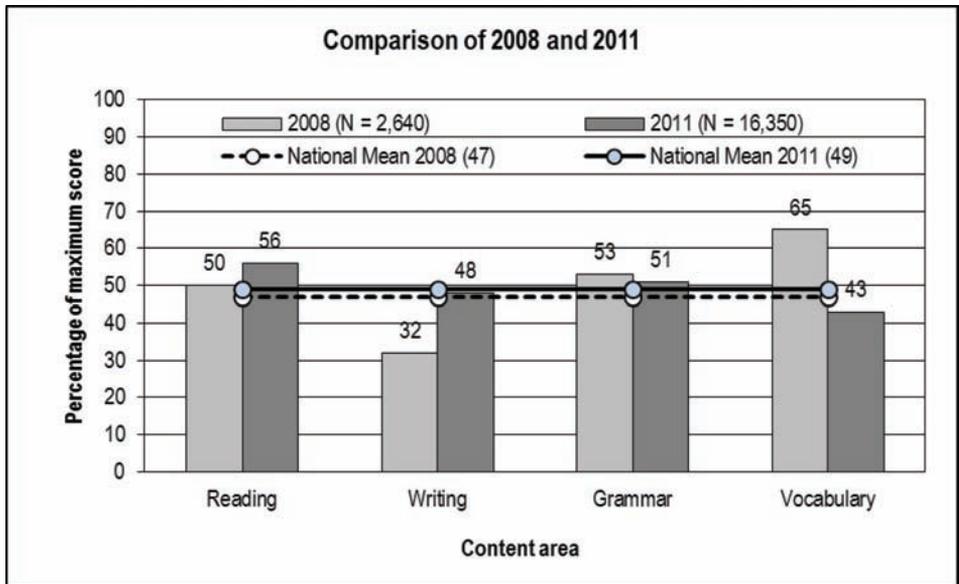
Figure 3.2.7 Content-wise differences between years 2008 and 2011

Except in the area of Writing, it was possible to equate the scores of the content areas by using the linking items. In Figure 3.2.7, Writing is, however, shown and it is good to take the comparison as a more or less civilized guess of the change between 2008 and 2011. The figures in Writing are based on modeling the Reading, Vocabulary, and Grammar first and then extrapolating the proficiency level of the students to the area of Writing. The achievement level in 2011 is higher than that in 2008 in Reading (+6 percent points) and Writing (+16 percent points) but it is

notable lower in Vocabulary (22 percent points) and somewhat lower in Grammar (2 percent points). The reasons are not discussed here.

The dataset gives a signal that the Nepali language skills have been developed in a positive way. The Nepali results rose 2 percent points within four years. Especially, the level rose in Writing (+16) and Reading (+6) but it has dropped dramatically in Vocabulary (−22).

3.2.3.1.6 Comparison with the international standard

The NASA 2011 was made comparable with the international PISA assessment. A good number of items (12) of the released PISA 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; when a student's ability level in NASA 2011 would be zero, it corresponds to the average level of the international students.

Figure 3.2.8 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 indicates the international average. As the ability is below the average, the bars are going down whereas when the ability is above the international average, the bars are going upwards.
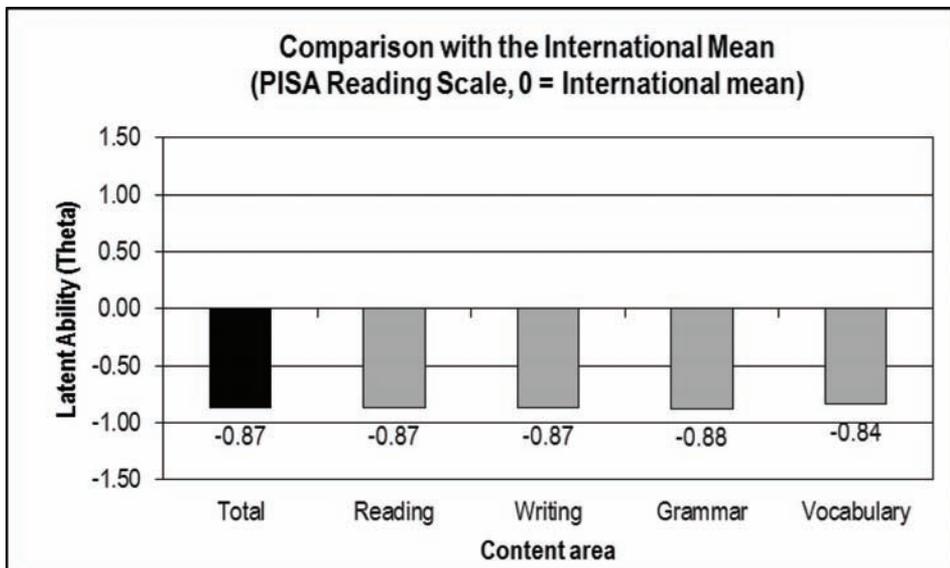
Figure 3.2.8 Student achievement in the international PISA reading scale

Figure 3.2.8 shows that the average ability shown by Nepali students in Nepali as a whole (*Total*) is below the average line. This indicates that the students in Nepal are somewhat poor in language subject when compared with the international average. Similarly, the students show much lower ability in *all* the content areas.

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 the Grammar and Vocabulary because, in PISA, these areas are not measured at all. However, they are modeled on the basis of proficiency in the reading test.

Another note may be worth highlighting in this context: another type of international comparison was done on the basis of Common European Framework for language testing (CEFR) was used to assess the criterion-based proficiency in Nepali language. This analysis is presented in a separate article on Section 4.6.

Dataset gives a signal that, from the international comparison viewpoint, the average reading proficiency in Nepal is much lower than the international average.

3.2.3.2 Results based on diversity factors

Diversity is a relative and contextual term. In the context of Nepal, some of the experts have defined eight diversities in Nepal: geographical/ecological-, language-, gender/sex-, religious-, ethnic-, cultural-, disability- and economic diversity. NASA 2011 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) diversity. These factors can be taken as equity factors; all children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goal.

3.2.3.2.1 Districts and student achievement

Out of 75 districts, 25 were randomly selected to represent the Ecological zones and Developmental regions and ultimately 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 condensed in Table 3.2.9 and Figure 3.2.9. The table shows the achievement in ascending order according to the achievement. The *mean* represents the average achievement percentage of the particular district.

Table 3.2.9 Average achievement score in the selected districts

Districts	N	Mean	SD	Districts	N	Mean	SD
Mustang	36	70.0	10.0	Jhapa	1169	45.7	19.8
Bhaktapur	245	69.3	11.8	Rautahat	873	44.9	21.7
Lalitpur	336	63.9	15.3	Bajhang	296	43.1	16.7
Kathmandu	1441	62.7	14.0	Nawalparasi	1009	42.6	20.4
Dhading	886	60.2	14.9	Ilam	927	42.3	15.9
Tanahu	947	57.4	15.1	Kanchanpur	584	42.1	16.3
Gorkha	775	57.2	15.5	Doti	365	42.0	13.8
Sindhupalchowk	384	50.6	15.4	Banke	720	41.8	17.2
Syanjha	941	50.2	17.4	Bhojpur	532	39.7	16.3
Sankhuwasabha	279	50.1	16.2	Bara	732	39.5	22.6
Pyuthan	580	46.5	17.3	Morang	975	38.8	18.2
Kalikot	194	46.4	20.7	Jajarkot	437	38.3	19.7
Ramechhap	687	46.4	16.9				
			Total	16350	48.56		19.3
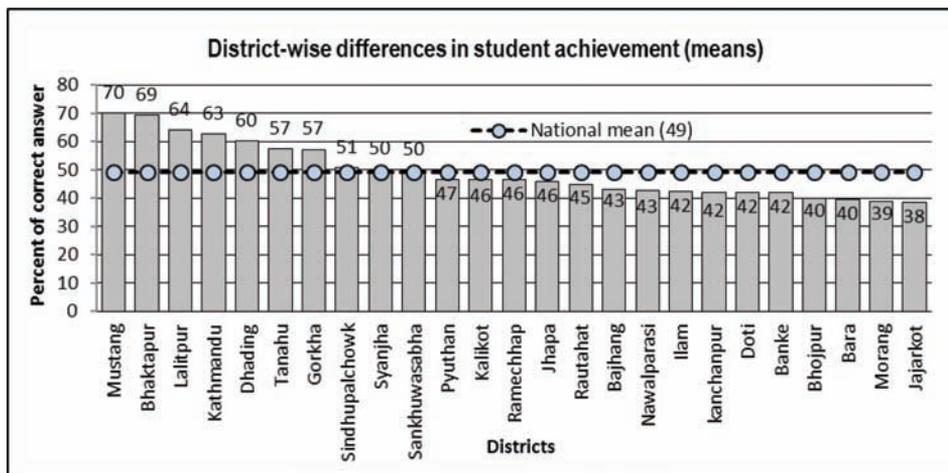
Figure 3.2.9 Achievement in the districts in the sample.

Of the randomly selected districts in the sample, the student performance was very low in Morang (39), Bhojpur (40), and Ilam (42) from the Eastern region, in Bara (39) from the Central, in Jajarkot (38) and Banke (42) from the Mid-Western-,

and Doti (42) and Kanchanpur (42) from the Far-Western region. Except for Mustang district (70), the best performing four schools come from the Central region and specifically in the Valley area: Bhaktapur (69), Lalitpur (64) and Kathmandu (63).

The difference in achievement due to the different (low and high performing) district is statistically significant ($p < 0.001$). The variation explained in achievement due to the district is $\eta^2 = 0.191$, that is, the district explains 19% of the variation in the data. Effect size is $f = 0.49$ indicating that the difference between the lowest performing district (38) and highest performing district (70) is remarkably high.

The dataset gives a strong signal 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 25 districts selected randomly in the sample; even lower-performing districts could be found if other districts would have been selected. In any case, the results in Jajarkot (38), Morang (39), Bara (39), Bhojpur (40), Banke (42), Doti (42), Kanchanpur (42), and Ilam (42) is so low that raising the standard in these districts would arise the standard in the whole country.

3.2.3.2.2 Ecological zone and student achievement

Access to educational facilities can play a vital role in students' achievement. The Mountain, Hill and Tarai are three geographical features in Nepal though the Valley can be 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, weather conditions, economic activities, aggressive development, as well as the dense human capacity make the Valley a unique fourth geographical area in the analysis. The variation in the Ecological zones in NASA 2011 is condensed in Table 3.2.10 and Figure 3.2.10.

Table 3.2.10 Achievement in the Ecological zones

Region	N	Mean	SD	SE	Minimum	Maximum
Mountain	1189	49	17.4	0.5	0	87
Hill	7077	50	17.9	0.2	0	89
Valley	2022	64	14.1	0.3	7	89
Tarai	6062	42	19.9	0.3	0	90
Total	16350	49	19.34	0.15	0	90
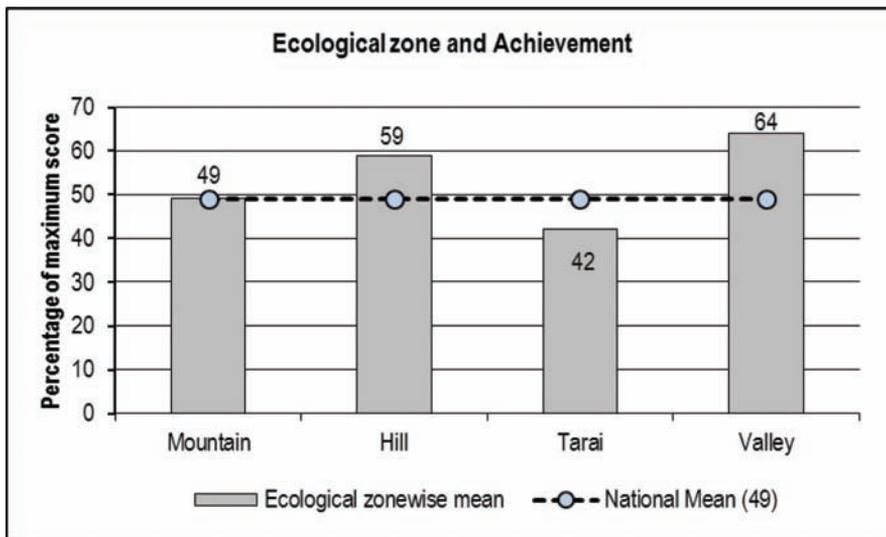
Figure 3.2.10 Differences in achievement in different Ecological zones

The data shows that, on average, the students from the Valley (64) outperform the students from all the other Ecological zones. The students from the Tarai area perform the lowest (42). It is noteworthy that some of the lowest performing districts, such as Morang (39), Bara (39), Doti (42), Banke (42), and Kanchanpur (42) are from Tarai.

The achievement in the regions differs significantly ($p < 0.001$). Tukey's *post hoc* test tells that all the zones deviate from each other in a statistically significant manner of at least $p = 0.050$ level. The effect size is $f = 0.35$ showing moderate or high difference between the highest and lowest performing Ecological zones; Ecological zone explains 11% of the variance in the data. As a comparison, remember that the district explains more than 19% of the variation.

Dataset gives a signal that there is a moderate difference between the student performances in four Ecological zones. Students in the Kathmandu Valley outperform the other students. The achievement is the lowest in Tarai area.

3.2.3.2.3 Developmental region and student achievement

The student achievement varies according to the Developmental regions which are divided into Eastern-, Central-, Western-, Mid-Western-, and Far-Western. Additionally, the Kathmandu Valley is taken as the 6th Developmental region though technically it falls under the Central Developmental region. The mean achievements in the Developmental regions are given in Table 3.2.11 and illustrated Figure 3.2.11.

Table 3.2.11 Achievement in the developmental regions

Region	N	Mean	SD	SE	Minimum	Maximum
Eastern	3882	42.7	18.1	0.29	1	90
Central	3562	48.5	20.3	0.34	0	87
Western	3708	51.6	18.5	0.30	0	89
Mid-Western	1931	42.9	18.5	0.42	1	86
Far-Western	1245	42.3	15.7	0.45	3	81
Valley	2022	63.7	14.1	0.31	7	89
Total	16350	48.6	19.3	0.15	0	90
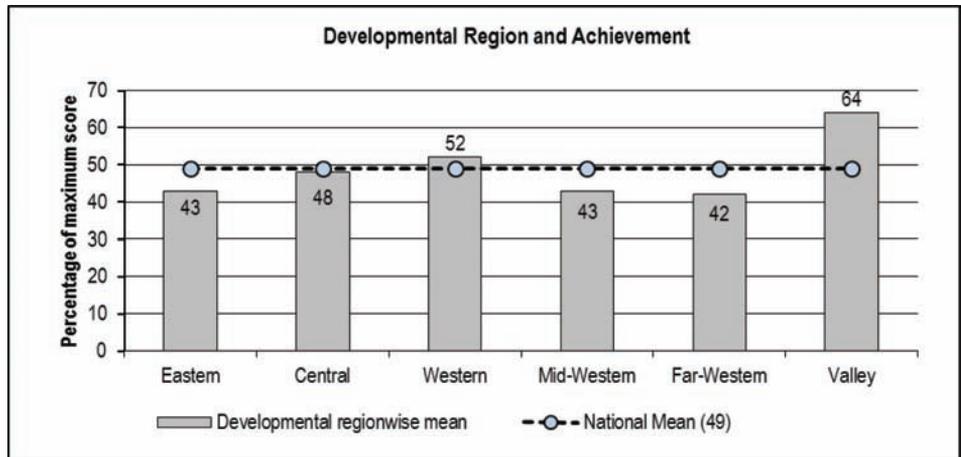
Figure 3.2.11 Comparison of student achievement in the developmental regions

The best performance can be found in the Valley (64) and in the Western region (52). The performance is the lowest especially in the Far-Western (42) though the Mid-western (43) and Eastern (43) regions are also areas were student performance is lower than the average. The difference between the regions is statistically significant ($p < 0.001$) and Tukey's *post hoc* test shows that all the Developmental regions differ from each other significantly with at least at the 5% significant level. Effect size is moderate or high ($f = 0.37$); Developmental region explains 12% of the student variation.

The dataset gives a strong signal that there is waste inequity in the Developmental regions of children's opportunities to reach an adequate level of Nepali. The difference between the lowest performing area (Far-Western,

133

42) and the highest performing region (Kathmandu Valley, 64) is remarkable – more than 20 percent points.

3.2.3.2.4 School type and student achievement

All the schools are categorized into community- and institutional schools. In what follows, these are referred to as community schools and institutional schools; the latter could also be called private schools. The difference in the Nepali achievement within the two types of schools is condensed in Table 3.2.12.

Table 3.2.12 Type of school and the average achievement

Type of school	N	Mean	SD	SE
Community	13663	46	19.0	0.16
Institutional	2687	62	14.5	0.28
Total	16350	49	19.3	0.15

The achievement levels in the community schools and institutional schools differ from each other remarkably. The average performance in the private schools is 62 whereas in the community schools it is 46; 16 percent point's difference is remarkable. The difference is statistically significant ($p < 0.001$) and the effect size is moderate or high ($f = 0.33$) showing a wide difference between the community- and institutional schools. Division of the students to the community- and institutional schools explains 10% of the student variation ($\eta^2 = 0.099$). From Figure 3.2.4 it is known that the deviance within the community schools is remarkable ranging from 20% to 80%; contrarily, most private schools in the sample show very high performance. Natural reason for this is the student selection. The selection of the students evidently takes the best students from the community schools and hence the average performance of the community schools may be lower when there are possibilities to select the school. It may be worth noting that the best community schools are not in Kathmandu Valley. It is also worth noting that the number of the students in the private schools is not very high; the main educational work is done in the community schools.

The dataset gives a strong signal that, on average, the students in the institutional schools outperform the students in the community schools. Most probably this deviance can be explained by the student selection, however, this cannot be known on the basis of the dataset.

3.2.3.2.5 School location and student achievement

One of the strata of sampling in NASA 2011 was the school location. The schools were divided into two: rural- and urban schools. The achievements of the students in rural and urban schools are condensed in Table 3.2.13.

Table 3.2.13 Student achievement on the basis of location of school

Type of school	N	Mean	SD	SE
Rural	12105	46	19.1	0.17
Urban	4245	55	18.5	0.28
Total	16350	49	19.3	0.15

The achievement level of the students in the urban schools (55) is remarkably higher than that of rural schools (46). The high level of the schools in Valley raises the level of the urban schools in general. The difference in average score is significantly different ($p < 0.001$) and the effect size is moderate ($f = 0.20$); division into rural and urban schools explain 4% of the student variation ($\eta^2 = 0.041$).

Data gives a strong signal that the students in the urban schools would gain 9 percent point more than the students in the rural areas. Without the Valley, the difference is not notable. However, from the educational equity point, this is not a good sign.

3.2.3.2.6 Language at home and student achievement

In the context of Nepal, the student achievement may depend on the language spoken in their homes i.e., the mother tongue of the students. The mother tongue reflects, in many cases, the ethnical background and hence any difference may be taken as a possible source for inequity in society. In Nepali language assessment (see also Section 4.6) the results are obvious.

On the basis of the total data, 30.4% of the 8th 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 Magars (3.2%), Tamangs (3.1%), and Tharus (2.2%). After dividing the languages into ten groups excluding Nepali, there were still 18.5% of the students classified into the group "else". 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 were grouped into "non-Nepali speakers". The results are condensed in Tables 3.2.14 and 3.2.15 and illustrated in Figure 3.2.12.

Table 3.2.14 Student achievement on the basis of home language

Language group	N	Mean	SD	SE
Nepali	11382	51	18.6	0.17
Non-Nepali	4968	43	19.9	0.28
Total	16350	49	19.3	0.15

When combining all the minor language groups as "Non-Nepali", there is a notable difference between the language groups (8 percent points favoring the Nepali speakers). However, on the basis of Table 3.2.15 with the original categorization of the minor languages, the issue looks quite much complex: It is evident that the students from Newar- (59) and Tamang (55) background perform higher than Nepali students (51). On the other hand, the students from Limbu- (40), Tharu- (40), "other"- (41), and Rai (42) background perform much lower.

Table 3.2.15 Achievement in the different language/ethnic groups

Language/Ethnicity[1]	N	Mean	SD	SE	Minimum	Maximum
Newar	178	59	18.0	1.35	6.4	84.0
Tamang	500	55	16.5	0.74	5.3	85.1
Nepali	11382	51	18.6	0.17	0.0	90.4
Sherpa	32	48	13.7	2.43	10.6	77.7
Magar	525	46	17.4	0.76	6.4	83.0
Gurung	111	45	18.8	1.78	6.4	83.0
Rai	36	42	16.4	2.73	7.5	68.1
Tharu	364	40	18.0	0.94	5.3	74.5
Limbu	83	40	17.0	1.86	6.4	73.4
Other	3032	41	20.0	0.36	0.0	85.1

1)Those language groups in which number of the students was less than 13 and missing are omitted.

136
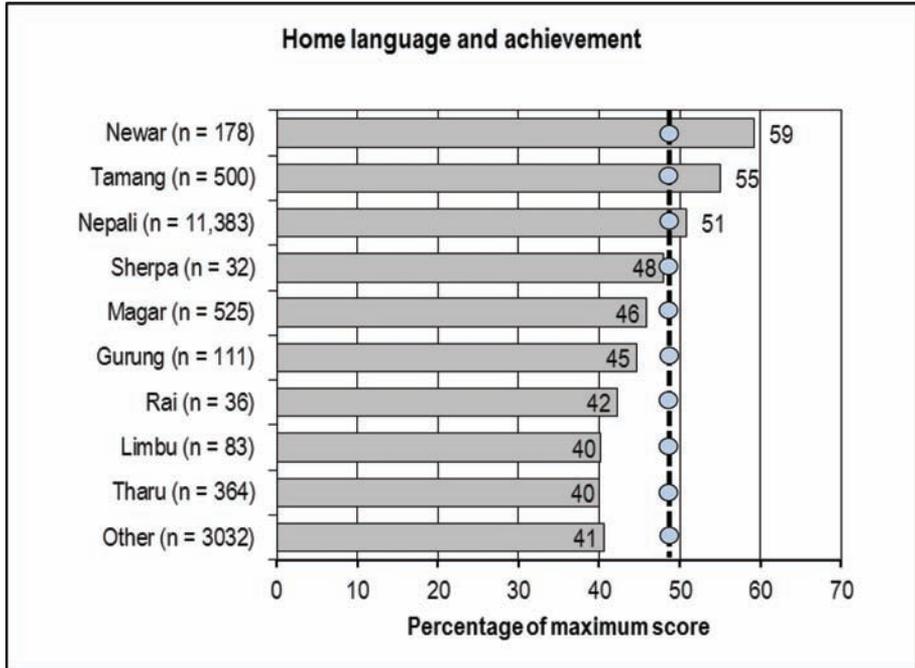
Figure 3.2.12 Relation between language at home and achievement

The difference between the language groups is statistically significant ($p < 0.001$). However, the effect size is medium ($f = 0.24$) because the minority groups are small; division into smaller language group explains somewhat 6% of the variation in the data ($\eta^2 = 0.058$). When analyzing only the minority languages and hence, excluding the Nepali speakers and the group "else", the effect size is high ($f = 0.40$) indicating really remarkable difference between the highest performing group (Newar, 59) and the lowest performing group (Tharu, 40).

Language and developmental region

When combining the results from the Developmental region and mother tongue, one notices that the achievement score of the students within a certain language group varies radically between the different regions. All language groups have a high score in Valley.

137

Table 3.2.16 Achievement in the different language groups in different regions

Developmental region	Nepali[1]	Magar	Tharu	Tamang	Newar	Rai	Gurung	Sherpa	Limbu
Eastern	44.8[2]	39.7	31.7	31.0	33.1	41.9	39.6	46.9	39.5
Central	52.8	35.2	40.2	54.4	47.9		47.5	47.9	
Western	54.4	46.9	36.8	25.5	52.7		44.4		56.4
Mid-Western	42.9	48.5	51.1				31.9		
Far-Western	39.4	44.9	40.3	27.1			57.2		
Valley	63.8	69.5	53.2	61.3	64.0	55.3		54.0	73.4
Total	50.8	45.9	39.9	55.0	59.2	42.2	44.6	47.9	40.1

1) The language groups of less than 14 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.

Language and Ecological zone

In all language groups, except the Limbu population, the results in Tarai region are lower than the mean and in all the language groups, except Tharus and Rais, the highest scores are found in the Valley (Table 3.2.17). The reason for the phenomenon stays open. However, it means that the home language alone does not determine the future of the children. As a single example, one may take the only Limbu speaking student in the valley whose score 73.

Table 3.2.17 Achievement in the different language groups in different zones

Ecological zone	Nepali[1]	Magar	Tharu	Tamang	Newar	Rai	Gurung	Sherpa	Limbu
Mountain	48.2			50.2[2]	45.9	56.0	53.2	46.7	53.2
Hill	50.0	48.9	69.7	57.4	44.4	40.4	45.3	48.7	38.3
Valley	63.8	69.5	53.2	61.3	64.0	55.3	57.2	54.0	73.4
Tarai	46.0	39.4	39.7	39.9	43.1	40.9	37.5		40.9
Total	50.8	45.9	39.9	55.0	59.2	42.2	44.6	47.9	40.1

1) The language groups of less than 14 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.

The Dataset shows that there is an educational inequity within the language groups. The students from Newar- (59) and Tamang (55) backgrounds perform higher than Nepali students (51). On the other hand, the students from Limbu- (40), Tharu- (40), "other"- (41), and Rai (42) backgrounds perform much lower than the national average. Ecological zone-wise analysis raises an important question: Why are the results in Nepali so low in Tarai region and what can be done to raise the standard?

3.2.3.2.7 Caste and student achievement

Modern education in Nepal has been influenced in several ways by legacy of the historical caste system. Though the caste system is officially abandoned it still lives in the mind-sets of most Nepali people. Historically, the Brahmans and Cheetris have been heavily involved in education, but Dalits, for example, have been practically outside or less participate of the educational system. Hence, modern society has made lots of efforts to make the education possible and accessible for all children. The latest household survey (2012) shows that the number of Hill Dalits has been increased remarkably in the lower education system but their number in the secondary and higher education is still very small. The results concerning the casts and achievement are condensed in Table 3.2.18 and illustrated in Figure 3.2.13.

Table 3.2.18 Achievement in the different language groups in different castes

Cast	N	Mean	SD	SE	Minimum	Maximum
Brahman/Cheetri	6176	51.8	18.4	0.2	0.0	90.4
Janjati	6615	49.4	18.3	0.2	0.0	88.3
Dalit	1523	44.9	19.0	0.5	3.2	84.0
Madhesi	1690	37.8	22.0	0.5	0.0	86.2
Alpasankhyak	306	45.1	20.1	1.2	0.0	79.8
Total	16310	48.6	19.3	0.2	0.0	90.4
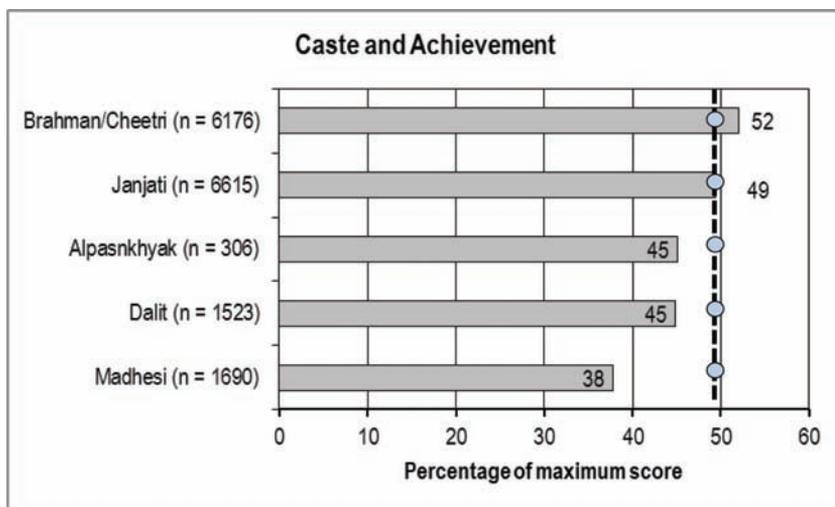
Figure 3.2.13 Relation between caste and achievement

On the basis of NASA 2011, the Madhesi students are performing the lowest (38) in Nepali, but in Mathematics the Dalits are performing lower than the other casts. The overall difference between the groups is statistically significant ($p < 0.001$) and the effect size is medium ($f = 0.22$); dividing the students according to their caste background explains 4.7% of the student variation ($\eta^2 = 0.047$). Tukey's *post hoc* test reveals that Alpasankhyaks and Dalits do not differ from each other. Otherwise, all the means of the caste groups differ from each other statistically with at least a 5% risk ($p < 0.05$).

A positive sign from the equity viewpoint is that the Dalit students perform remarkably better than the national mean (49) in the Western Mountain- (67), Western Hill- (54), and Central Hill area (54) (Table 3.2.19a). Unfortunately, though, it seems that the results are much lower than the average in the Far-Western Mountain area (30) and all regions in Tarai (ranging 35–42). The number of the students in certain strata is small and hence it may be wise not to make the too strong and farfetched implications of the results. On the other hand, from the equity point of view it is a negative sign that Madhesi students' performance in Nepali language is very low (Table 3.2.19b) – it is especially low in the Western- (29) and Eastern regions (33); but in Valley they perform exceptionally high (62).

Table 3.2.19a   Dalit students' achievement in different Ecological- and develop mental areas

	Eastern	Central	Western	Mid-Western	Far-Western	Total
Mountain	49.4	46.1	67.2	41.6	29.9	41.7
Hill	41.6	54.1	53.9	44.5	40.5	49.9
Tarai	35.2	42.1	38.2	36.6	35.9	37.5
Total	38.4	47.3	50.7	41.0	35.6	44.9

Table 3.2.19b Madhesi students' achievement in different developmental regions

	Eastern	Central	Western	Mid-Western	Far-Western	Valley	Total
Madhesi	33.1	39.2	28.6	40.8	38.0	62.2	37.8
other	43.9	51.9	52.2	43.0	42.4	63.7	49.8

In Section 4.6, it is noticed that the reading- and especially writing skills within the Madhesi students are so low that more than 50% of the students are at the lowest levels of the writing skills measured by the test. Practically speaking, a notable percentage of Madhesi students have not reached the required level of reading and writing Nepali to be able to continue in higher education in the Nepali language.

The reason for the low performance may be caused by social reasons; the Madhesi population has historically had a very strong connection to the Indian population and they may think that learning Nepali is not useful when most connections are with Indians.

Dataset gives a signal that the Madhesi students' performance is radically lower than the results of the other castes. Their performance is especially low in the Western- and Eastern Developmental regions. A notable percentage of Madhesi students have not reached the required level of reading and writing Nepali to being able to continue in higher education in the Nepali language.

3.2.3.2.8 Sex and student achievement

Lots of effort has been put globally into reducing the difference between boys' and girls' school achievement. Because the sex- or gender-wise equity seems to be important in the modern discourse, the matter is handled somewhat more extensively than the previous sections of equity. Basic results are condensed in Table 3.2.20 and Figure 3.2.14.

Table 3.2.20 Student achievement of boys and girls

Sex	N	Mean	SD	SE	Minimum	Maximum
Boys	7898	48	19	0.2	0.0	89.4
Girls	8421	49	19	0.2	0.0	90.4
Total	16319	49	19	0.2	0.0	90.4
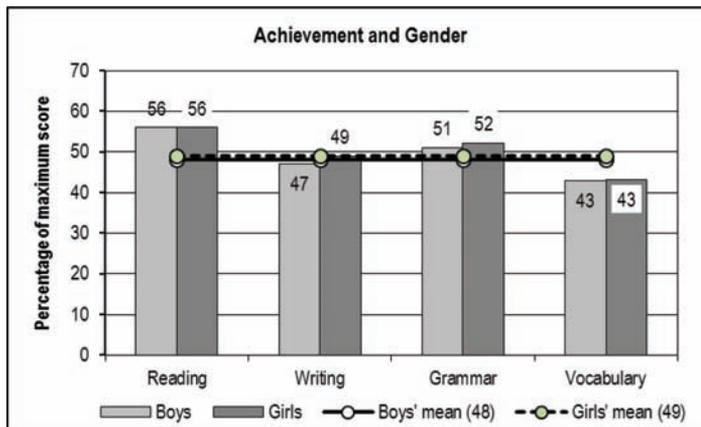
Figure 3.2.14 Comparison of boys and girls in different content areas

141

There is a statistically significant difference between boys (48) and girls (49) ($p < 0.001$) in the total score though the real difference is in the area of writing; girls are better writers (49) than boys (47). Girls are also slightly better in the area of Vocabulary ($p = 0.005$). In all areas, the differences are small (effect sizes are $d < 0.10$). Sex explains only 0.8% of the student variation. From the equity point of view the signal is positive thought there is still work to do to reduce the gap.

Sex and caste

The difference between boys and girls is the highest in the Madhesi caste (difference is 5 percent points) where the boys outperform the girls. Tukey's post hoc test shows that the differences are statistically significant at $p < 0.01$ risk level. Actually, the Madhesi boys outperform the girls only in the Central- and Western regions (by 7% and 9 percent points respectively); in the Far-Western region and the Valley the girls are remarkably better than the boys (19% and 14 percent points respectively). The sample sizes in the latter two regions are very small though.
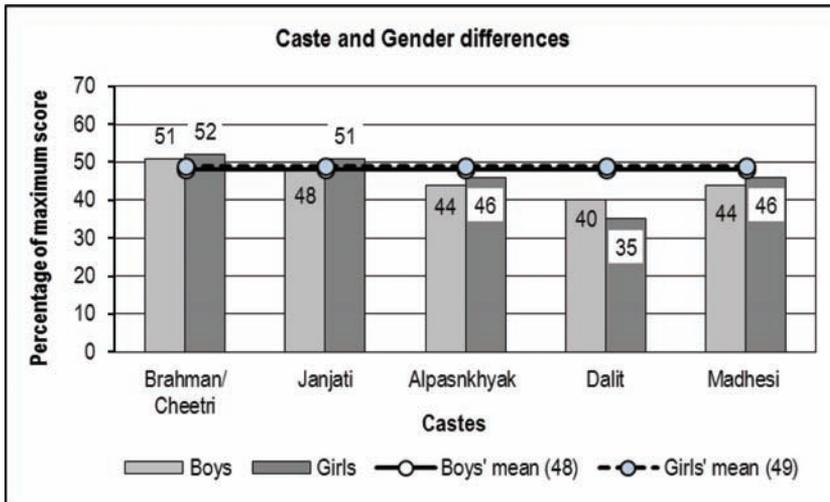
Figure 3.2.15 Caste/ethnicity-wise comparison of achievement of boys and girls

Sex and ecological zone

The achievement of girls and boys differs significantly between the ecological belts. Namely, in Mountain-, Hill-, and Valley regions girls outperform boys ($p < 0.001$) but in Tarai, there is no statistically significant difference between boys and girls though girls in the sample are somewhat better than the boys. (Figure 3.2.16).

142
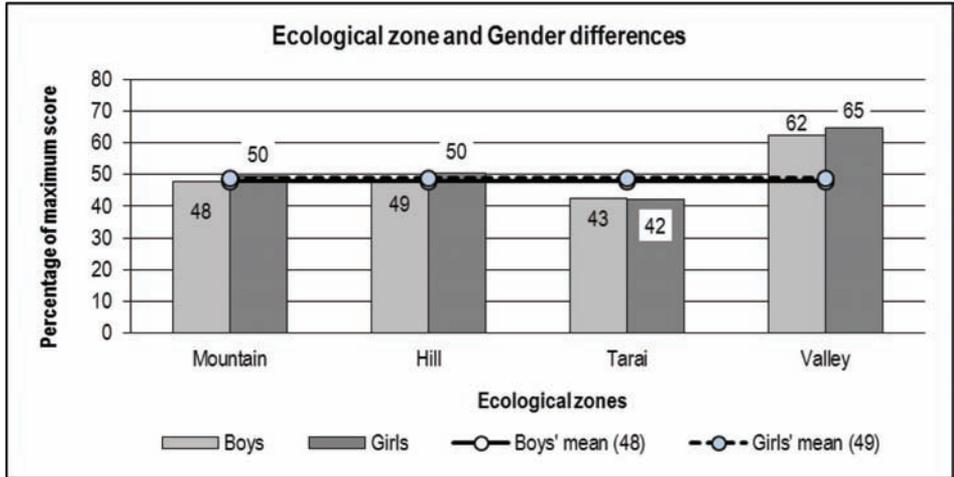
Figure 3.2.16 Ecological zone and gender-wise differences

When it comes to the Ecological zones, the differences between boys and girls are very small *(f = 0.03)*.

Sex and developmental region

There are no notable differences between the Developmental regions when it comes to boys' and girls' equal opportunities to reach the same educational goals. The difference between boys and girls seems somewhat wider in the Kathmandu Valley though.
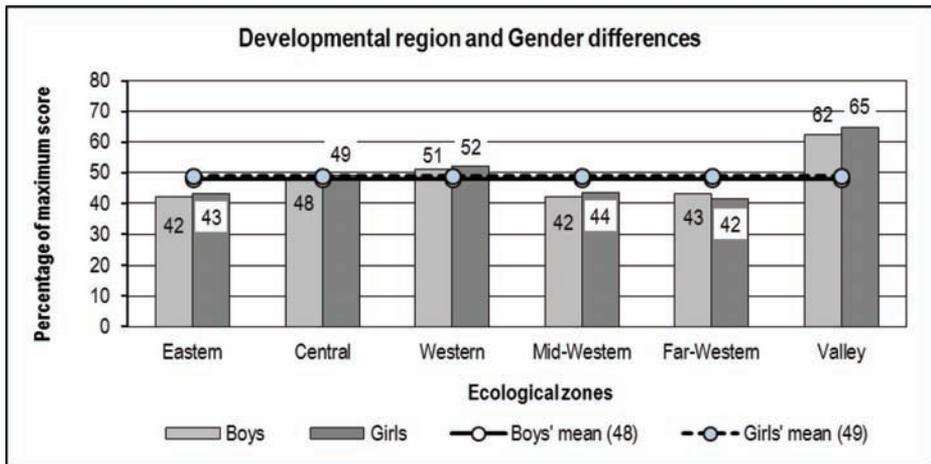
Figure 3.2.17 Developmental region and gender-wise differences

143

Dataset shows that the girls are slightly out-performing boys in writing and in the total score. Differences are not wide though they are significant. Sex explains only 1% of the student variation; from the equity point of view the signal is positive thought there is still work to do to reduce the gap. There are no notable differences between sexes in Ecological zones or Developmental regions.

3.2.3.3 Selected explanatory factors and achievement

The simplistic model in Section 2.3.1 represents several possible factors, which may explain the differences in student achievement. Many of the factors have already been handled in Section 3.2.3.2: geographical factors, such as Districts, the Ecological zone, and Developmental region, as well as School-related technical factors, such as school Type and school Location. Also some individual related factors were handled, such as Home language, Caste and Sex/Gender. In this section, several other factors are taken into consideration. The socioeconomic status (SES) of the students' families, paid work after school, Students' attitude towards Nepali as a school subject, Age of the student, and Help given to the studies are mainly family- and individual related factors. As a sample of deepening school- and teacher-related factors, also the Availability of schoolbooks, Homework given by the teacher, and Selected activities in the school are handled. Many other factors could be selected; the background questionnaire is rich.

3.2.3.3.1 Parents' education and occupation and student achievement

There are many variables indicating the socioeconomic status. In NASA 2011, 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, cars or bathrooms there is 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 2.4.2). In this section, the education of the parents is further elaborated on, so that the 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, Decision tree analysis (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. Some examples of this are

handled with parents' educational background and its relation with students' achievement in Nepali.

Parents' education

In NASA 2011 background questionnaire, the parents' education is divided into nine categories: 1) illiterate, 2) literate, 3) grades 1–5 pass, 4) grades 6–10 pass, 5) SLC pass, 6) IA pass, 7) BA pass, 8) MA pass, and 9) Above MA pass. The question was asked from the students and hence there may be some impurities embedded in the data. However, with the huge dataset the results seem credible.

DTA classifies mother's education into four groups with statistically significant differences in students' achievement levels: illiterate mothers (students' average is 45% of the maximum score), just literate mothers (49%), grade 1–5 passed mothers (51%), grade 6–10; SLC pass or higher mothers (52%), (Figure 3.2.18, of the explanation of the elements in the figure, see Section 2.7.1). In each group, the number of mothers is high enough to make a credible prediction. The difference between each group is statistically significant ($p < 0.001$). In practical words, the results mean, for example, that when the mother has passed at least grade 10, she can give +8 percent point advance in the national test compared with illiterate mothers.
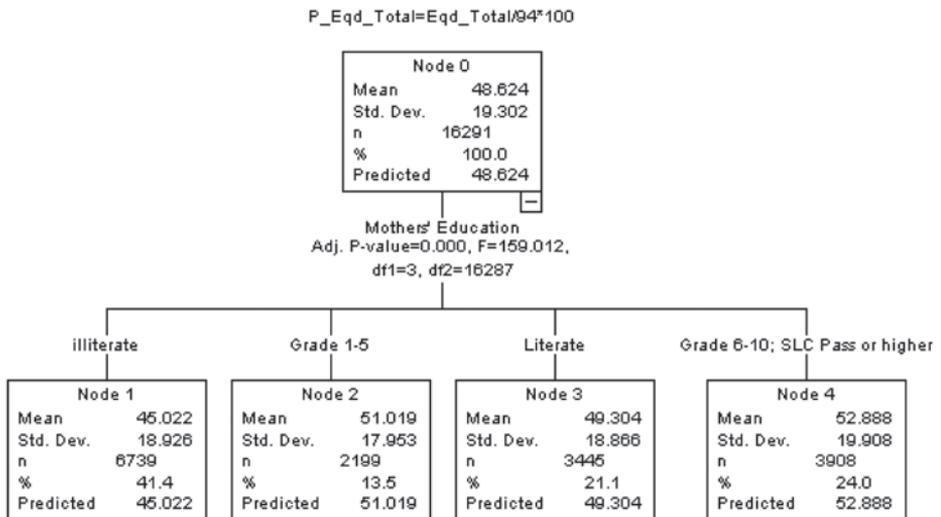
Figure 3.2.18 DTA of mother's education and students' achievement in Nepali

In parallel, DTA divides the father's education into five categories: illiterate (43), literate (47), grade 1–5 passed (49), grade 6–10 passed fathers (50), and SLC passed or higher fathers (53) (Figure 3.2.19).
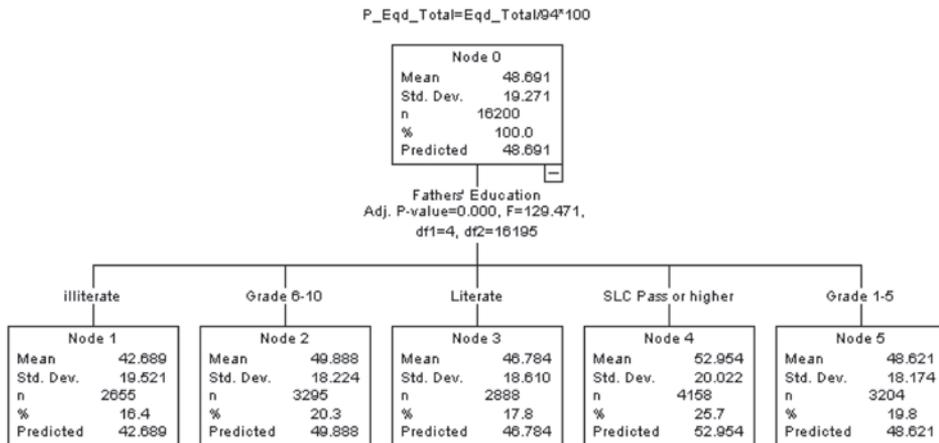
Figure 3.2.19 DTA of father's education and students' achievement in Nepali

One can infer that when the father has passed grade 10, the family can give +7 percent points' advance to their children compared with families with an illiterate father. When the father has passed the SLC- or higher, the advance would be 10 percent points.

After combining the mother's and father's education, the poorest prediction for the children's future achievement in Nepali comes when the mother or father is illiterate (43%). The highest results are in the group where both the father and mother have passed the SLC (56%) It seems 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, the cut-off for parental education was set to "SLC-passed", that is, when being SLC-passed (or higher), the indicator for mother's (and father's) education for SES was set to 1, and the lower education than SLC-passed gave the value 0.

Parent's occupation

The occupation of parents was categorized into five groups: 1) agriculture, 2) teaching, 3) services, 4) business, and 5) others. While comparing the mean by ANOVA, the student achievement is the lowest when the mother's occupational

background comes from agriculture (47). It is statistically significantly lower than when the mother comes from teaching (49), services (53), business (52) and other occupations (55) in round figures. The result is seen in Figure 3.2.20.

P_Eqd_Total=Eqd_Total/94*100

```
                    ┌─────────────────────────┐
                    │        Node 0           │
                    │ Mean          48.854    │
                    │ Std. Dev.     19.207    │
                    │ n             16029     │
                    │ %             100.0     │
                    │ Predicted     48.854    │
                    └─────────────────────────┘
                q22 Mother's occupation
              Adj. P-value=0.000, F=171.164,
                    df1=2, df2=16026
```

	<= 1.0	(1.0, 4.0]	> 4.0

```
┌─────────────────────────┐ ┌─────────────────────────┐ ┌─────────────────────────┐
│        Node 1           │ │        Node 2           │ │        Node 3           │
│ Mean         47.092     │ │ Mean         51.613     │ │ Mean         54.908     │
│ Std. Dev.    18.639     │ │ Std. Dev.    20.353     │ │ Std. Dev.    18.783     │
│ n            11078      │ │ n            3175        │ │ n            1776        │
│ %            69.1       │ │ %            19.8        │ │ %            11.1        │
│ Predicted    47.092     │ │ Predicted    51.613     │ │ Predicted    54.908     │
└─────────────────────────┘ └─────────────────────────┘ └─────────────────────────┘
```
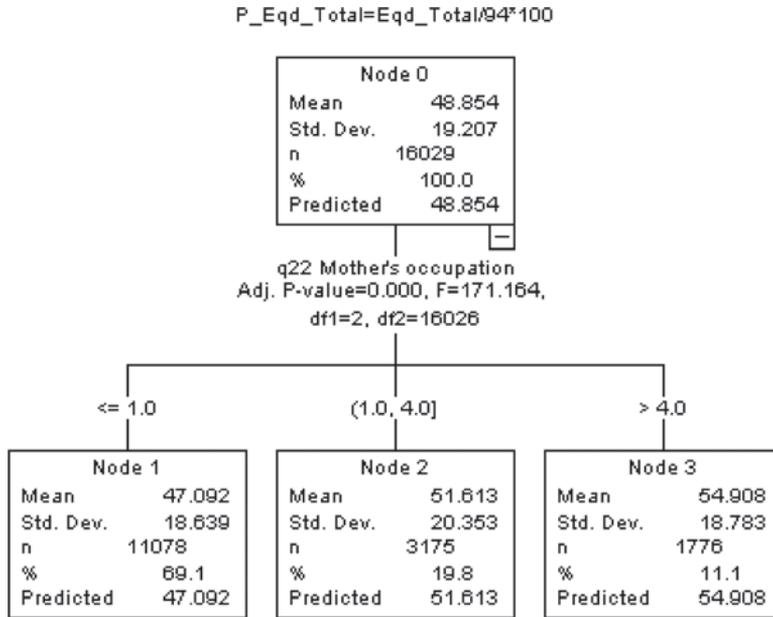
Figure 3.2.20 DTA of mother's occupation and students' achievement in Nepali

In a similar manner as with the parents' education, DTA was used to find the statistically the most deviating groups related to student achievement. This analysis finds three significant nodes. The lowest achievement is in the group where the mother comes from agriculture (47). Significantly higher achievement is seen when the mother is a teacher, in services, or in business (52). The highest achievement is in the group "other" (55). Because it is not precisely defined the kind of occupation which lies in "other", it is not easy to use the result. However, the result indicates that the mothers who are not in agriculture may be very beneficial for the achievement of the students.

When it comes to the father's occupation, the main division is whether the father works in agriculture (45%) or not. If the father were in a teaching or service profession, the students would gain 7 percent points more in the test (52%).

147

P_Eqd_Total=Eqd_Total/94*100

```
                          ┌─────────────────────────┐
                          │         Node 0          │
                          │ Mean           48.936   │
                          │ Std. Dev.      19.168   │
                          │ n              15916    │
                          │ %              100.0    │
                          │ Predicted      48.936   │
                          └─────────────────────────┘
```

q23 Father's occupation
Adj. P-value=0.000, F=176.209,
df1=3, df2=15912

| <= 1.0 | (1.0, 3.0] | (3.0, 4.0] | > 4.0 |

Node 1		Node 2		Node 3		Node 4	
Mean	45.381	Mean	52.429	Mean	48.999	Mean	54.022
Std. Dev.	18.812	Std. Dev.	18.734	Std. Dev.	19.950	Std. Dev.	17.649
n	7074	n	5131	n	2319	n	1392
%	44.4	%	32.2	%	14.6	%	8.7
Predicted	45.381	Predicted	52.429	Predicted	48.999	Predicted	54.022

Figure 3.2.21 DTA of father's occupation and students' achievement in Nepali

When combining the mother's and father's occupation, low achievement is found in the families where either the father or both parents come from an agricultural background (46%). The highest achieving students come from the families where the father comes from teaching or services and the mother from "other" (59); or the father comes from the "other" occupation and the mother has any other occupation except agriculture (56).

For the later use as a SES-indicator, the cut-off for the 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 may 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 the 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, poetry books, or artistic things like pictures, and books that help them for study such as a 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, computers, cars, and bathrooms.

There are 12 questions in the student background questionnaire related to home possessions. Each was scored 1 if the student had an access to this possession (e.g. having a separate room or a table to study). Adding these items up, the maximum score was 12 indicating that the student expressed to have access to *all* of the possessions and the lower the score the less possessions they have at home. Figure 3.2.22 shows the connection of home possessions and achievement level: Except the highest category, the achievement level of the students' raises logically the more there is access to home possessions. Pearson product moment correlation coefficient between the achievement level and the factor ($r = 0.09$) is statistically significant ($p < 0.001$) though the value is small.
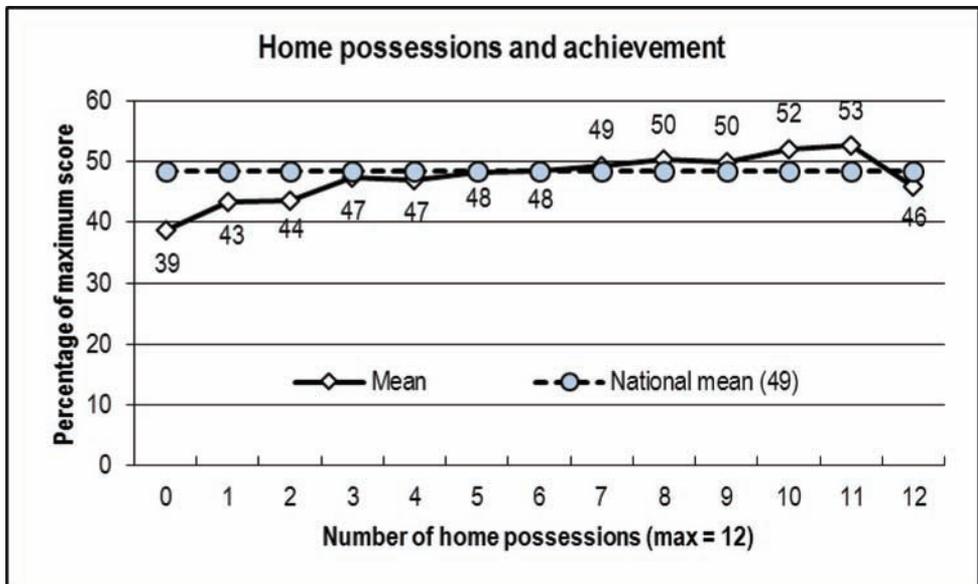
Figure 3.2.22 Connection of home possessions and achievement in Nepali

For the later use in SES, the cut-off for the factors was set on 5 possessions: if 5 items or more 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 3.2.23. 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). Results are similar in other accessories

except the computer i.e., when the number of computer increases from 1, correlation is negative however there is no change in the total result when the computer is excluded from the analysis. Therefore, the availability of the home accessories is dichotomized in the same way. After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level, see Table 3.2.21), all the five indicators were summed. The maximum score was 5 indicating that the student's home possessed a set number of *all* of the accessories.[15]

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
Car	1–3	0, missing
Bathroom	2, 3	0, 1, missing
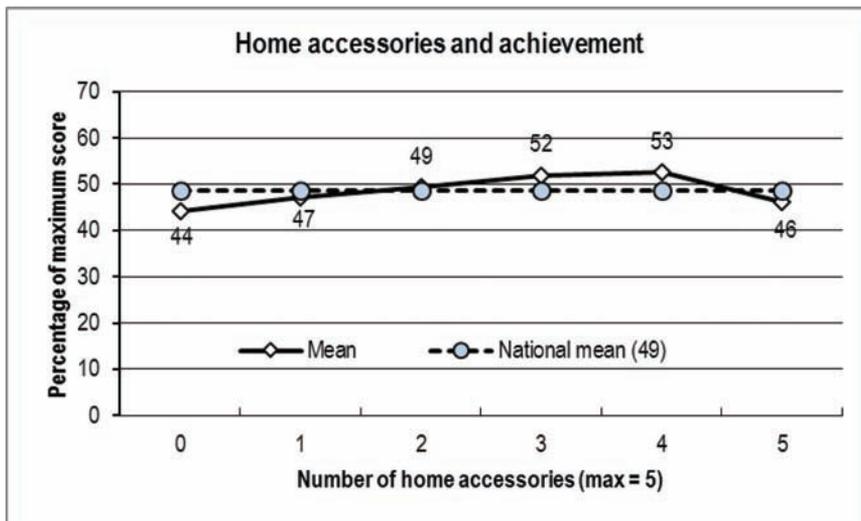
Figure 3.2.23 Connection of number of home accessories and achievement in Nepali.

---

[15] The analysis is bound the fact that the values were given by the students – they are, in many cases, credible. However, as with the home possessions (see Figure 3.2.22), 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. Both possessions and accessories show that the achievement level is quite low in this group.

Figure 3.2.23 clarifies how the number of home possession or accessories increases, students' achievement increases from 44 (if none of them are available) to 53 (if any four of them are available) however when all five are available, the achievement decreases. Availability of all the stated facilities indicates the high SES of the family (see also the footnote 3). Correlation between home accessories and achievement is $r = 0.04$ ($p < 0.001$) which is certainly positive though very small.

Dataset gives a strong signal that the educational level of the parents predicts the children's future achievement level in Nepali. Especially harmful for the achievement level seems to be the situation where the father is illiterate.

Data shows that when children have very few home possessions – zero to two out of the 12 – the achievement level is statistically lower (<43) than if there are more than two (> 47). With ten to eleven possessions, the average score is very high (> 52%) compared with the national average. The same is true of home accessories: When none or only one accessory indicator out of five is met, the results are lower than average (44–47%) and when there are two or more met, the results are remarkably higher (> 49%).With three or four indicators met, the results are the best (52–53%).

The dataset gives a strong signal that either economic- or intellectual capacity or both at home helps children to increase their educational standard. 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 other occupational groups. In the next assessment, it should be identified that the' other' category significantly influences achievement.

3.2.3.3.2 SES and achievement

The socioeconomic status was formed on the basis of seven indicators which were all first dichotomized (see the opening section of the Section 3.2.3.3.1). 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 (as SES) and changed into the percentage of the maximum score (P_SES). Deeper description of the transformations is seen in Section 2.7.2. 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.
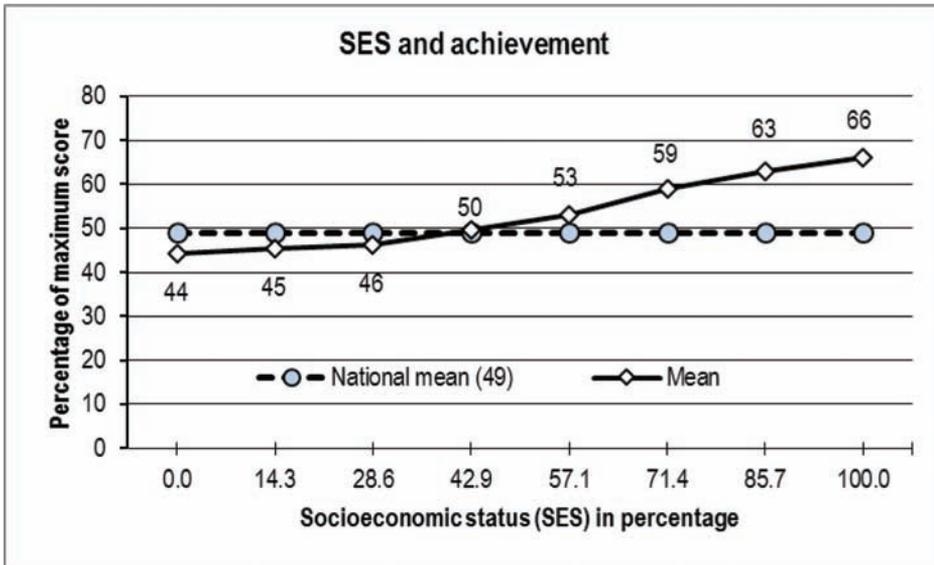
Figure 3.2.24 Connection of SES and achievement in Nepali

Figure 3.2.24 shows a strict positive relationship between SES and the Nepali achievement; the correlation between the variables is $r = 0.298$ which is a statistically significant association ($p < 0.001$). The differences between the SES groups are statistically significant (ANOVA, $p < 0.001$), the effect size is moderate or high ($f = 0.28$); that is, the highest and lowest group differ from each other moderately. SES explains somewhat 7% of the student variation ($\eta^2 = 0.072$) which is not, though, very high a percentage.

The dataset gives a strong signal that the socioeconomic status plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (22 percent points). This means that if the social economic standard of the lowest performing students is raised into a decent level, that is, in practice, that the problems in four out of seven indicators would be solved, the results in these groups would raise remarkably. Especially challenging is the situation in the families where the father or both parents are illiterate or they both work in agriculture.

3.2.3.3.3 Working after school and achievement

Several questions were set in the student background questionnaire of the students' activities outside the school. Two of them are briefly handled here: Working after

the school in a paid capacity and Participating in household work/chores. The values of the variables are divided into five categories: 0 (no paid work 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 GLM Univariate (regression modeling) indicates that cut-off is on whether the students work in a paid capacity or not. The relationship is strictly negative when students need to be engaged in paid work before and after school (Fig. 3.2.25).
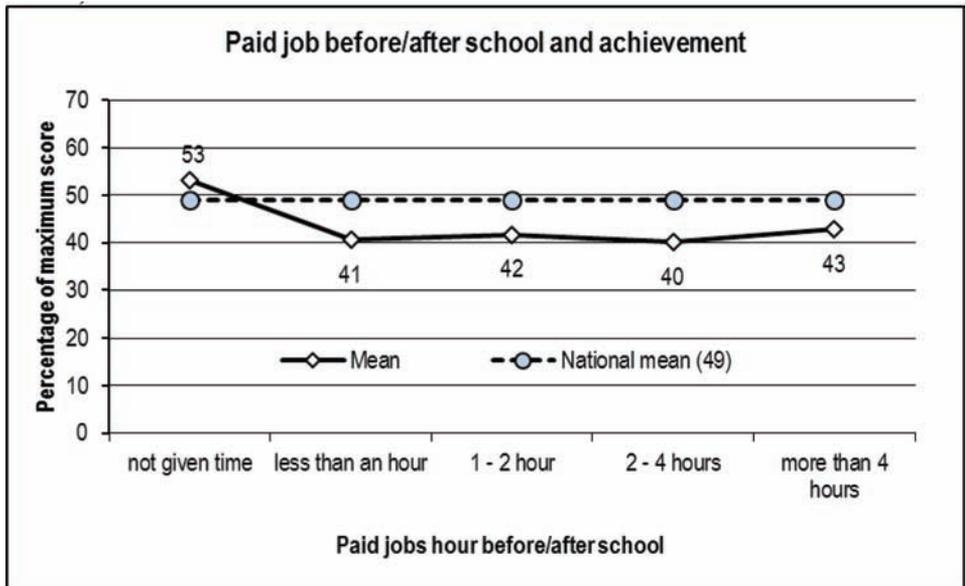
Figure 3.2.25 Relationship between achievements with paid job before and after school

The DTA shows that when the children have no paid work *at all*, the results are notably above the national average (53) (Fig. 3.2.26). *If the students are working on a paid basis – even less than one hour, the results are remarkably poorer than the average (41)*. The differences are statistically significant ($p < 0.001$) though the effect size is moderate – most of the children do not need to be engaged in paid work. Working after school indicates that the family is poor and the extra salaries are needed. It is obvious that when the student needs to work more than 4 hours per day there is no time or energy to handle school homework.

P_Eqd_Total=Eqd_Total/94*100
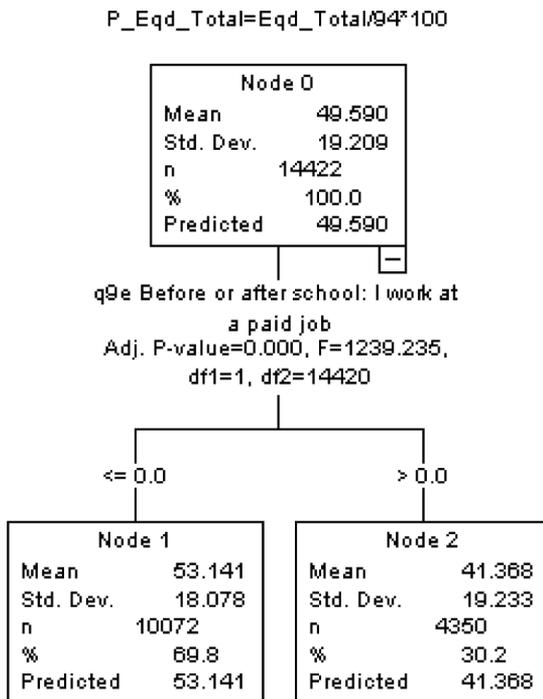
Figure 3.2.26 DTA of paid work and achievement in Nepali

It is usual – and a supported – practice in families that the children take part in household work at home; this is part of the socializing process of the children. The DTA shows that when the child needs to spend more than 4 hours per day doing household work, the results are almost as poor (43) as if the student would need to work in a paid capacity (Figure 3.2.27). However, when the amount of household work is decent – practically less than two hours per day – the achievement level is higher than the average (51). Differences are significant ($p < 0.001$) though the effect size is small or moderate ($f = 0.15$). It is somewhat interesting that more than 10% of the Nepali students (n = 1,691) inform us that they spend more than 4 hours per day doing household work. In the rural area, this may be obligatory when stock raising and when the cattle if far from the home. It is understandable, that in these cases there is not much energy to decently concentrate on their school work.

P_Eqd_Total=Eqd_Total/94*100

```
                        ┌─────────────────────────┐
                        │        Node 0           │
                        │ Mean         49.057     │
                        │ Std. Dev.    19.261     │
                        │ n            15245      │
                        │ %            100.0      │
                        │ Predicted    49.057   ─ │
                        └─────────────────────────┘
```

q9d Before or after school: I do jobs
at home
Adj. P-value=0.000, F=108.133,
df1=3, df2=15241

<= 0.0   (0.0, 2.0]   (2.0, 3.0]   > 3.0

```
┌────────────────────┐ ┌────────────────────┐ ┌────────────────────┐ ┌────────────────────┐
│      Node 1        │ │      Node 2        │ │      Node 3        │ │      Node 4        │
│ Mean      45.343   │ │ Mean      51.019   │ │ Mean      48.028   │ │ Mean      43.190   │
│ Std. Dev. 21.675   │ │ Std. Dev. 19.086   │ │ Std. Dev. 17.885   │ │ Std. Dev. 18.230   │
│ n         1512     │ │ n         9337     │ │ n         2705     │ │ n         1691     │
│ %         9.9      │ │ %         61.2     │ │ %         17.7     │ │ %         11.1     │
│ Predicted 45.343   │ │ Predicted 51.019   │ │ Predicted 48.028   │ │ Predicted 43.190   │
└────────────────────┘ └────────────────────┘ └────────────────────┘ └────────────────────┘
```

Figure 3.2.27 DTA of household work and achievement in Nepali

The dataset gives a strong signal that either working in a paid capacity or for four hours per day unpaid on 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 learning for the students in Nepali.

3.2.3.3.4 Attitude and achievement

In the context of the Nepali language achievement assessment, 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 connection is not always clear, the correlation between achievement and attitudes toward 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 attitudes and achievement may can be near zero, like in Macedonia (Kadijevich, 2008), the Philippines (Wilkins, 2004), 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, the same shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) as is in use in several international

comparisons like in TIMSS and PISA studies was in use. The original scales included nine dimensions but in these international comparisons only three is used with four items on each dimension and two negative items on each of the first two dimensions (see in detail in Section 2.4.2 and 4.5). The names of the factors can be "Liking Nepali", "Self-Efficacy in Nepali", and "Experiencing utility in Nepali" (compare naming in, e.g., Kadijevich, 2006; 2008). Factor analysis was used to identify the factors of the responses in FSAS and the negative items were reversed to make the whole test unidirectional. As in several countries in Asia, the expected factor structure cannot be found in Nepal (for a deconstruction of the test scales, see Metsämuuronen, 2012a; 2012b). Hence, only the total score is used to show the connection of attitudes and achievement. The relation between the attitude (divided into ten groups with somewhat an equal number of the students, that is, deciles) and achievement score is shown in Figure 3.2.28.
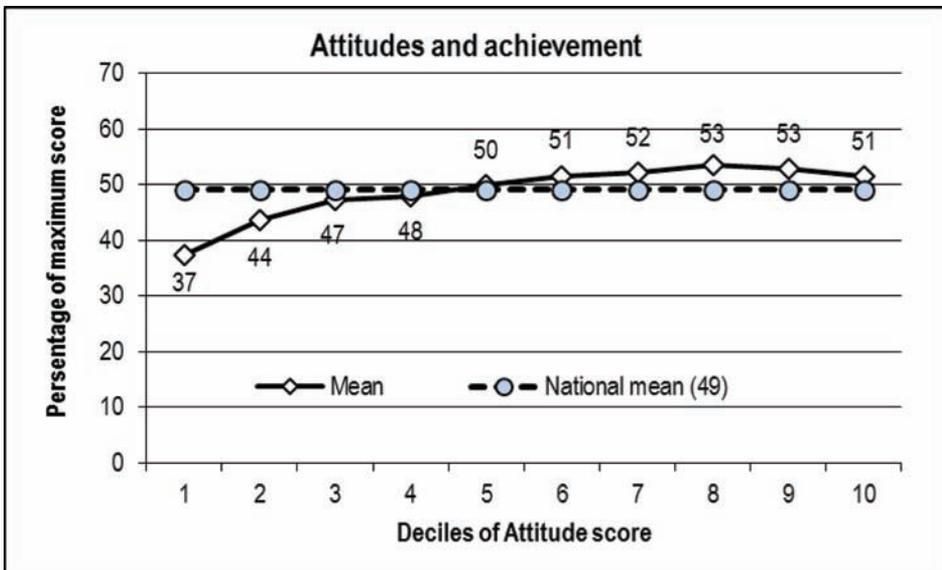
Figure 3.2.28 Connection of attitudes and achievement in Nepali

There is a clear positive correlation between the attitudes and Nepali achievement ($r = 0.20$). The connection is moderately high ($f = 0.25$); the division of attitudes to ten groups explains the achievement level somewhat 6% ($\eta^2 = 0.060$). The difference between the lowest and highest attitude group is 14 percent points. The connection is clear though it is not known whether the positive attitude is a consequence of high achievement or the other way round.

Data gives a signal that positive attitudes toward the subject correlate with positive achievement.

3.2.3.3.5 Age and student achievement

In the Nepalese context, the age of the students attending to grade 8 studies varies widely. Some students have given their age as below ten years and some above 20. All the ages of the students below 13 were encoded as 'up to 13 years', and all students above 19 were encoded as '19 years & above'. The descriptive statistics of the mean in each year are given in Table 3.2.22 and visualized in Figure 3.2.29.

Table 3.2.22 Descriptive statistics of the students' achievement in different age groups

Age	N	Mean	SD
up to 13 years	3310	54	19.0
14 years	5110	52	18.9
15 years	4401	46	18.9
16 years	2050	43	18.7
17 years	896	40	17.9
18 years	320	39	17.4
19 years & above	230	41	19.3
Total	16317	49	19.3
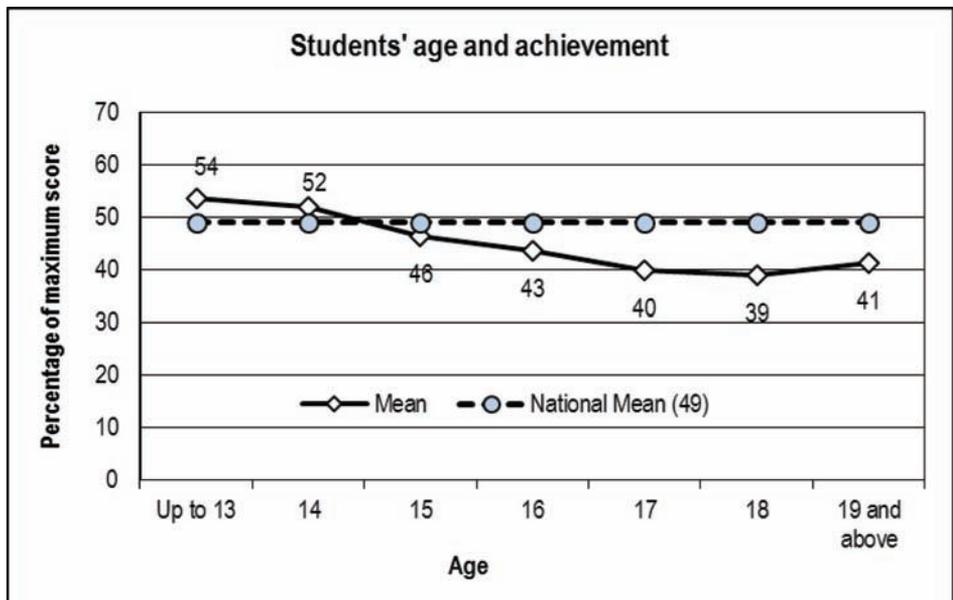
Figure 3.2.29 Connection of age and achievement in Nepali

157

It seems evident that the best achievers are those students who are at the proper age for grade 8 studies (13 to 14 years old, scoring 54 and 52 respectively). The higher the age is – meaning that the students have either started much later than they should have, or they have doubled the classes – the weaker the results are. The achievement level is remarkably lower than the average when the students are of age 16 or higher (39–43). Correlation between the variables is -0.22 ($p <$ 0.001) indicating moderate effect size (f = 0.23). The ANOVA hints that the age (that is, the prolonged studies) explains more the achievement level (14%) than the achievement level the prolonged studies (5%). Another side of the matter is that it is good that these "over-aged" students are at school to learn; although they should have been identified at a much earlier age for extra tuition or support.

Dataset gives a signal that the highest performance is with those students studying with their normal age group, that is, at the age of 13 and 14 years. Otherwise the achievement decreases as the age increases.

3.2.3.3.6 Help in study and student achievement

The relation between the help in 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 helpers are given in Table 3.2.23.

Table 3.2.23 Descriptive statistics of helpers of the students and related achievement level

Who helps	N	Mean	SD
Tuition	924	52	18.2
Mother	518	50	21.1
Brother/Sister	3139	50	19.1
Teacher	10334	48	18.9
No one	220	48	20.9
Father	884	47	21.4

It seems that help is necessary for the students to gain better than average marks on the test. There is about 4% score difference between those who don't get any kind of help and those who receive (private) tuition. Mother's help at home raises the achievement level (50.2), which is the same as the achievement of the students

having help from brothers and sisters (49.8). It is interesting to note that those who were helped by their fathers gained as low (47.1) as those who did not get any help (47.5). Those helped by the teacher represent the main data and hence their average is very close to the National average (48.4). The extra (private) tuition seems to bring 3 percent point advance compared with those helped by the teacher. However, the advance is not at all at the same level as when comparing the illiterate and literate families, for example. It is also unknown whether the time spent on the tuition has something to do with the matter?

The dataset gives a slight signal that the help given by the mother and by brothers and sisters raises the achievement level more than help is given by the father. The highest achieving group is the one who receive private tuition. It is possible that is group also spent more time on their homework, which may explain the high score.

3.2.3.3.7 Availability of textbook and student achievement

The data shows that there are still some students who don't have the Nepali textbook up to the last academic session. Table 3.2.24 shows the descriptive statistics of availability of the Nepali textbook and the achievement (mean).

Table 3.2.24 Availability of textbook of Nepali and the achievement (mean)

Do you have textbook of Nepali?	N	Mean	SD
Yes	15695	49	19.1
No	385	40	20.7
Total	16080	49	19.3

Out of 16,080 students who responded to the question, 2.4% don't have a textbook available at school. The relation between the textbook and achievement is significant ($p < 0.001$) though the effect size is small ($f = 0.07$) due to the small group size in one group. The difference in achievement is anyhow quite big (9% of maximum marks).

Data shows that 2.4% of the students lack the proper text book in Nepali. The achievement level of these students is significantly lower than those who have access to the text book.

3.2.3.3.8 Homework given/checked and achievement

Homework is one way to enhance teaching; it can be used as drill-, exercise-, and as an evaluation tool. When homework is systematically checked, it may boost

achievement levels. Now, the related results in NASA 2011are based on students' reports; it may be possible that within the same classroom some students have given a slightly deviant response than the other students from the same class. However, in the wide scope, the results are credible. Statistics related to homework given and checked is condensed in Table 3.2.25 and visualized in Figure 3.2.30.

Table 3.2.25 Homework given and the achievement

Status of homework	N	Mean	SD
Not given	121	35	19.8
Given Some day - not checked	114	39	19.0
Given Every day - not checked	233	38	19.4
Given Some day - Checked some day	1515	48	19.3
Given Some day - Checked every day	1369	48	18.8
Given Every day - Checked some day	1801	49	19.6
Given Every day - Checked every day	10701	50	18.9
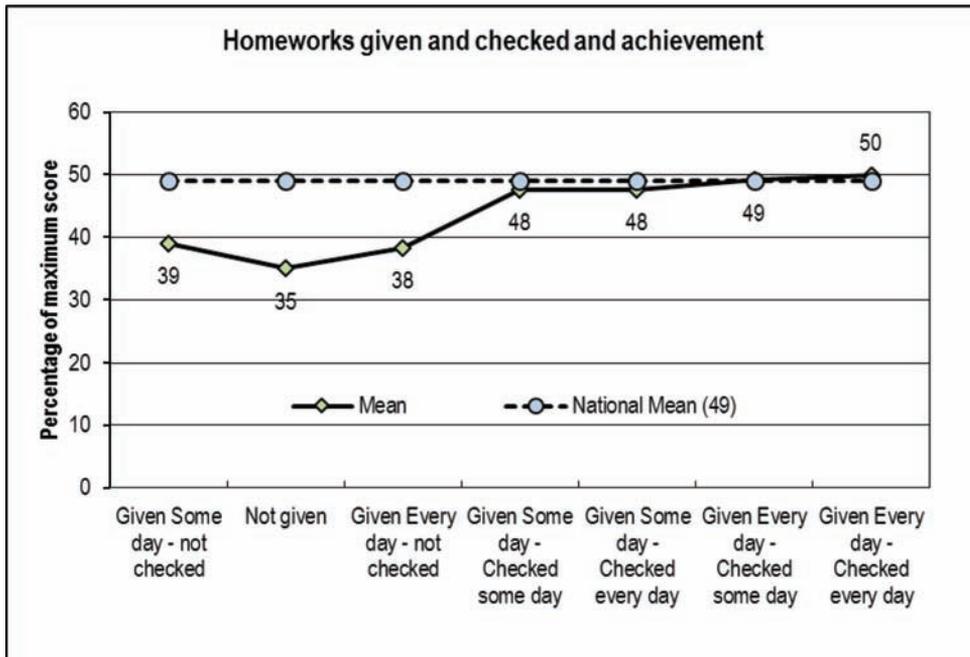
Figure 3.2.30 Connection of homework given and checked and achievement in Nepali

160

It is evident, on the basis of the dataset, that if the teachers give homework every day to the students and it is checked – even only now and then, student achievement is higher (50) than when homework is issued non-systematically (48) or homework is not given or they are not checked at all (35–39). The differences are statistically significant ($p < 0.001$). Those groups with no checking or homework are, however, very small and hence, the effect size is small ($f = 0.09$); grouping explains only 0.8% of the variance in the data ($\eta^2 = 0.008$).

Dataset gives a solid signal that if the teacher gives the homework daily and checks it, the achievement is higher than without checking or issuing of homework. By giving daily homework daily and by checking it, even not every day, the teacher may have raised the scores up to 15 percent points.

3.2.3.3.9 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 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' 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 learning environment for the students. International Studies like TIMSS and PISA give a specific emphasis to identify such indicators. In the NASA 2011 student questionnaire, five questions indicate the varieties of bullying that may 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 condensed in Table 3.2.26 and visualized in Figure 3.2.31. '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 the last month. Alone, the fact that 22% of the students mention that, during the last month, something of their own was stolen is an alarming sign of the system.

Table 3.2.26 Bullying and the achievement

Type of Bullying	No (%)	Yes (%)
Something of mine was stolen	75.9	21.9
I was made fun of or called names	83.5	12.6
I was hit or hurt by other student(s)	82.4	14.1
I was made to do things I didn't want to do by other students	87.3	9.0
Fellow students kept outside without involving me in activities	88.8	7.9
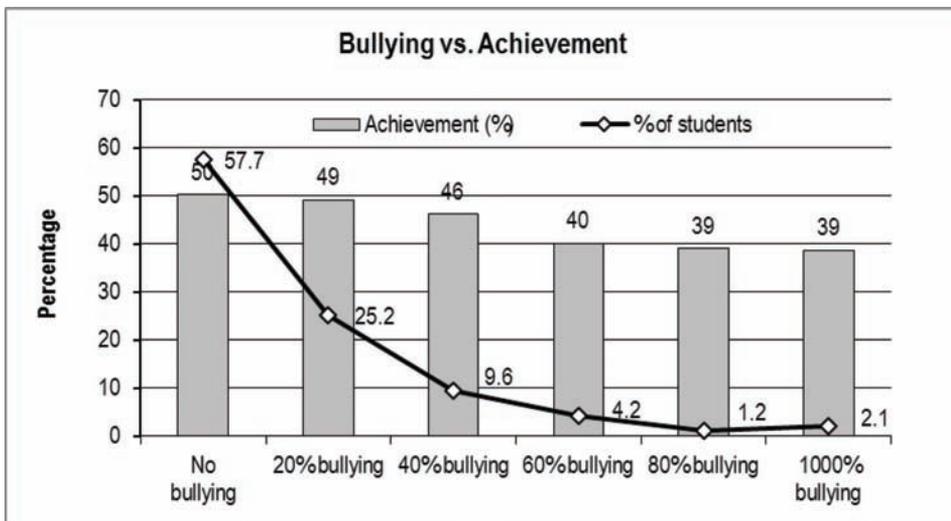
Figure 3.2.31 Relation of bullying and achievement in Nepali

The sum of all five items is done as an indicator of bullying. Figure 3.2.31 shows the extent of bullying with the percentage of the students and achievement of the students in each category of bullying is shown in the line graph as bars. 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 57.7% of the students did not encounter any bullying during the last month, one can infer that the remaining 42.3% *did* encounter at least one type of bullying. This is a remarkable number of the students. About 3.3% of the students are experiencing a severe kind of bullying (the sum of 80% and 100% bullying). It seems that learning outcomes are remarkably lower with that 7.6% of the students who have encountered more than two different types of bullying (39–40). Students

162

who don't experience bullying and students who encountered extremely bullying of four or five kinds have 11 percent points' achievement gap; though there are a few number of the students who reported this kind of bullying (n = 512). The difference is statistically significant ($p = 0.001$) though the effect size is small or medium ($f = 0.17$). 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 the schools.

The dataset gives a signal that an alarmingly high number of the students (42%) have encountered bullying in schools within the last month. The phenomenon seems to affect the learning outcomes in almost all the groups of the students who felt bullying, all possible efforts have to be done to root the phenomenon from the schools.

Positive activities in school

The activities that can boost the learning and achievement of the students are categorized as positive activities. Such positive activities about the school were asked from the students in two sets of questions collected in Table 3.2.27. The table shows the responses of the students in all four categories; the responses are in the 4-point rating scale anchored to fully agree and fully disagree.

Table 3.2.27 Students' response towards teacher- and school-related activities in the schools

Teacher and Students activities	Respondents in % (valid percentage)			
	Fully agree	Partially agree	Partially disagree	Fully disagree
q27a School: Students get along well with most teachers	78.2	18.2	2.4	1.2
q27b School: Most teachers are interested in student's well-being	86.7	9.6	2.1	1.6
q27c School: Most of the teachers really listen to what I have to say	57.5	33.6	6.6	2.3
q27d School: If I need extra help. I will receive it from my teacher	77.7	17.9	2.6	1.8
q27e School: Most of my teachers treat me fairly	57.4	22.2	8.3	12.0
q28a School: I like come and stay in school	90.1	6.7	1.6	1.6
q28b School: Students in my school try to do their best	76.6	19.5	2.7	1.2
q28c School: Teacher in the school care about the students	81.2	15.0	2.6	1.2
q28d School: Teacher wants the students to do their best	91.0	6.2	1.3	1.5
Average	77.4	16.6	3.4	2.7

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 4 attitude groups in the indicator. These boundaries and descriptive statistics are seen in Table 3.2.28 and illustrated in Figure 3.2.32. The overall result is that the feeling of the positive actions in the school relates positively with the student achievement. The correlation between the sum of nine positive activities and achievement is positive ($r = 0.11, p < 0.001$).

Table 3.2.28 Students' response towards teacher- and school-related activities in the schools

percentage of positive action	N	Mean	SD
1 (-66.66)	1191	39	20.6
2 (<77.77)	1125	44	20.1
3 (<88.88)	3810	48	18.6
4 (>=88.88)	10112	50.5	18.8
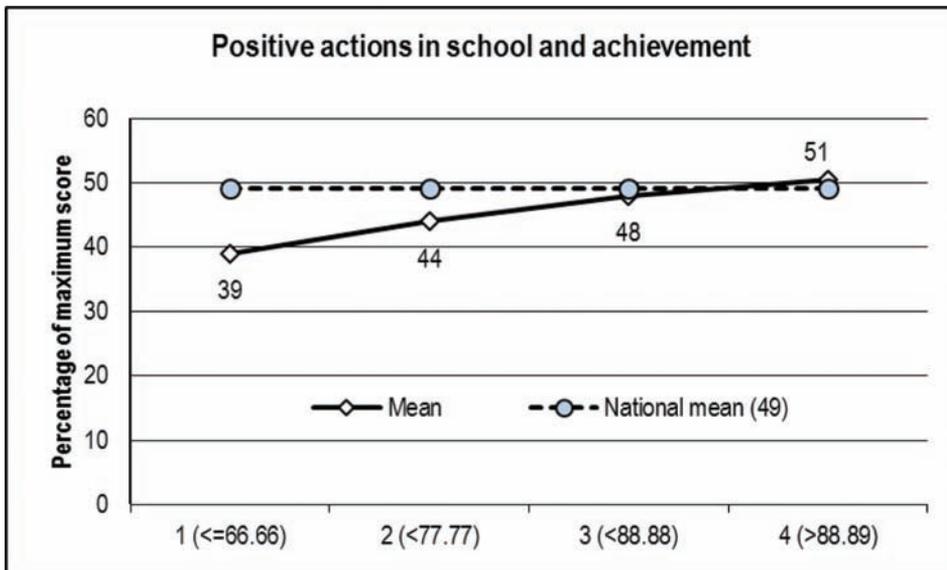
Figure 3.2.32 Relation of positive actions in school and achievement in Nepali.

164

The data shows that there is a positive connection between the students' feeling of the school activities and the achievement. The increase in achievement is directly proportional to the increase in intensity of such activities. After dividing the indicator into six four groups on the basis of DTA, the differences between the groups are statistically significant ($p < 0.001$), however, the effect size is small ($f = 0.14$). Only when the students are extremely positive towards school and teachers' behavior, is the learning achievement much higher than the average (51). However, the difference between the most positive group and the most negative group is notable (12%).

Dataset gives a signal that when the students think that the actions of the teachers and the schools are ultimately good the results are better than average (51). At the other extreme of feeling that such actions are ultimately negative, the results are far below the average (39).

3.2.4 Conclusions

The main findings of the Nepali achievement at the grade 8 are condensed as follows:

Basic results

- The Nepali proficiency in grade 8 is normally distributed.

- The learning outcomes are the highest in the content area of Reading and weakest in Vocabulary.

- Students' ability to solve complex problems is quite low; only 42% of the maximum scores were reached. Students are much better in the recalling type of questions (74%), which may be explained by the educational system, which seems to be geared more of remembering the things than solving novel problems.

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

- The Nepali language skills have been developed in a positive way. The

Nepali results rose 2 percent points within four years. Especially, the level in Writing (+16) and Reading (+6) rose but it dropped dramatically in Vocabulary (-22).

- From the international comparison viewpoint, the average reading proficiency in Nepal is much lower than the international average.

Equity indicators

- 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 in Jajarkot (38), Morang (39), Bara (39), Bhojpur (40), Banke (42), Doti (42), Kanchanpur (42), and Ilam (42) are so low that raising the standard in these districts would raise the standard in the whole country.

- There is a moderate difference between the student performances in four Ecological zones. Students in the Kathmandu Valley outperform the other students. The achievement is the lowest in Tarai area.

- There is a waste inequity in the Developmental regions in children's opportunities to reach an adequate level of Nepali. The difference between the lowest performing area (Far-western, 42) and the highest performing region (Kathmandu Valley, 64) is remarkable – more than 20 percent points.

- On average, the students in the institutional schools outperform the students in the community schools. This deviance may be explained by the student selection though this cannot be known on the basis of the dataset.

- The students in the urban schools would gain 9 percent point more than the students in the rural areas. When excluding Valley, the difference is not remarkable. From the educational equity viewpoint, this is not, though, a good sign.

- There is an educational inequity within the language groups. The students from Newar- (59) and Tamang (55) background perform higher than Nepali students (51). On the other hand, the students from Limbu- (40), Tharu- (40), "other"- (41), and Rai (42) background perform much lower than the national average. Ecological zone-wise analysis raises an important question: why are the results in Nepali so low in Tarai region and what can be done to raise the standard?

- The Madhesi students' performance is radically lower than the results of the other castes. In particular, their performance is low in the Western-

and Eastern Developmental regions. A notable percentage of Madhesi students have not reached the required level of reading and writing Nepali for being able to continue in the higher education in Nepali language.

- The girls are slightly out-performing boys in writing and in the total score. Differences are not wide though they are significant. Sex explains only 1% of the student variation; from the equity point of view the signal is positive thought there is still work to do to reduce the gap. There are not notable differences between sexes in Ecological zones or Developmental regions.

Selected explanatory factors

- The parents' educational level strongly predicts the children's future achievement level in Nepali. Especially harmful for the achievement level seems to be the situation where the father is illiterate.

- Either the economic- or intellectual capacity or both at home helps children to raise their educational standard. If the father or mother or both are coming from the agricultural or related occupation, the students' achievement in Nepali is significantly lower than with the other occupational groups.

- When the children have only very few home possessions (like their own table or room or dictionary) – 0 to 2 out of the 12 – the achievement level is statistically lower (<43) than if there are more than two (> 47). With ten to eleven possessions, the average score is very high (> 52%) compared with the national average. The result is the same with home accessories (like mobiles, cars, bathrooms): With less than two accessory indicator out of five is met, the results are lower than average (44–47%) and when there are two or more met, the results are remarkably higher (> 49%). With three or four indicators met, the results are the best (52–53%).

- The socioeconomic status (SES) plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (22 percent points). This means that if the social economic standard of the lowest performing students can be raised to a decent level, the results in these groups would raise remarkably. Especially challenging is the situation in the families where the father or both parents are illiterate or they both work in the agricultural field.

- Either paid work or working for four hours per day unpaid on household work outside of school effectively reduces the school achievement of the student. However, a decent amount of household work up to two hours per day supports the learning for the students in Nepali.

- The positive attitude toward the subject correlates with positive achievement.

- Achievement decreases as the age increases. The highest performance is with those students studying within their normal age group, that is, at the age of 13 and 14 years.

- Tuition given by the mother, brother or sister raises the achievement level more than when given by the father. The highest achieving group are those who receive private tuition. It is possible that those with the private tuition also spent more time on their homework, which may explain the high score.

- 2.4% of the students lack the proper text book in Nepali. The achievement level of these students is significantly lower than those who have the access to the text book.

- If the teacher gives the homework daily and checks it, the achievement is higher than if this does not occur. By giving daily homework and by checking it, even not every day, the teacher may have raised the scores up to 15 percent point.

- An alarmingly high number of the students (42%) have encountered bullying in schools within the last month. The phenomenon seems to affect the learning outcomes in almost all the groups of the students who experienced bullying. All possible effort has to be done to root the phenomenon from the schools.

- When the students think that the teachers and schools actions are ultimately good, the results are better than average (51). At the other extreme, if feeling is ultimately negative, the results are far below the average (38).

References for Section 3.2

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

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

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

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

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

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

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

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

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

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

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

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

Fullbright (2008). *A final report on national assessment of grade-V students*, Fulbright Consultancy & Cooperative Hands in restoration (CHIRAG). Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education.

House, J. D. &Telese, J. A. (2008). Relationships between Student and Instructional Factors and Algebra Achievement of the students in the United States and Japan: An Analysis of TIMSS 2003 Data. *Educational Research and Evaluation*, **14**(1), 101−112.

Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on Mathematics attitude.*The Teaching of Mathematics*, **9**(2), 41−51.Retrieved from http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf.

Kadijevich, Dj. (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.

Metfessel, N., Michael, W. B., &Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the Writing of behavioral Objectives. *Psychology in the Schools*, **6**, 227−231

Metsämuuronen J (2012b). Comparison of Mental Structures of Eighth-Graders in Different Countries on the basis of Fennema-Sherman test*.* **International Journal of Psychological Studies,** 4(4) December.

Metsämuuronen J (2012a). *Challenges of the Fennema-Sherman Test in the International Comparisons.* **International Journal of Psychological Studies,** 4(3) September.

Papanastasiou, C. (2000). Effects of Attitudes and Beliefs on Mathematics Achievement. Studies in Educational Evaluation, **26**(1), 27−42.

Papanastasiou, E. (2002). Factors That Differentiate Mathematics Students in Cyprus, Hong Kong, and the USA. *Educational Research and Evaluation: An International Journal on Theory and Practice,* **8**(1), 129−146.

PEDP (1998). Primary Education Development Project: An assessment. 1998. Kathmandu.

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.

TIMSS (2009). TIMSS 2007 User Guide for the international Database. Released Items, Mathematics – Eight grade. IEA, TIMSS &Pirls International Study Center. Lynch School of Education, Boston College.

Wilkins, J. L. M. (2004). Mathematics and science Self-Efficacy: An international investigation. *The Journal of Experimental Education*, **72**(4), 331−346.

3.3 Achievement in Social Studies in NASA 2011

Ms. Sunita Shakya[1],
Dr. Jari Metsämuuronen[2] &
Prof. Dr. Sriram P. Upadhyaya[3]

1. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal

2. Helsinki University, Faculty of Behavioral Sciences, Finland

3. Tribhuwan University, Kathmandu, Nepal

Abstract

Achievement in Social Studies of 8th graders was assessed by utilizing student achievement tests and the related background questionnaires of 16,299 students from 413 randomly selected schools from 25 randomly selected districts. The schools represented all Ecological zones and Developmental regions, rural- and urban areas as well as community- and institutional schools.

Achievement in Social Studies is normally distributed. The learning outcomes are notably better than average in Economics (58) and Geography (57) and weaker than the average in politics (46). Students' ability to solve complex problems is low; only 34% of the maximum scores were reached. Students are good in recognizing the correct answer and in very fundamental knowledge/contents, true and false, matching texts, puzzles, selection of words and fill in the gaps. They are much weaker in reasoning, problem solving, and constructing the information. From the international comparison viewpoint, the average achievement of Geography in Nepal is better than the international average.

From the equity viewpoint, there are many good signs. With the exception of the Valley region, there is no difference between the Ecological zones. There are no practical differences between the developmental regions (46–49). Though the results are somewhat better in cities than in rural areas, the difference is not remarkably high. The difference between the boys and girls is small. On the other hand, students in the institutional schools perform very well and the students in the community schools form two kinds groups of schools: high-performing schools and low-performing schools. There are also remarkable differences between the language groups

in the achievement of Social Studies. In Social Studies, Madhesi- (46), Dalit- (47), and Alpasankhyak (47) students' performance is lower than the results of the other castes;

The educational level of the parents highly predicts the children's future achievement level in Social Studies. Especially challenging is the situation when either the father or both parents are illiterate. If the father or mother or both were coming from an agricultural or related occupation, the students' achievement in Social Studies is significantly lower than with the other occupational groups. The socioeconomic status (SES) as a whole plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (23 percent points). Working outside of school effectively reduces the school achievement of the student. 3.8% of the students lack the proper text book in Social Studies. An alarmingly high number of the students (43.4%) have encountered bullying in schools within the last month.

3.3.1 Introduction

Social Studies as a school subject – contrary to Mathematics and Nepali – is assessed less systematically and less frequently in the national assessments of student achievement (NASA) in Nepal. Unfortunately, the results of the previous national assessments (see BPEP, 1997; EDSC, 1997; BPEP, 1998; PEDP, 1998; EDSC, 1999; CERID, 1999; EDSC, 2001; 2003; CERSOD, 2001; EDSC, 2008; Fulbright, 2008) are not comparable with each other because of the missing linking procedure between the tests. Of the previous datasets, like in Mathematics and Nepali, only the year 2008 (EDSC, 2008) could be used as a basis for comparison because it is the one where an item-wise data was found. Hence, that dataset is used as a basis for the comparison of 2011 results with the previous one. Several items from the 2008 test were used as the linking items between the years 2008 and 2011. In theory, with these linking items it is possible to assess what kind of changes has happened between the two measurement points. However, showed up that although, like in Mathematics and Nepali, some items from the year 2008 were used as linking items, due to an incomparable marking schemes those items were discarded from the analysis of Social Studies (see in detail Sections 2.2.5 and 3.3.2).

Another linking procedure was administered to link the results in Nepal to the international Trends in Mathematics and Science Study (TIMSS) data. Several published items from the international bank of Geography were borrowed in order to compare the results in Nepal with an international standard (see in detail Section 2.4). By using the Item Response Theory (IRT) modeling, the 2008 dataset, TIMSS datasets, and the 2011 NASA dataset were linked together to give unique new information on the achievement level in Social Studies. However, only the TIMSS comparison is reported.

In what follows, Section 3.3.2 describes the specific methodological solutions in the Social Studies testing – more detailed information on the process is found in Section 2. Section 3.3.3 reveals the basic results in the different content areas of Social Studies in general (Section 3.3.3.1), this deepens to the effects of different equity matters from the diversity view point (Section 3.3.3.2) and gives some hints on the deeper factors explaining the differences in the Social Studies achievement (Section 3.3.3.3). Finally, the results are condensed in Section 3.3.4.

3.3.2 Methodological solutions

In the process of testing the Social Studies achievement of 8th graders, 16,299 students, 401 schools, 402 head teachers and 402 teachers from 25 districts participated in the survey and testing (for more details of procedures and sampling, see Chapter 2.1). Three test versions (in what follows, S1, S2, and S3) were administered. S1 and S2 were administered in twenty-three districts so that both versions were administered simultaneously in the same classroom. S3 was administered in two Himalayan districts, Mustang and Kalikot, two months earlier in a way that all other subjects (Mathematics and Nepali) were also administered simultaneously in the same classroom. Hence, more schools were encountered from the Himalaya region though the number of the students was proportionally fewer than in the Hill and Tarai regions.

3.3.2.1 Item analysis and characteristics of the test versions

NASA 2011 is the methodological shift in the National Assessment of Student Achievement. The modern test theory, that is, IRT modeling was used from the beginning of the item construction and preparing the marking scheme for the analysis of the data. In practice, this means, first, certain restrictions on the marking of the papers: no decimal number scores are allowed in the assessment during the answer sheet marking. If the students are not qualified to secure full score, 0.5 score is not provided in any case i.e., students' responses are marked in whole numbers. Obviously, the marking scheme in 2008 differed from this logic. Hence the 2008 data was adjusted to meet the requirement of IRT modeling: half points were lowered to the *lower* full mark because the students did not show enough achievement to gain the upper full mark. Secondly, the marking scheme has to be more rigorously prepared because of the need to make the exact the same judgments in the years to come with the linking items. Thirdly, IRT modeling requires that all the possible marks have to be observed in the dataset; this made some difficulties when analyzing some of the productive items because of the markers' tendency *not* to give full marks to the students. This was not a severe problem in

173

Social Studies compared with the Nepali subject (see Section 3.2). A more severe challenge was the different marking schemes of 2008 and 2011 markers (see Section 3.3.3.1.5). Finally, IRT modeling requires a linking procedure between the different versions in the test. Hence, the common items for each test version, the linking items, were carefully selected from the 2008 test booklet, TIMSS dataset, as well as in the pretest items banked for the item selection. Classical item- and test analysis methods were used in the pretest phase for finding the percentage of correct answers, item discrimination power, and the test reliability whereas IRT was used for item calibration, finding the latent ability (Theta, θ), comparing and equating the versions S1 – S2, the dataset from the year 2008, TIMSS database. SPSS software was used for classical Analysis and One Parametric Logistic Model (OPLM) software was used for the IRT modeling.

On the basis of the pretested items, the versions S1 and S2 were parallel whereas S3 was shorter and easier than the other two. S3 was administered earlier than the other two; because of the willingness to avoid any leaking of the items. All the versions were linked with each other by the use of the identical linking items. The longer versions S1 and S2 were scored out of a maximum of 74 and 76 maximum marks and S3 has a maximum score of 63 marks. There were 8 TIMSS items as the linking items all over three versions. The parameters of the international items were fixed during the item calibration so that all the test items of the year 2011 and 2008 were calibrated in the international TIMSS scale. After the calibration of the items, all the scores in the versions S1–S3 and the version of the year 2008 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 different difficulty levels of the tests have been modeled by using IRT modeling. The original output is the latent ability (Theta, θ) which is a standardized normal score ranging usually from -4 to +4. These values in each test versions were later transformed to equated scores and further, the equated scores were converted into the percentage of maximum score so that the score 100 means that the student made a perfect score and 0 means that no items were successfully answered. Now onwards, marks or average or mean score refers to the percentage of the maximum marks ranging from 0 to 100.

Table 3.3.1 shows the average marks (mean) calculated in three versions S1, S2 and S3 separately.

Table 3.3.1 Comparison of the characteristics of Social Studies test versions

Version	N	Mean1	SD	Remarks
S1	8066	51	14.5	Administered in 23 districts, in the same classroom
S2	8007	48	16.1	
S3	226	45	13.0	Administered in Mustang and Kalikot districts
Total	16,299	49	15.3	Social Studies as a whole (S1, S2 & S3)

1) percentage of the equated maximum score rounded in whole number

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 items were classified into five categories: *Geography, Civic, Economic, History* and *Politics*. The number of items asked and the weighting of the items is based on the recommendations of the curriculum for grade 8. Overall internal consistencies (given by reliability) of the whole tests on each version were very high ($\alpha = 0.86$–0.88), however, some of the categories (mainly Politics) contain a few items and hence the reliability is low. The overall summary of the content wise test analysis is given in Table 3.3.2. The reliability of the score in the total sample cannot be given in a classical way because it can only be estimated version-wise.

Table 3.3.2 Construct validities and reliabilities of the scores

Topics	Marks			Percentages			Percentage in Curriculum	Reliability		
	M1	M2	M3	M1	M2	M3		V1	V2	V3
Civic	39	38	36	45,3	44,2	44,4	45	0.71	0.76	0.71
Economics	18	17	17	20,9	19,8	21,0	20	0.51	0.59	0.54
Geography	11	13	11	12,8	15,1	13,6	13	0.41	0.57	0.33
History	9	9	10	10,5	10,5	12,3	12	0.51	0.40	0.61
Politics	9	9	8	10,5	10,5	9,9	10	0.40	0.28	0.40
Total	74	76	63					0.86	0.88	0.87

3.3.2.2 Analytical tools used in the analysis

Basic statistical methods are used throughout the report. The basic tools of statistical description (means, Standard Deviations, percentages, and frequencies), correlations (Pearson's product moment correlation coefficient and Spearman's rank order correlation), and comparison of two means (t-test) as well as statistical inference

(p-values, effect sizes) are used when appropriate. These methods are described in all basic books of statistical description and -inference. The Analysis of Variance (ANOVA) is used in the General Linear Modeling (GLM) way when several means are compared. All the p-values are corrected by using Multilevel Modeling (or Hierarchical Linear Modeling) by using SPSS Linear Mixed models module. Decision Tree Analysis (DTA) is used in finding the cut-offs of the continuous variables. All analyses are done in SPSS20 environment.

3.3.2.3 Some statistical concepts used in the text

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

When the sample size is huge – as the sample of 16,000 students is – the p-value very easily gives a signal that the difference between the groups is real in population. P-value does not, however, tell whether the difference is small or big. For this purpose there is another statistic, Effect Size (ES). *Effect size* tells how far the lowest and highest groups come from each other. Commonly used indicators of ES are Cohen's d used for two means and Cohen's f for several means (Cohen 1988). Cohen has given boundaries for small, medium and large effect sizes. During the text, these boundaries are used as a "measurement stick" to indicate whether the difference is small, medium/moderate or high. The rough boundaries of the small, medium, and high effect sizes are collected in Table 3.3.3.

Table 3.3.3 Rough boundaries of effect sizes

ES	Cohen's d	Cohen's f
small	0.2 or below	0.1 or lower
medium	round 0.45	round 0.2
high	0.8 or higher	0.4 or higher

Hence, one may read that "the difference between boys and girls is statistically significant ($p < 0.001$) but the effect size is small" in the text. This means that first, the difference between the boys and girls is real, but second, the difference is, in reality, very small or not notable.

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

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

Hence, if $\eta^2 = 0.30$, then $f = \sqrt{(0.3/0.7)} = \sqrt{0.43} = 0.65$ showing very high effect size.

3.3.3 Results

3.3.3.1 Basic results

3.3.3.1.1 Population Analysis

When assessing the student achievement, the population is usually normally distributed in a large sample. Then, 50% of the students lie below average and 50% of the students lie above the average. This kind of distribution can be found in Social Studies (Figure 3.3.1).
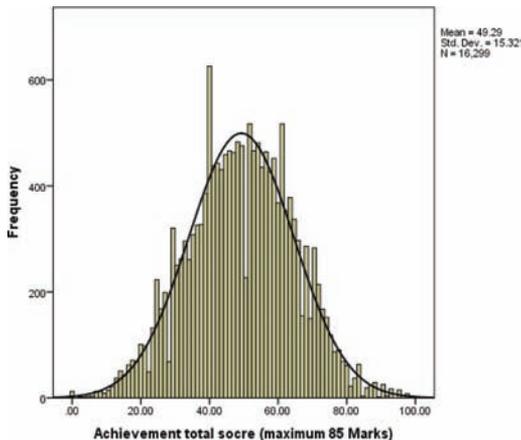
Figure 3.3.1 Normally distributed student population of Social Studies test

177

The Social Studies sample was big enough to form a normal distribution (16,299 students). Most of the students have achievement score in between 30–70%. It is notable, that, in contrast to Mathematics and the Nepali language, also the schools' means follow the normal distribution (Figure 3.3.2). In Mathematics and Nepali, the corresponding graphs show that the populations are obviously *not* normally populated. In Mathematics, there seems to be three populations: low-performing schools, medium-performing schools, and high-performing schools and, in Nepali, two populations. In Social Studies, there is only one population.
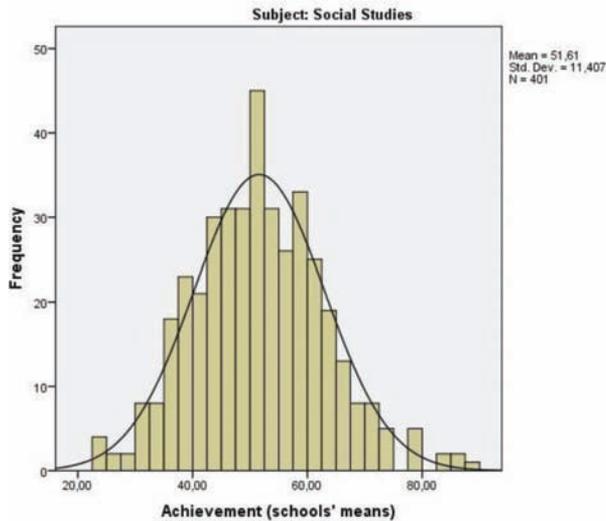
Figure 3.3.2 Normally distributed school population in the Social Studies test

After separating the schools into community and institutional school and plotting the population chart, a chart is given as in Figure 3.3.3.
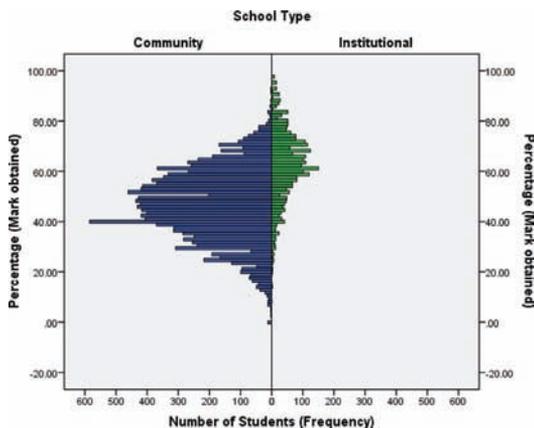
Figure 3.3.3 Distribution of the students' means in community schools and institutional schools

In Figure 3.3.3, the left hand side distribution shows that the community school students are distributed distinctly normal in the same way as the total student population. There are a remarkably high number of the students below the 30% of the maximum marks. The right hand side distribution shows the population of institutional school students, which are performing at a higher level even though there are quite a number of the students getting very low marks. This indicates that the students in the institutional schools vary from the low-performer to the highest performer though most of the students are the higher performers.

By analyzing the matter further with the scatter plot, Figure 3.3.4 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 schools to very low-performing schools.
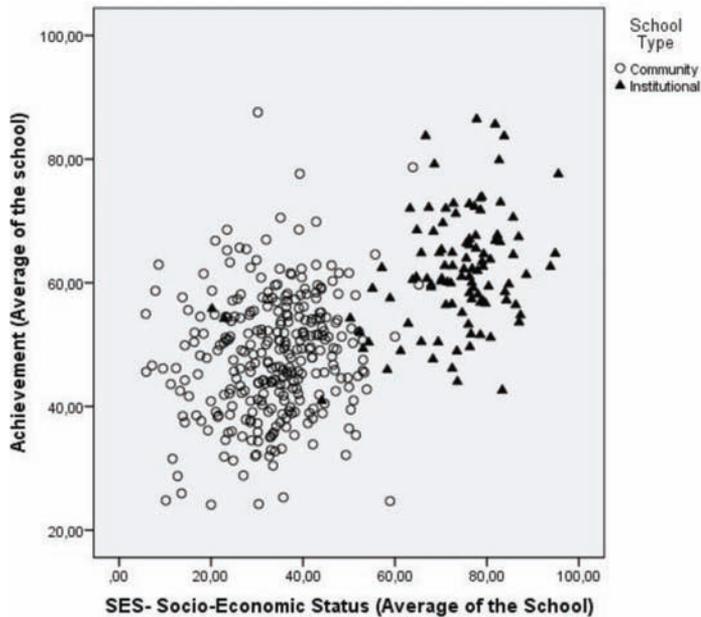
Figure 3.3.4 Achievement and social economic status and type of schools

According to the data, the achievement in Social Studies is normally distributed in both the student population as well as in schools' population.

The dataset gives a strong signal that the students in the institutional schools perform very well and the students in the community schools form two kinds groups of schools: high-performing schools and low-performing schools. The variety between the community schools is remarkable.

3.3.3.1.2 Different content areas and achievement

The whole Social Studies test was the combination of five content areas: *Geography*, *Civic*, *Economics*, *History*, and *Politics*. The maximum marks of the five content areas were proportionally equal to the weighting given by the curriculum. To compare the achievement in all the topics, these sub-scores are converted into a percentage of the maximum score of the content area. Table 3.3.4 shows the students achievement in Social Studies as a whole (total) and the achievement level in five content areas and Figure 3.3.5 illustrates the differences.

Table 3.3.4. Comparison of the total score average with different topics

Content area	Mean	SD	Min	Max
Geography	57	26.3	0	100
Civic	51	15.4	0	100
Economics	58	21.8	0	100
History	50	19.7	0	100
Politics	46	20.3	0	100
Social Studies as a whole (total)	49	15.3	0	98
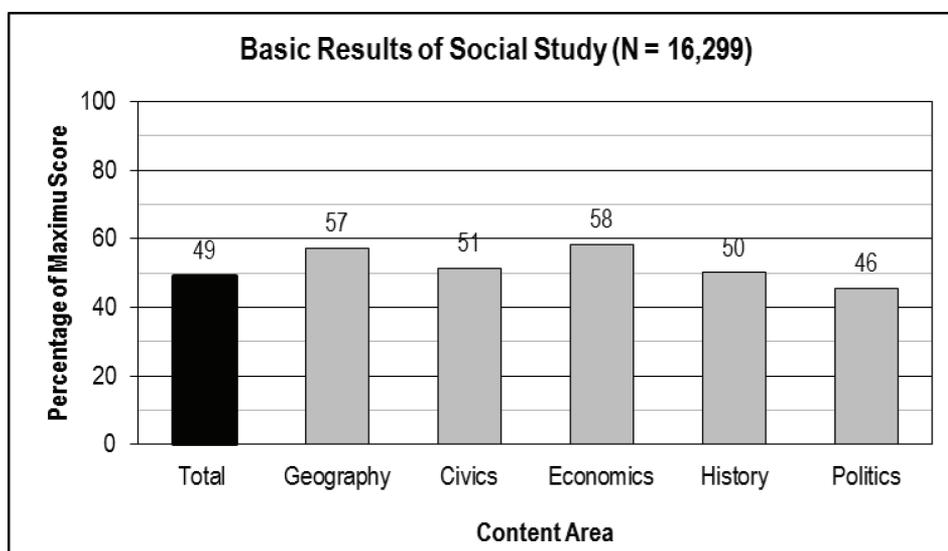
Figure 3.3.5 Basic results in the content areas of Social Studies

The percentage of achievement score shows that the national average of Social Studies is 49. Of the different content areas, students are the weakest in politics (46). They perform better than average in economics (58), geography (57), civics (51) and History (50).

   Dataset gives a signal that the learning outcomes are notably better than average in Economics (58) and Geography (57) and weaker than the average in politics (46).

3.1.3.1.3 Hierarchical level of cognitive domain and achievement

The Social Studies test as 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 is shown in Table 3.3.5 and illustrated in Figure 3.3.8.

Table 3.3.5 Achievement level based of hierarchical level of the items

Hierarchical level of item	Mean	SD	SE
Knowledge	52	22.0	0.2
Comprehension	66	21.7	0.2
Application	40	17.7	0.1
Higher Ability	34	17.0	0.1
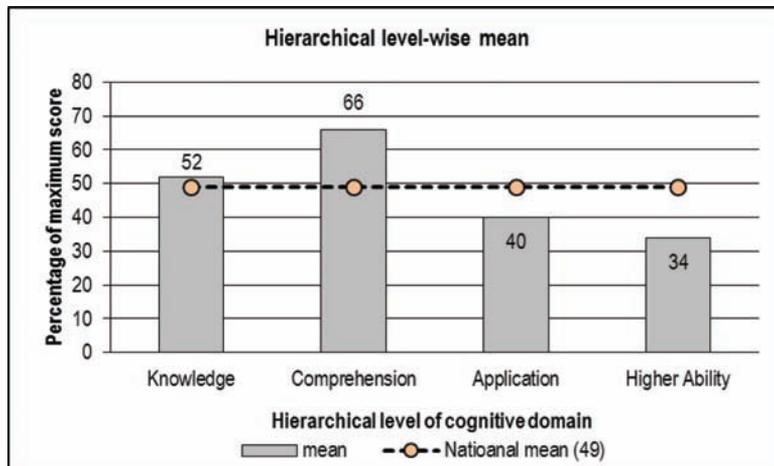
Figure 3.3.6 Achievement in different hierarchical levels of cognitive domain

181

Remarkably high a number of students were able to solve only 15% or less of maximum score of the practical problems, that is, the application type of items (8% of the students in Social Studies). In Social Studies, 12% of the students could solve just less than 15% of the tasks requiring the higher cognitive abilities.

The dataset shows that the students' ability to solve complex problems is low; only 34% of the maximum scores were reached. Students are much better in the comprehensive type of questions (66%), which may be explained by the educational system which seems to be grasping the meaning of given things than solving novel problems.

3.3.3.1.4 Type of item and achievement

There were basically two types of questions in the test: objectively marked and subjectively marked items. Objective items covered a wide range of content areas and were very specific to judge because there were 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 require a longer procedure to get the full marks. Both the objective and subjective types of items were made at 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 level. Table 3.3.5 comprises the basics statistics of the item type-wise achievement levels.

Table 3.3.6 Item type-wise means

Type of items	N	Mean	SD	SE
Objective	16299	69	16.1	0.13
Subjective	16299	36	17.2	0.13

It is evident that the subjectively scored tasks – usually those with more demanding requirements for the correct answer – are solved much poorer (36%) than the objective items (69%). Most of the objective items were knowledge-, comprehension- and application type whereas subjective items were application- and higher ability type.

Dataset gives a strong signal that the students are good in recognizing the correct answer and in very fundamental knowledge/contents, true and false, matching texts, puzzles, selection of words and fill in the gaps. They are much weaker in reasoning, problem solving, and constructing the information.

3.3.3.1.5 Comparison of 2011 with 2008 achievement result

The national assessment carried out in various years aims to assess the achievement and the progress over a period of the years. The National Assessment of grade 8 students carried out by the Educational Development Service Centre in 2008 (EDSC, 2008) shows the national average of the students is 53.4 ($\approx$ 53). The National Assessment of 2011 carried out by the NASA Unit of ERO (MOE) shows that the national average of Social Studies on grade 8 is 49. These figures are coming from *Classical Test Theory (CTT) 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 different difficulty level of the tests.*

The Ministry of Education has made the decision of using IRT modeling in NASA to make the national assessment results comparable with the previous and to-come datasets. In NASA 2011, the tests were linked with the test of 2008 so that the comparison was possible. Tests of 2008 and 2011 were equated to the same scale by using international parameters (fixed difficulty parameters) from the TIMSS dataset. However, Table 3.3.7 shows that there seems to be a radical difference between the marking schemes of 2008 and 2011; several items show such a remarkable difference in percentages of correct answers that the decision was made not to equate the 2008 and 2011 datasets. On the basis of the common items it seems (faulty) that the achievement level was dramatically lowered within four years. This cannot happen within the given timeframe and can only be explained by the change in the marking schemes; with the markers in 2008 marking more leniently compared with the markers in 2011.

Table 3.3.7 Comparison of the linking items of the years 2008 and 2011

	Percentage of correct answers		
	2008 (N = 2640)	2011 (N = 8091)	Change (2011 2008)
S08_1	0.86	0.41	-0.45
S08_7	0.59	0.29	-0.29
S08_8	0.63	0.31	-0.32
S08_11	0.89	0.28	-0.61
S08_12	0.84	0.07	-0.77
S08_13	0.49	0.43	-0.06
S08_14	0.63	0.41	-0.22
S08_15	0.63	0.55	-0.08

On the basis of the dataset it is not possible to say anything firm of the development in Social Studies between the years 2008 and 2011.

3.3.3.1.6 Comparison with the international standard

NASA 2011 results were made comparable with the international TIMSS assessment. A good number of Geography items (8) of the released TIMSS items were used as linking items. Their known difficulty parameters were fixed in the calibration of the local items. Hence the international average of $\theta = 0$ was fixed in the Nepalese datasets; when a student's ability level in NASA 2011 would be zero, it corresponds to the average level of international students.

Figure 3.3.7 shows the comparison of the students' achievement with the international standard. In the figure, the x-axis shows the content areas of Social Studies and y-axis shows the ability shown by the students. The middle horizontal line indicates the international average. As the ability is below the average, the bars are going down whereas when the ability is above the international average, the bars are going upwards.
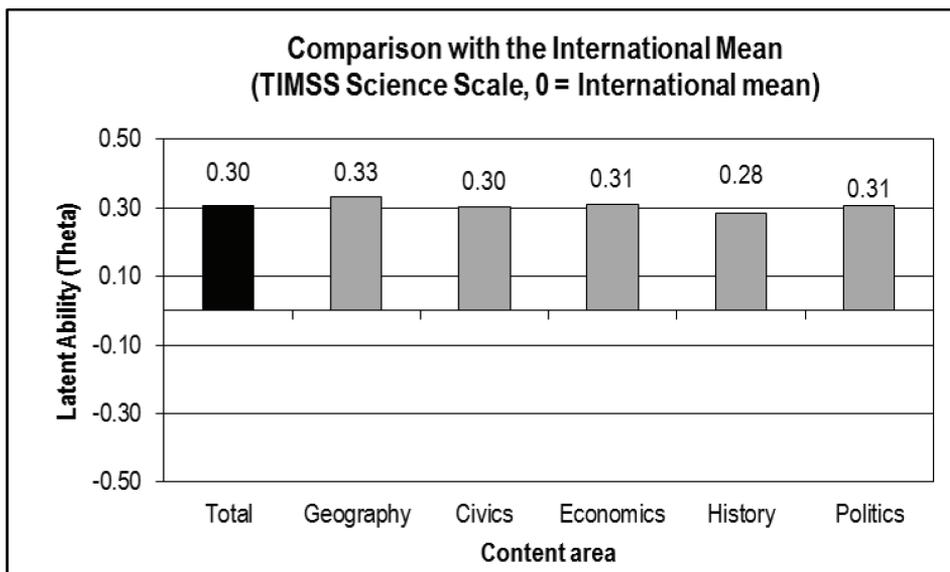
Figure 3.3.7 Student achievement in the international TIMSS scale

Figure 3.3.10 shows that the average ability shown by Nepali students in Social Studies as a whole (*Total*) is above the average line.[16] This indicates that the students in Nepal are somewhat better in Social Studies subject when compared with the international average. It seems that if the students would have participated in the TIMSS study, they would be quite good in Geography. It is good to remember that *all the linking items came from the content area of Geography* and hence there actually is no real equating in the other areas because the other areas are not measured in TIMSS. However, they are modeled on the basis of proficiency in the Geography test. More comparable research may be needed to confirm the result.

Dataset gives a signal that, from the international comparison viewpoint, the average achievement of Geography in Nepal seems to be better than the international average. It is, however, good to be cautious of the possible error in the estimation because of the process of the item selection.

3.3.3.2 Results based on diversity factors

Diversity is a relative and contextual term. In the context of Nepal, some of the experts have defined eight diversities in Nepal (see Section 4.1): geographical/ecological-, language-, gender/sex-, religious-, ethnic-, cultural-, disability- and economic diversity. NASA 2011 background information questionnaire included six of the above diversities; two diversities, 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) diversity. These factors can be taken as equity factors; all children regardless of their sex, language, birth place, or family background should have equal opportunities to reach the same educational goal.

3.3.3.2.1 Districts and student achievement

Out of 75 districts, 25 were randomly selected to represent the Ecological zones and Developmental regions and ultimately 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 condensed in Table 3.3.8 and Figure

---

[16] This comparison is based on item parameters published by PISA. Although the comparison is based on equating the tests with a good number of selected linking items, the result of the Indian students in 2012 PISA inquiry casts a shadow over the high (or average) result in Geography in Nepal. Assuming the results in Nepal are not much higher than in India – and India was clearly below the average in Mathematics and Reading. Now, in Reading the level in Nepal is the same kind as in India, that is, much lower than the average (see Section 3.2.3.1.6). The method of equating used in Geography was exact the same as in Reading (see Section 2.4.1). Hence, what may explain the difference in Indian and Nepalese results? One good explanation is that the selection of the items was done on the different basis in Reading than in Mathematics and Geography. In the latter subjects, only such items were selected which fits the local curriculum where in Nepali, the Reading items were selected more or less randomly. It is advisable to use the same idea also in the assessments to come.

3.3.8. The table shows the achievement in ascending order according to the achievement. The *mean* represents the average achievement percentage of the particular district.

Table 3.3.8 Average achievement score in the selected districts

District	N	Mean	SD	District	N	Mean	SD
Bhojpur	471	38	12.7	Tanahu	838	48	11.9
Jajarkot	423	40	15.0	Bara	726	49	14.1
Sankhuwasabha	242	44	12.8	Ramechhap	597	50	14.4
Kalikot	192	45	13.6	Syanjha	871	51	13.3
Morang	921	45	14.6	Lalitpur	400	51	12.3
Dhading	872	45	12.5	Doti	454	51	15.1
Ilam	1055	46	14.8	Jhapa	1230	51	15.3
Kanchanpur	622	46	13.1	Gorkha	713	53	11.6
Pyuthan	527	46	11.5	Sindhupalchowk	352	54	13.6
Nawalparasi	1117	47	16.4	Banke	637	55	16.1
Rautahat	880	47	17.6	Bhaktapur	335	57	15.8
Bajhang	259	47	15.2	Kathmandu	1531	61	15.2
Mustang	34	48	9.1				
				Total	19,299	49	15.3
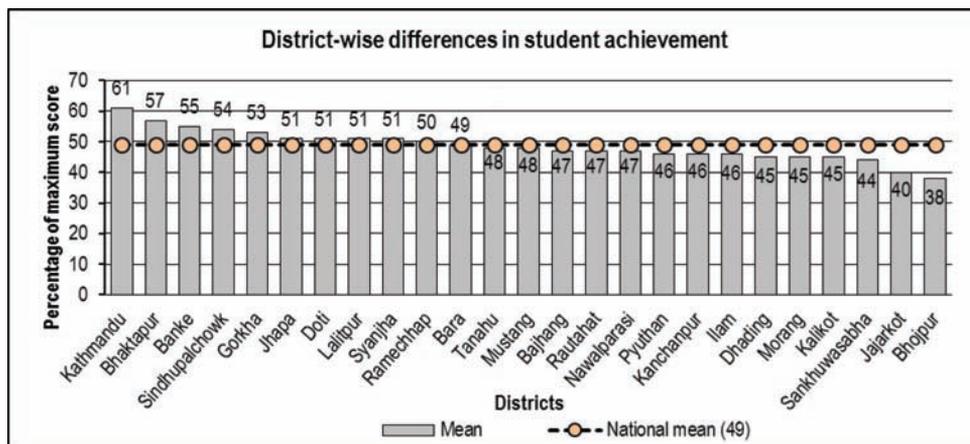
Figure 3.3.8 Achievement in the districts in the sample

Of the randomly selected districts in the sample, the student performance in Sicial Studies was very low in Bhojpur (38) from the Eastern region, and in Jajarkot (40) from the Mid-Western; the other districts had a mean score of 5 and above. Except for Banke (55) and Gorkha (53) districts, out of five best performing schools three come from the Central region and specifically two districts from the Valley area: Kathmandu (61) and Bhaktapur (57) and one came from the Central Mountain district of Sindhupalchowk (54). The other two best performing schools are from the Mid-Western and Western regions. The data mining module of the SPSS software, Decision tree analysis, indicates that the lowest performing schools can be found in the Eastern and Mid Western Hill areas.

The difference in achievement due to the different district is statistically significant ($p < 0.001$). The variation explained in achievement due to the district is $\eta^2 = 0.12$; the district explains 12% of the variation in the data. Effect size is $f = 0.37$ indicating that the difference between the lowest performing district (31) and highest performing district (61) is remarkably high.

The dataset gives a strong signal that there is a wide difference between the districts when it comes to the equal opportunities of children to reach the preset goals in Social Studies. The results are bound to the 25 districts selected randomly to the sample; even lower-performing districts could be found if other districts would have been selected. In any case, the results in Bhojpur (38), Jajarkot (40), Sankhuwasabha (44) Kalikot, Morang and Dhading (45 each) are so low that raising the standard in these districts would raise the standard in the whole country.

3.3.3.2.2 Ecological zone and student achievement

Access to educational facilities can play a vital role in students' achievement. The Mountain, Hill and Tarai are three geographical features in Nepal though the Valley can be 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, weather conditions, economic activities, aggressive development, as well as the dense human capacity make the Valley a unique fourth geographical area in the analysis. The variation in the Ecological zones in NASA 2011 is condensed in Table 3.3.9 and Figure 3.3.9.

Table 3.3.9 Achievement in the ecological zones

Region	N	Mean	SD	SE	Minimum	Maximum
Mountain	1079	48	14.3	0.4	7.1	87.1
Hill	6821	47	13.9	0.2	0.0	90.6
Tarai	6133	48	15.8	0.2	0.0	97.6
Valley	2266	59	15.3	0.3	4.7	97.6
Total	16,299	49	15.3	0.1	0.0	97.6
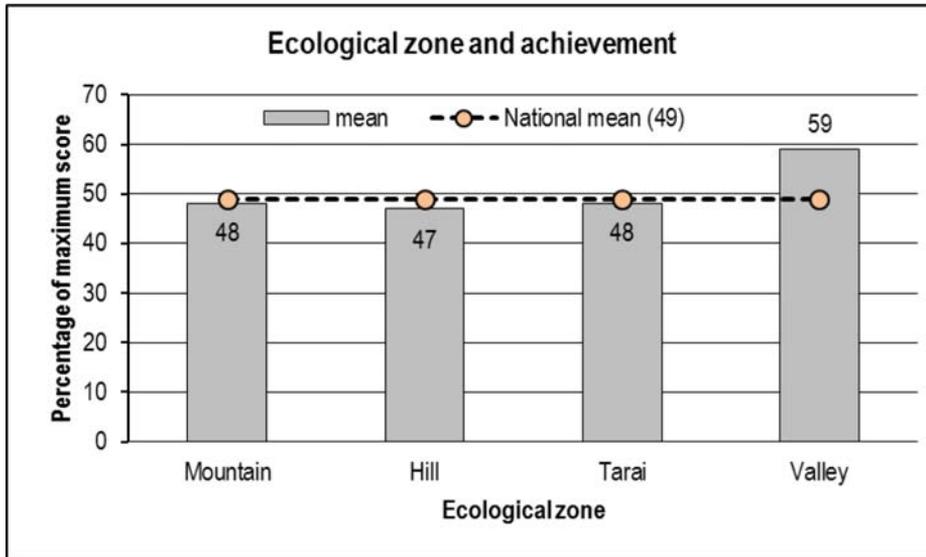
Figure 3.3.9 Differences in achievement in different Ecological zones

The data shows that, on average, the students from the Valley (59) outperform the students from all the other Ecological zones. The students from the Hill area perform the lowest (47). The achievement in the regions differs significantly ($p < 0.001$). Tukey's *post hoc* test tells that the Valley differs from all other regions and Tarai differs from Hill; otherwise there are no differences between the regions. The effect size is $f = 0.25$ showing moderate difference between the highest and lowest performing Ecological zones; Ecological zone explains 6.1% of the variance in the data. When excluding the Valley, there is practically no difference between the regions which is a good sign from the equity viewpoint ($f = 0.03$). As a comparison, remember that the district explains more than 10% of the variation.

Dataset gives a signal that with the exception of the Valley region, there is no difference between the Ecological zones. Students in the Kathmandu Valley (59) outperform the other students (47–49).

3.3.3.2.3 Developmental region and student achievement

The student achievement varies according to the Developmental regions which are divided into Eastern-, Central-, Western-, Mid-Western-, and Far-Western regions. Additionally, the Kathmandu Valley is taken as the sixth developmental region though technically it falls under the Central Developmental region. The mean achievements in the Developmental regions are given in Table 3.3.10 and illustrated Figure 3.3.10.

Table 3.3.10 Achievement in the developmental regions

Region	N	Mean	SD	SE	Minimum	Maximum
Eastern	3919	46	15.1	0.2	0.0	90.6
Central	3427	48	14.9	0.3	0.0	87.1
Western	3573	49	13.9	0.2	2.4	90.6
Mid-Western	1779	48	15.4	0.4	3.5	97.6
Far-Western	1335	48	14.4	0.4	2.4	82.4
Valley	2266	59	15.3	0.3	4.7	97.6
Total	16,299	49	15.3	0.1	0.0	97.6
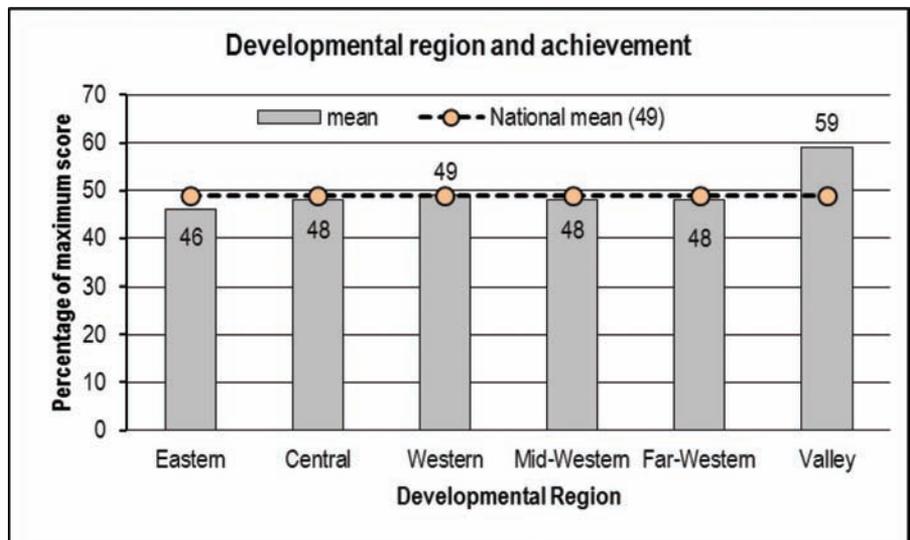
Figure 3.3.10 Comparison of student achievement in the developmental regions

The best performance can be found in the Valley (59) and in the Western region (49). The performance is the lowest especially in the Eastern region (46) though the Center, Mid-West and Far-West are also regions in which students' performance is lower than the average (i.e. 48 in each Developmental regions). The difference between the regions is statistically significant ($p < 0.001$) and Tukey's *post hoc* test shows that there is no difference between the Central-, Mid-Western-, and Far-Western regions; but the Eastern region differs from all regions. Effect size is moderate or high ($f = 0.26$); Developmental region explains 64% of the student variation. If the Valley is

189

taken away from the analysis, the effect size if very low ($f = 0.03$) indicating that there is actually no differences between the regions. This is a good sign form the equity viewpoint.

The dataset gives a signal that, when it comes to the achievement in Social Studies, with the exception of the students in the Valley (59), there are no practical differences between the Developmental regions (46–49). From the equity viewpoint, this is a good sign. However, the difference between the lowest performing area (Eastern region, 46) and the highest performing region (Kathmandu Valley, 59) is remarkable – around 13 percent points.

3.3.3.2.4 School type and student achievement

All the schools are categorized into community-managed- and institutional-managed schools. In what follows, these are referred to as community schools and institutional schools; the latter could also be called private schools. The difference in the achievement of the two types of schools is condensed in Table 3.3.11.

Table 3.3.11 Type of school and the average achievement

Type of school	N	Mean	SD	SE
Community	13486	46	14.1	0.1
Institutional	2813	63	13.7	0.3
Total	16299	49	15.3	0.1

The achievement levels in the community schools and institutional schools differ from each other remarkably. The average performance in the private schools is 63 whereas in the community schools it is 46; 17 percent point's difference is remarkable. The difference is statistically significant ($p < 0.001$) and the effect size is high ($f = 0.44$) showing a wide difference between the community- and institutional schools. Division of the children to the community- and institutional schools explains 16% of the student variation ($\eta^2 = 0.16$). From Figure 3.3.4 above, it is known that the deviance within the community schools is remarkable ranging from 25 to 85; contrarily, all private schools in the sample show higher mean than 40. Natural reason for this is the student selection. The selection of the students evidently takes the best students from the community schools and hence the average performance of the community schools may be lower when there are possibilities to select the school. It may be worth noting that the best community schools are not in Kathmandu Valley. It is also worth noting that the number of the

students in the private schools is not very high; the main educational work is done in the community schools.

The dataset gives a strong signal that, in Social Studies, the students in the institutional schools outperform the students in the community schools. Most probably this deviance can be explained by student selection, however, this cannot be known on the basis of the dataset.

3.3.3.2.5 School location and student achievement

One of the strata of sampling in NASA 2011 was the school location. The schools were divided into two: rural schools and urban schools. The achievements of the students in rural and urban schools are condensed in Table 3.3.12.

Table 3.3.12 Student achievement on the basis of location of school

Type of school	N	Mean	SD	SE
Rural	12022	48	15.0	0.1
Urban	4277	52	16.0	0.2
Total	16299	49	15.3	0.1

The achievement level of the students in the urban schools (52) is somewhat higher than that of rural schools (48). The difference in average score is significantly different ($p < 0.001$) though the effect size is not high ($f = 0.10$); division into rural and urban schools explain only 1.1% of the student variation ($\eta^2 = 0.011$).

Data gives a positive signal that though the results are somewhat better in cities than in the rural areas, the difference is not remarkably high. From the equity point, this is a positive thing.

3.3.3.2.6 Language at home and student achievement

In the context of Nepal, the student achievement may depend on the language spoken in their homes i.e, the mother tongue. The mother tongue reflects, in many cases, the ethnical background and hence any difference may be taken as a possible source for inequity in society. In Nepali language assessment (see Sections 3.2 and 4.6) the results are obvious. However, it is not known what the effect may be in the subject of Social Studies.

On the basis of the total data, 27.7% of the 8th 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 Magars (2.1%), Tamangs (2.6%), and Tharus (2%). After dividing the languages into ten groups excluding Nepali, there were still 17.2% of the students classified into the group "else". 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 were grouped into "non-Nepali speakers". The results are condensed in Tables 3.3.13 and 3.3.14 and Figure 3.3.11.

Table 3.3.13 Student achievement on the basis of language group

Language group	N	Mean	SD	SE
Nepali	11783	50	15.1	0.1
Non-Nepali	4516	47	15.5	0.2
Total	16299	49	15.3	0.1

When combining all the minor language groups as "Non-Nepali", there is difference of 3 percent point between the language groups. Hence the difference is not very wide. However, on the basis of Table 3.3.14 with the original categorization of the minor languages, the issue looks very different: It is evident that the small groups of Sherpa (57) and Newari (57) speakers perform very well in Social Studies. It is also noteworthy that the equally small language groups, such as Rai-, Limbu-, and Tharu speakers perform very low (36, 42, and 45 respectively). Especially, the Rai speakers would need remarkable support to raise their standard.

Table 3.3.14 Achievement in the different language/ethnic groups

Language/Etnicity[1]	N	Mean	SD	SE	Minimum	Maximum
Sherpa	14	57	8.9	2.4	43.5	78.8
Newar	189	57	15.6	1.1	16.5	97.6
Gurung	125	52	11.9	1.1	20.0	87.1
Nepali	11783	50	15.1	0.1	0.0	97.6
Tamang	419	47	15.2	0.7	12.9	90.6
Magar	343	46	13.1	0.7	8.2	75.3
Tharu	329	45	14.3	0.8	4.7	83.5
Limbu	89	42	10.7	1.1	12.9	70.6
Rai	70	36	13.2	1.6	12.9	77.6
Other language	2768	47	15.9	0.3	0.0	95.3

1) those language groups are omitted with less than 13 students
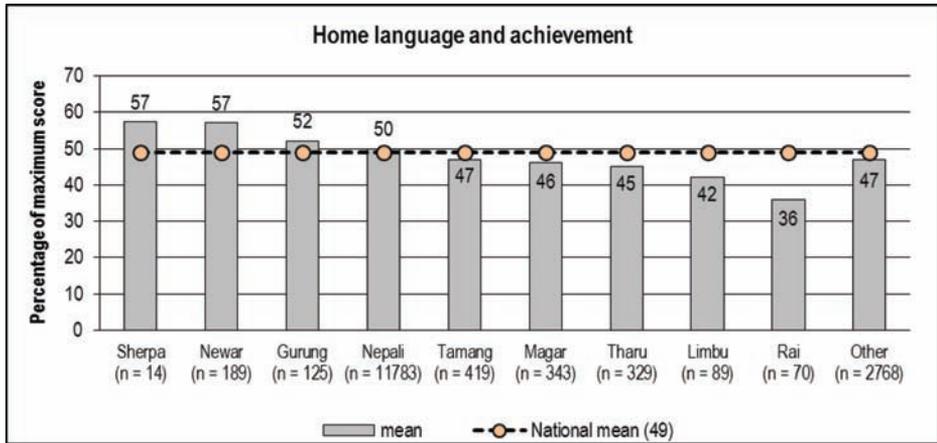
Figure 3.3.11 Relation between language at home and achievement

The difference between the language groups is statistically significant ($p < 0.001$). However, the effect size is small ($f = 0.14$) because the minority groups are small; division into smaller language group cannot explain the variation in the data ($\eta^2 = 0.02$). When analyzing only the minority languages and hence, excluding the Nepali speakers and the group "else", the effect size is high ($f = 0.33$) indicating really remarkable difference between the highest performing group (Sherpa, 57) and the lowest performing group (Rai, 36).

Language and developmental region

When combining the results from the developmental region and the mother tongue, the Valley seems to carry a couple of interesting characteristics (Table 3.3.15). In Nepali, Tharu, Newar, Gurung, Sherpa, and Limbu groups the results in the Valley is higher than in any of the regions. This means that the language group itself does not determine the achievement level of the students. One example of this is the fact that though the Rai students in the Central region (n = 13) perform very low in Social Studies (average 26), one student in the Western region had a very high mark (66) and three students in the Valley had an average of 55. Another related fact is that, in the groups of Newar, Rai, Gurung, and Limbu, the lowest achievement is found in the Central region.

Table 3.3.15 Achievement in the different language groups in different regions

Development region	Nepali	Magar	Tharu	Tamang	Newar	Rai	Gurung	Sherpa	Limbu
Eastern	48	52	36	48	49	37	61		42
Central	48	46	44	46	45	28	34	59	32
Western	51	46	39	55	52	66	52	48	42
Mid-Western	48	38	50	44	41				
Far-Western	45	75	46	59			20	48	
Valley	59	45	50	48	62	55	79	59	68
Total	50	46	45	47	57	36	52	57	42

1) the main population is highlighted by the gray shade. In some cases there is only one student behind the mean.

Language and Ecological zone

As appears to happen in the Central region, also happens in Tarai region. In the groups of Tharu, Tamang, Rai, Gurung, and Sherpa, the lowest results come from the Tarai region (Table 3.3.16). The differences within the language groups may be very wide; for example, when the average of Gurung students is quite high (52), the average of two Gurungs in the Valley was 79 but in Tarai the student gained only 20. The reason for the phenomenon stays open. However, it means that the mother tongue alone does not determine the future of the children.

Table 3.3.16 Achievement in the different language groups in different zones

Ecological zone	Nepali	Magar	Tharu	Tamang	Newar	Rai	Gurung	Sherpa	Limbu
Mountain	48	42		44	62	41	61	59	39
Hill	47	46		48	44	37	51	48	42
Valley	59	45	50	48	62	55	79	59	68
Tarai	51	46	45	42	52	27	20	48	43
Total	50	46	45	47	57	36	52	57	42

1) the main population is highlighted by the gray shade. In some cases there is only one student behind the mean.

Rai speaking students (score 36) deviates from the national mean (49) remarkably. Students who speak the Rai language are located in seven sampled districts namely: Bhojpur, Sankhuwasabha, Morang, Ilam, Rautahat, Tanahu and Kathmandu. Out of seven, three districts (Bhojpur, Sankhuwasabha and Morang) are located in the

Eastern region and scored below 42. Likewise, Limbu speakers who are the second lowest performing language group in Social Studies (42), who are also situated in two Eastern districts have a low achievement score (Sankhuwasabha, 39 and Morang, 40 respectively). However, those Rai speakers from Kathmadu (55) and Tanahu (66) and Limbu speakers from Bhaktapur (68) are outperforming, i.e. above the national average (49). Sherpa speakers are outperforming in all the five districts where they have been located. Newars are located in 12 sample districts and only in Dhading district are they performing lower than 40. This scenario hints that different language speaking students who were performing low are mostly located in the low performing districts and regions.

The Dataset arises two difficult questions: What has been done differently – or has not done at all – in the Central Tarai area where Gurung, Rai-, Tamang-, and even Sherpa students get very low scores? Why are the results in Tarai area lower than in other regions?

3.1.3.2.7 Caste and student achievement

Modern education has been influenced in several ways by the legacy of the historical caste system. Though the caste system is officially abandoned it still lives in the mind-sets of most Nepali people. Historically, the Brahmans and Cheetris have been heavily involved in education, but Dalits, for example, have been practically outside of the educational system. Hence, the modern society has made lots of efforts to make the education possible and accessible for all children. The latest household survey (2012) shows that number of Hill Dalits has been increased remarkably in the lower education system, but their number in the secondary and higher education is still very small. The results concerning the castes and achievement are condensed in Table 3.3.17 and illustrated in Figure 3.3.12.

Table 3.3.17 Achievement in the different castes

Caste	N	Mean	SD	SE	Minimum	Maximum
Brahman/Cheetri	6466	52	15.4	0.2	0.0	97.6
Janjati	6376	48	14.9	0.2	0.0	97.6
Dalit	1648	46	13.8	0.3	5.9	92.9
Madhesi	1495	47	16.8	0.4	0.0	95.3
Alpasankhyak	221	47	15.2	1.0	3.5	90.6
Total	16206	49	15.3			
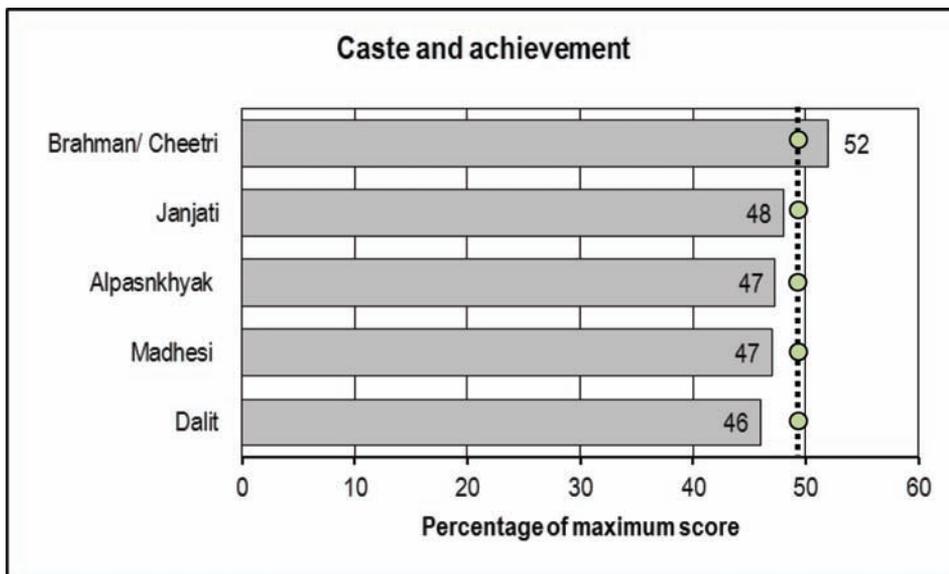
Figure 3.3.12 Relation between caste and achievement

On the basis of NASA 2011, it is obvious that the Dalits are still performing lower (46) than the other castes and especially lower than Brahmans/Cheetris (52). The overall difference between the groups is statistically significant ($p < 0.001$) but the effect size is small or moderate ($f = 0.14$); dividing student according to their caste background explains only 2% of the student variation ($\eta^2 = 0.020$). Tukey's *post hoc* test reveals that Madhesis, Dalits, and Alpasankhyaks do not differ from each other.

A positive sign from the equity viewpoint is that Dalit students perform remarkably better than the national mean in the Central- (55) and Eastern- (52) Mountain; Western (50) and Far-Western-Hill (50); and Mid-Western Tarai (54) (Table 3.3.18). The results are also just above the average in the Far-Western Mountain region (50). Unfortunately though, it seems that the results are much lower than the average in the Mid-Western Mountain (42); Eastern- (41) and Mid-Western Hill (42); and Western- (43) and Far-Western Tarai areas (42). The number of the students in certain strata is small and hence it may be wise not to make too strong and farfetched implications on the basis of these results.

Table 3.3.18 Dalit students' achievement in different Ecological zone and developmental regions

Ecological zone	Developmental region					
	Eastern	Central	Western	Mid-Western	Far-Western	Total
Mountain	521	55	49	42	50	49
Hill	41	45	50	42	50	46
Tarai	47	45	43	54	42	46
Total	45	46	48	45	45	46

1)   the highest averages per each region are highlighted by the gray shade.

The Dataset gives a signal that, in Social Studies, Madhesi- (46), Dalit-(47), and Alpasankhyak (47) students' performance is lower than the results of the other castes; the highest achievement is in the group of Brahman/Cheetri students (52). The Dalits performance in particular is low in the Eastern Hill- (41), Mid-western Mountain and Hill- (42) and Far-Western Tarai (42) areas. A positive signal is that the Dalits' performance is high in Eastern- (52) and Central Mountain- (55) and the Mid-Western Tarai area (54)

3.3.3.2.8 Sex and student achievement

Lots of efforts have been put globally into reducing the difference between boys' and girls' school achievement. Because the sex- or gender-wise equity seems to be important in the modern discourse, the matter is handled somewhat more extensively than the previous sections of equity. Basic results are condensed in Table 3.3.19 and Figure 3.3.13.

Table 3.3.19 Student achievement of boys and girls

Sex	N	Mean	SD	SE	Minimum	Maximum
Boys	7882	50	15.3	0.2	0.0	97.6
Girls	8354	49	15.3	0.2	0.0	97.6
Total	16236	47	15.9			
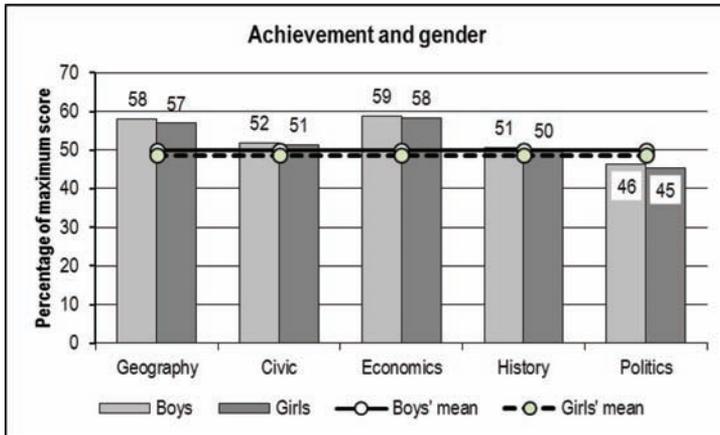
Figure 3.3.13 Comparison of boys and girls in different content areas

There is a statistically significant difference between boys (50) and girls (49) ($p <$ 0.001). It is noteworthy that in all content areas boys outperform the girls. However, the effect size is small ($f = 0.04$) indicating that the difference is not remarkable. Sex explains only 0.2% of the student variation. From the equity point of view the signal is positive though there is still work to do to reduce the gap.

Sex and caste

On the basis of NASA 2011 data, the difference between boys and girls is the highest in Dalit- (difference is 3 percent points) and Madhesi societies (2 percent points). Tukey's *post hoc* test shows that the differences are statistically significant at least at 5% risk level. There is practically and statistically no difference within Brahman/Cheetri-, Janajati- and Alpasankhyak castes (1 percent point) (Figure 3.3.14). From the equity point, the signal is positive.
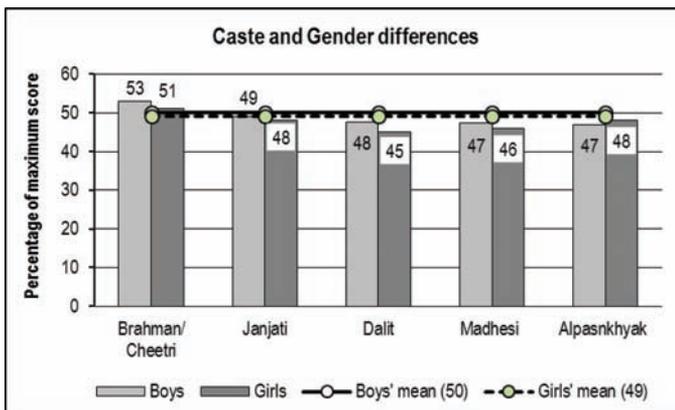
Figure 3.3.14 Caste/ethnicity-wise comparison of achievement of boys and girls

Sex and Ecological zone

The achievement of girls and boys differs significantly among ecological belts *(p < 0.001).* There is no statistically significant difference between boys and girls though boys in the sample are somewhat better than girls in all the regions (Figure 3.3.15). From the equity point of view the signal is positive.
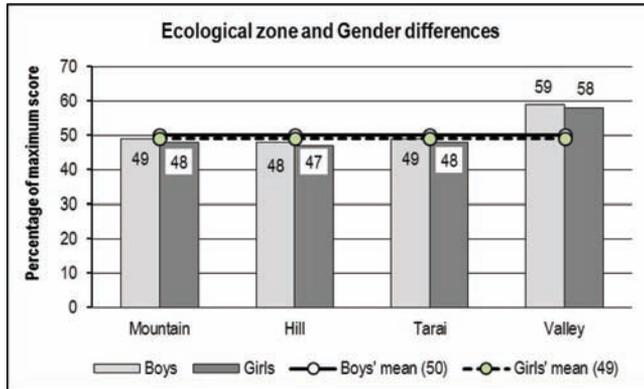
Figure 3.3.15 Ecological region and gender-wise differences

Sex and developmental region

Among the Developmental regions, there are statistically significant differences between the boys and girls *(p < 0.001)* though the differences are small. The difference is the highest in the Far-Western region (3 percent points) and in the Western region (2%) and notably smaller (1%) in the Eastern- and Mid-Western regions (Figure 3.3.16). However, there is no difference in the Central region.
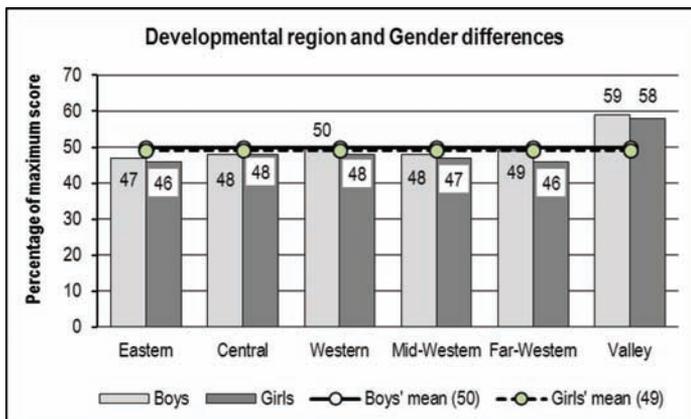
Figure 3.3.16 Developmental region and gender-wise differences

199

Dataset shows that the boys are out-performing girls in all content areas of Social Studies. Differences are not wide though they are significant. Sex explains only 0.2% of the student variation; from the equity point of view the signal is positive though there is still a need to reduce the gap. Differences are the widest in the Far-Western region; in the Central region there is no difference between the genders. All in all, in Social Studies, the differences between the boys and girls are small.

3.3.3.3 Selected explanatory factors and achievement

The simplistic model in Section 2.3.1 represents several possible factors which may explain the differences in student achievement. Many of the factors have already been handled in Section 3.3.3.2: geographical factors, such as Districts, the Ecological region, and Developmental region, as well as School-related technical factors, such as school Type and school Location. Also some individual related factors were handled, such as Home language, Caste and Sex/Gender. In this section, several other factors are taken into consideration. The socioeconomic status (SES) of the students' families, paid work after school, students' attitude towards Social Studies as a school subject, age of the student, and help given to the studies are mainly family- and individual related factors. As a sample of deepening school- and teacher-related factors, availability of schoolbooks, homework given by the teacher, and selected activities in the school are handled. Many other factors could be selected; the background questionnaire is rich.

3.3.3.3.1 Socioeconomic status (SES) and student achievement

There were many variables indicating the socioeconomic status. In NASA 2011, they were categorized into parents' education, parents' occupation, home possessions (whether or not the student has his own space to do homework, or a dictionary), home accessories (like how many mobile phones, cars or bathrooms there is 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 2.4.2). In this section, the education of the parents is further elaborated on, so that the illiteracy of the parents is analyzed in relation to the Social Studies achievement.

Several SES-related variables were analyzed by using a data mining tool of SPSS, Decision tree analysis (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. Some examples of this are

handled with parents' educational background and its relation with students' achievement in Social Studies.

Parents' education

In NASA 2011 background questionnaire, parents' education is divided into nine categories: 1) illiterate, 2) literate, 3) grades 1–5 pass, 4) grades 6–10 pass, 5) SLC pass, 6) IA pass, 7) BA pass, 8) MA pass, and 9) Above MA pass. The question was asked from the students and hence there may be some impurities embedded in the data. However, with the huge dataset the results seem credible.

DTA classifies mother's education into five groups with statistically significant differences in students' achievement levels: illiterate mothers (students' average is 46% of the maximum score), just literate and grade 1–5 passed mothers (50%), grade 6–10 passed mothers (52%), SLC passed mothers (56%) and grade IA and above (54) (Figure 3.3.17, of the explanation of the elements in the figure, see Section 2.7.1). In each group the number of mothers is high enough to make a credible prediction. The difference between each group is statistically significant ($p < 0.001$). In practical words, the results mean, for example, that when the mother has passed IA she can give 10 percent point advance in the national test compared with the illiterate mothers.
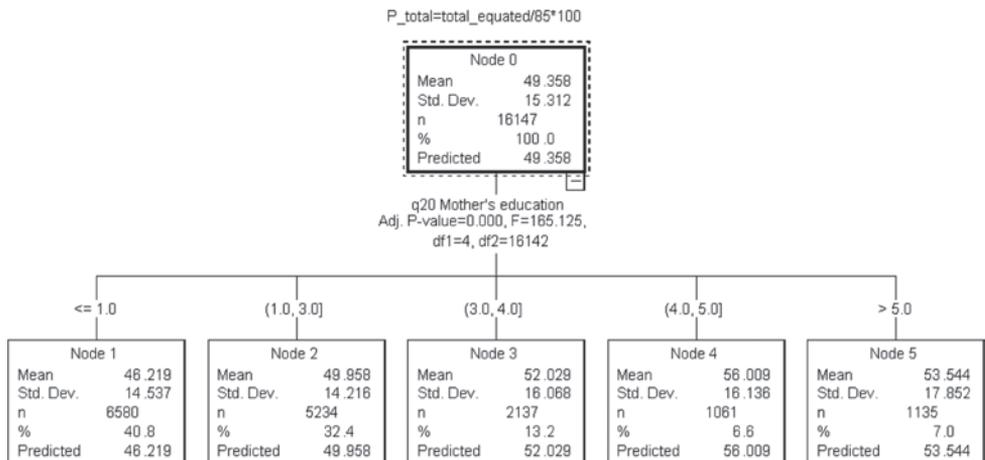
Figure 3.3.17 DTA of mother's education and students' achievement in Social Studies

In parallel, the DTA divides the father's education into six categories: illiterate or just literate fathers (average 44), grade 1–5 passed fathers (48), grade 6–10 passed fathers (51), and SLC passed fathers (53), IA passed fathers (56), and higher than

IA-passed fathers (54) (Figure 3.3.18). Compared with the mother's education, note that, if the mother has passed the grade 1–5, it raises the standard of the children more (50) than if the father is just literate (44). The same can be seen with mothers who have passed the SLC (56) and fathers (53). However, when the father has passed AI, the results are remarkably higher (56).

P_total=total_equated/85*100

Node 0	
Mean	49.372
Std. Dev.	15.321
n	16111
%	100.0
Predicted	49.372

q21 Fathers education
Adj. P-value=0.000, F=178.967,
df1=5, df2=16105

<= 1.0

Node 1	
Mean	43.502
Std. Dev.	14.397
n	2527
%	15.7
Predicted	43.502

(1.0, 3.0]

Node 2	
Mean	47.786
Std. Dev.	14.218
n	5598
%	34.7
Predicted	47.786

(3.0, 4.0]

Node 3	
Mean	50.531
Std. Dev.	14.661
n	3413
%	21.2
Predicted	50.531

(4.0, 5.0]

Node 4	
Mean	52.576
Std. Dev.	14.959
n	2163
%	13.4
Predicted	52.576

(5.0, 6.0]

Node 5	
Mean	56.423
Std. Dev.	15.765
n	670
%	4.2
Predicted	56.423

> 6.0

Node 6	
Mean	54.034
Std. Dev.	17.892
n	1740
%	10.8
Predicted	54.034

Figure 3.3.18 DTA of father's education and students' achievement in Social Studies

When combining the mother's and father's education, the poor prediction for the children's future achievement in Social Studies comes, if either the father is or both father and mother are illiterate (43). In the other extreme, the best results come when the father is IA-passed and the mother has passed grades 1"5 or higher (58), or the father is SLC and IA passed and the mother has passed grades 1"5 (58). It seems 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, the cut-off for parental education was set to "SLC-passed", that is, when being SLC-passed (or higher), the indicator for mother's (and father's) education for SES was set to 1, and the lower education than SLC-passed gave the value 0.

Dataset gives a strong signal that the educational level of the parents predicts the children's future achievement level in Social Studies. Especially harmful for the achievement in the Social Studies seems to be the situation that either the father or both parents are illiterate.

Parent's occupation

The occupation of parents was categorized into five categories: 1) agriculture, 2) teaching, 3) services, 4) business, and 5) others. Similar manner as with the parents' education, DTA was used to find statistically the most deviating groups related to student achievement. The student achievement is the lowest when the mother's occupational background comes from an agricultural background (47). It is statistically significantly higher when the mother comes from other services (57) than teaching, and service. The student's achievement is also better when the mother depends on business (53). It seems that either economic- or intellectual capacity or both at home helps children to raise their standard.

P_total=total_equated/85*100

Node 0	
Mean	49.355
Std. Dev.	15.323
n	16027
%	100.0
Predicted	49.355

q22 Mother's occupation
Adj. P-value=0.000, F=440.310,
df1=2, df2=16024

<= 1.0   (1.0, 4.0]   > 4.0

Node 1		Node 2		Node 3	
Mean	47.194	Mean	53.087	Mean	56.913
Std. Dev.	14.436	Std. Dev.	16.132	Std. Dev.	15.842
n	11316	n	2916	n	1795
%	70.6	%	18.2	%	11.2
Predicted	47.194	Predicted	53.087	Predicted	56.913
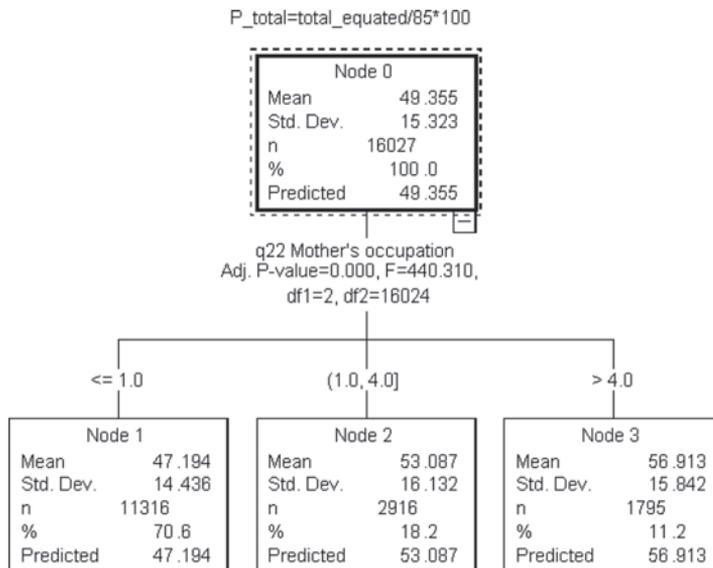
Figure 3.3.19 DTA of mother's occupation and students' achievement in Social Studies

When it comes to father's occupation, the main division is whether the father works in agriculture (46) or not? If the father's occupation was in Service, the students would gain 6 percent points more in the test (52) and if the father is either in business or any other occupation the results are also that good (51) and not that much different than from a business background.

203

P_total=total_equated/85*100
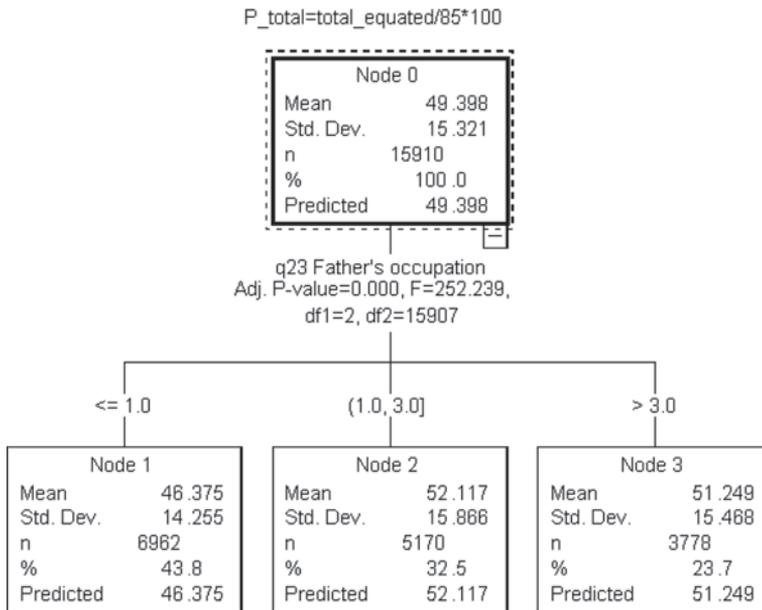
Figure 3.3.20 DTA of father's occupation and students' achievement in Social Studies

When combining the mother's and father's occupation the lowest achievement is found in the families where both parents come from an agricultural background (45) or mother alone in business (47) – here the business may refer to selling and serving the agricultural goods in markets. Still, another somewhat low-performing group are the students from the families where the mother is in agriculture and the father in "service or business" (47).

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

The dataset gives a strong signal that if the father or mother or both are coming from an agricultural or related occupation, the students' achievement in Social Studies is significantly lower than with the other occupational groups.

Parents illiteracy and agriculture

An additional question related to parents' education and occupation is whether the illiteracy of either parent is related to the agricultural occupation. On the basis of the previous analyses, the student performance is the lowest in the illiterate homes

and in agriculture. When focusing only on the illiterate parents, the DTA shows that there are two pockets of very harmful combinations for children's Social Studies achievement. Generally speaking, the results are better if the parents do not come from an agricultural background. Here, the achievement is low when both parents work in agriculture and the mother alone is illiterate (43). However, *if the mother works in agriculture and the father is not literate, student achievement is very low (41).* Of the analysis above, it is known that if both parents come from an agricultural background then the students' social studies achievement is very poor regardless of their parents' literacy levels (41). If the mother comes from a non-agricultural background and both parents are literate, student achievement van be very high (56). If the father is illiterate but the mother is literate and comes from a non-agricultural background then student achievement is also quite high (51). It seems that both the mother's education and economic status influences students' learning somewhat more than that of father's.

Home possessions and accessories

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

There are 12 questions in the student background questionnaire related to home possessions. Each was scored 1 when the student expressed his/her access to this possession (e.g. having a separate room or a table to study). Adding these items up, the maximum score was 12 indicating that the student expressed to have access to *all* of the possessions and the lower the score the fewer possessions they have at home. Figure 3.3.21 shows the connection of home possessions and achievement level: the achievement level of the students' raises logically the more there is access to home possessions. Pearson product moment correlation coefficient between the achievement level and the factor ($r = 0.22$) is statistically significant ($p < 0.001$) which is moderately high. Effect size is moderate ($f = 0.25$); the difference between the lowest (41) and highest (59) scores are remarkable.

For the later use in SES, the cut-off for the factors was set on five possessions: if there were five items or more mentioned in the background questionnaire, the student was given 1 otherwise 0.
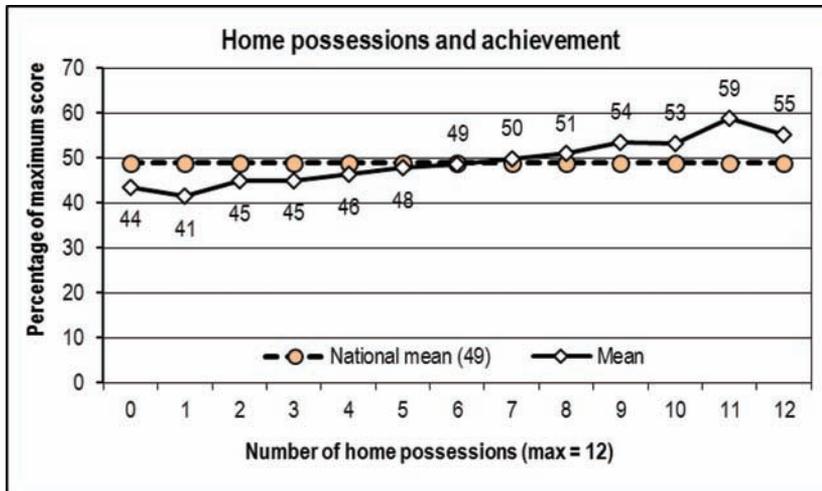
Figure 3.3.21 Connection of home possessions and achievement in Social Studies

The same pattern – the more possession, the better results – can be seen with home accessories, as seen in Figure 3.3.22. The question in the background questionnaire was set differently compared with the 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). After dichotomizing the items individually by using meaningful cut-offs found with ANOVA and DTA (and maximizing the differences in achievement level, see Table 3.3.20), all the five indicators were summed. The maximum score was 5 indicating that the student's home possessed a set number of all the accessories.

Table 3.3.20 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
Car	1–3	0, missing
Bathroom	2, 3	0, 1, missing
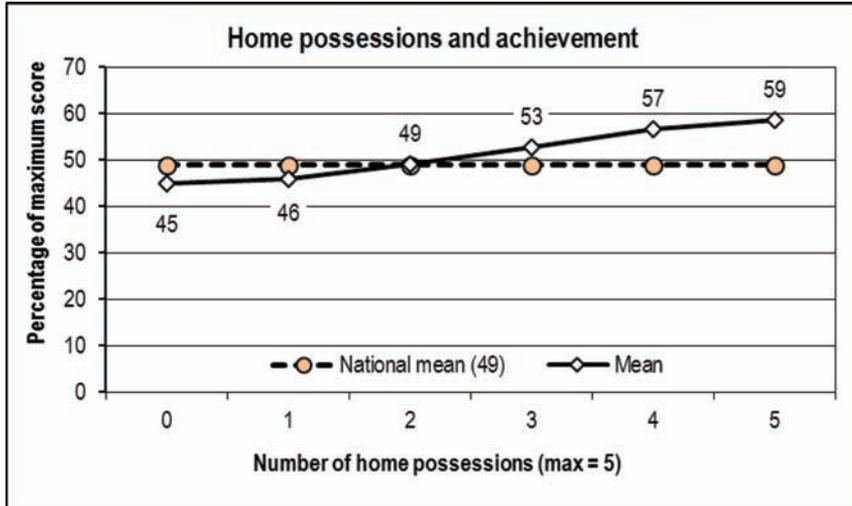
Figure 3.3.22 Connection of number of home possessions and achievement in Social Studies

Figure 3.3.22 clarifies how the number of home possession or accessories increases, students' achievement increases from 45 (if none of them are available) to 59 (if all of them are available). Availability of all the stated facilities indicates the high SES of the family. Correlation between home accessories and achievement is $r = 0.26$ ($p < 0.001$) certainly positive though not very high. The effect size of $f = 0.28$ indicates a medium size of difference between the groups. The difference between the lowest group with no mobiles, television, computer, cars, or bathrooms (45) and the highest group with all these accessories (59) is remarkable though; the parents with better economic possibilities can offer their children almost 14 percent points more marks than without the economic possibilities. It may be worth noting that 53.8% of those students, who have the access to all five accessories, live in Kathmandu Valley.

Data shows that when children have very few home possessions – 0 to 4 out of the 12 – the achievement level is statistically lower (<46) than if there are more than four (> 48). With nine to twelve possessions, the average score is very high (> 53%) compared with the national average. The same is evident of home accessories: With less than two accessory indicators met, the results are very poor (45−46%) and when there are two or more met, the results are remarkably higher (> 49%). With four or five indicators met the results are the best (57−59%).

SES and achievement

The socioeconomic status was formed on the basis of seven indicators which were all first dichotomized (see the opening section of the Section 3.3.3.3.1). 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 (as SES) and changed into the percentage of the maximum score (P_SES). The deeper description of the transformations is seen in Section 2.7.2. The P_SES represents the percentage of SES the student possesses; 100 means that the student has the highest SES possible measured with these variables and with these transformations (that is, all the seven indicators of SES 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 3.3.23 presents the relationship between SES of the students and the achievement.
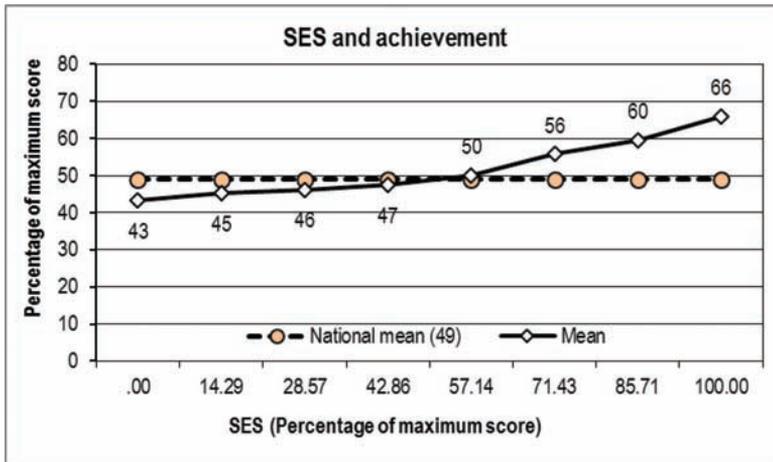
Figure 3.3.23 Connection of SES and achievement in Social Studies

Figure 3.3.23 shows a strict positive relationship between SES and the achievement; the correlation between the variables is $r = 0.33$ which is a significant association ($p < 0.001$). The differences between the SES groups are statistically significant (ANOVA, $p < 0.001$), the effect size is high ($f = 0.38$); that is, the highest and lowest group differ from each other remarkably. SES explains somewhat 12% of the student variation ($\eta^2 = 0.125$) which is not, though, a very high percentage. DTA suggests that three highest groups should be merged; there are not very many students in the highest SES groups. After combining the three highest SES groups, the mean of the students in that merged group is 62.

The dataset gives a strong signal that the socioeconomic status plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (23 percent points). This means that if the social economic standard of the lowest performing students was raised into a decent level, that is, in practice, that the problems in four out of seven indicators would be solved, the results in these groups would raise remarkably. Especially challenging is the situation in the families where both parents are illiterate or they both work in agriculture.

3.1.3.3.2 Working after school and achievement

Several questions were set in the student background questionnaire of the students' activities outside the school. Two of them are briefly handled here: Working after the school in a paid capacity and Participating in household work/chores. The values of the variables are divided into five categories: 0 (not 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). DTA shows that when the children do not work in a paid capacity *at all*, the results are notably above the national average (52). *If the students were working on a paid basis – even less than one hour, the results were remarkably lower than average (43–44).* The differences are statistically significant ($p < 0.001$) though the effect size moderate – most of the children do not need to work in a paid capacity. Working after school indicates that the family is poor and the extra salaries are needed and when the student needs to work more than 4 hours per day there is no time or energy to handle the school homework.

P_total=total_equated/85*100

Node 0	
Mean	50.073
Std. Dev.	15.385
n	13987
%	100.0
Predicted	50.073

q9e Before or after school: I work at a paid job
Adj. P-value=0.000, F=300.171,
df1=2, df2=13984

<= 0.0   (0.0, 2.0]   > 2.0

Node 1	
Mean	51.601
Std. Dev.	15.326
n	11270
%	80.6
Predicted	51.601

Node 2	
Mean	44.251
Std. Dev.	14.679
n	1461
%	10.4
Predicted	44.251

Node 3	
Mean	43.137
Std. Dev.	13.017
n	1256
%	9.0
Predicted	43.137
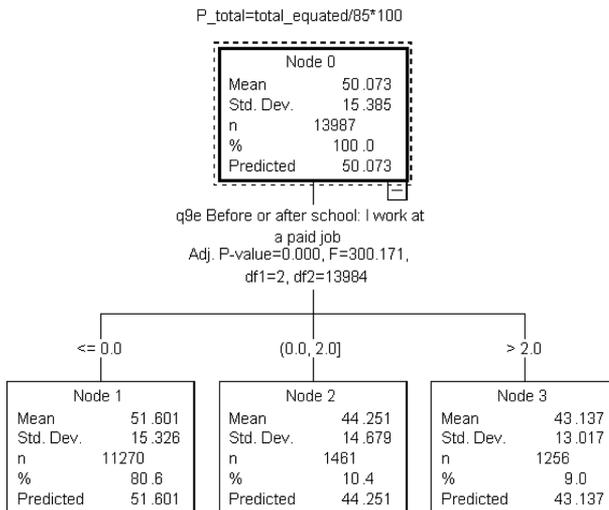
Figure 3.3.24 DTA of paid work and achievement in Social Studies

209

It is usual – and a supported – practice in families that the children take part in household work at home; this is part of the socializing process of the children. The DTA shows that when the child needs to spend more than 2 hours per day doing household work, the results are as poor (44) as if the student would need to work in a paid capacity (Figure 3.3.25). However, when the amount of household work is decent – practically less than one hour or one to two hours per day – the achievement level is close to the average (48). Differences are significant ($p < 0.001$) though the effect size is small or moderate – few children are participating in the household work for more than 2 hours per day.

P_total=total_equated/85*100

```
                    ┌─────────────────────────┐
                    │         Node 0          │
                    │ Mean          49.644    │
                    │ Std. Dev.     15.365    │
                    │ n             15015     │
                    │ %             100.0     │
                    │ Predicted     49.644    │
                    └─────────────────────────┘
```

q9d Before or after school: I do jobs
at home
Adj. P-value=0.000, F=159.758,
df1=2, df2=15012

| <= 1.0 | (1.0, 2.0] | > 2.0 |

```
┌──────────────────────┐ ┌──────────────────────┐ ┌──────────────────────┐
│       Node 1         │ │       Node 2         │ │       Node 3         │
│ Mean       50.892    │ │ Mean       48.330    │ │ Mean       44.164    │
│ Std. Dev.  15.717    │ │ Std. Dev.  14.266    │ │ Std. Dev.  13.365    │
│ n          10555     │ │ n          2703      │ │ n          1757      │
│ %          70.3      │ │ %          18.0      │ │ %          11.7      │
│ Predicted  50.892    │ │ Predicted  48.330    │ │ Predicted  44.164    │
└──────────────────────┘ └──────────────────────┘ └──────────────────────┘
```
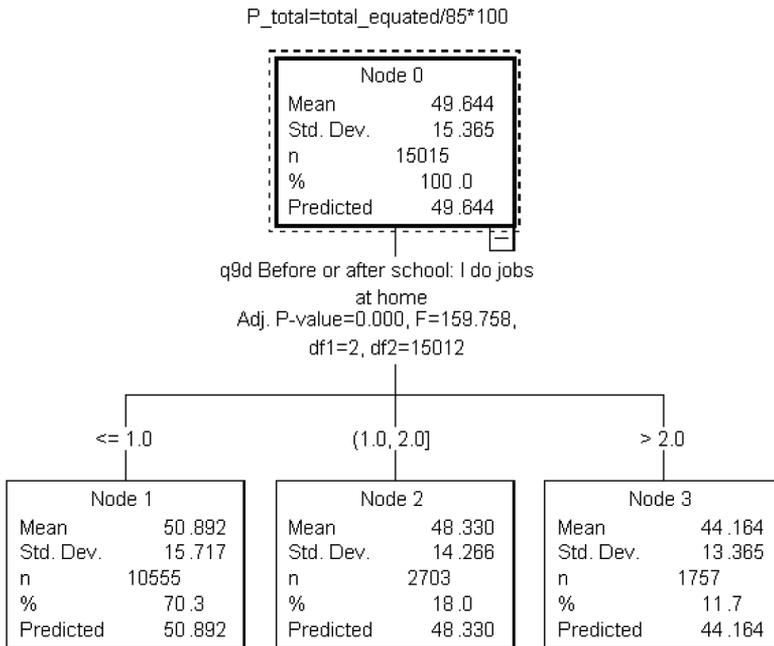
Figure 3.3.25 DTA of household work and achievement in Social Studies

The dataset gives a strong signal that working outside of school effectively reduces the school achievement of the student. The phenomenon is most probably connected with the poor economic situation in the family. Especially when the children need to work for more than 2 hours per day either in a paid capacity or unpaid doing household work, the achievement level is remarkably lower than in the families with no need for this much support from the 8th grader student.

3.3.3.3.3 Attitude and achievement

In the context of the Social Studies achievement assessment, the attitude tells us what the students think about Social Studies and its usefulness in their daily life

210

and future utility. 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 Social Studies achievement and attitudes toward Social Studies 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 attitudes and achievement may can be near zero, like in Macedonia (Kadijevich, 2008), Philippines (Wilkins, 2004), 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, the same shortened version of Fennema–Sherman Attitude Scales (FSAS, Fennema & Sherman, 1976) as is in use in several international comparisons, like in TIMSS and PISA studies, was in use. The original scales included nine dimensions but in these international comparisons only three is used with four items on each and two negative items on each of the first two dimensions (see in detail in Section 2.4.2). The names of the factors can be "Liking Social Studies", "Self-Efficacy in Social Studies", and "Experiencing utility in Social Studies" (compare naming in, e.g., Kadijevich, 2006; 2008). Factor analysis was used to identify the factors of the responses in FSAS and the negative items were reversed to make the whole test unidirectional. As in several countries in Asia, the expected factor structure cannot be found in Nepal (for a deconstruction of the test scales, see Metsämuuronen, 2012a; 2012b). Hence, only the total score is used to show the connection of attitudes and achievement. The relation between the attitude (divided into five groups of equal number of the students, that is, quintiles) and achievement score is shown in Figure 3.3.26.
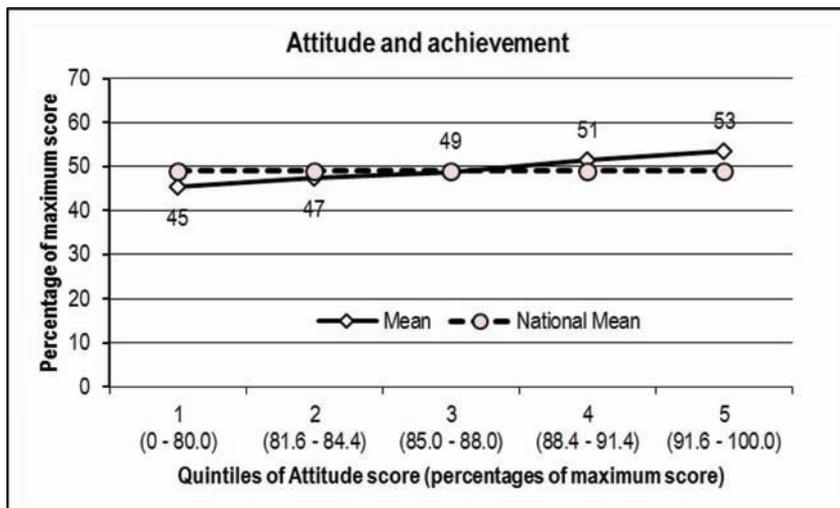
Figure 3.3.26 Connection of attitudes and achievement in Social Studies

Figure 3.3.26 shows that the difference between the lowest attitude group (<80% of the maximum scores) and highest attitude group (>91.6%) is notable (8 percent units). However, the real advantage comes in the two highest attitude groups (>88.4%). The correlation between the positive attitude towards Social Studies and achievement is $r = 0.07$ ($p < 0.001$); the effect size is moderate ($f = 0.19$).

A common challenge in using attitudes in explaining the achievement is that one does not know which is the hen and which the egg, that is, whether the attitudes are high because of the high performance or the performance is high because of positive attitudes. By explaining the variables both way with univariate ANOVA, it is more probable that the attitude is caused by the achievement (achievement explains 4.1% of attitude) and thus it is less probable that the attitude caused the achievement (attitude explains 3.4% of achievement).

Data gives a strong signal that the more positive the attitude is toward the Social Studies the higher is the achievement. Data also gives support to the fact that the positive achievement influences the positive attitude (rather than other way round).

3.3.3.3.4 Age and student achievement

In the Nepalese context, the age of the students attending to grade 8 studies varies remarkably. Some students have given their age as below ten years and some above 20. All the ages of the students below 13 were encoded as 'up to 13 years', and all students above 18 as '19 years & above'. The descriptive statistics of the mean in each year are given in Table 3.3.21 and visualized in Figure 3.3.27.

Table 3.3.21 Descriptive statistics of the students' achievement in different age groups

Students age	N	Mean	SD
up to 13 years	3560	53	15.4
14 years	5203	52	15.2
15 years	4032	47	14.6
16 years	2051	44	13.8
17 years	879	42	14.0
18 years	346	42	13.6
19 years & above	153	42	13.4
Total	16224	49	15.3
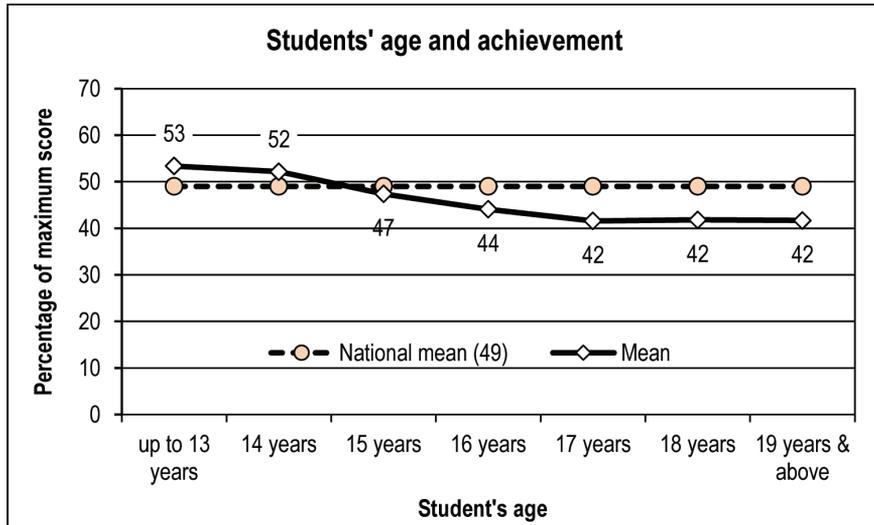
Figure 3.3.27 Connection of age and achievement in Social Studies

It seems evident that the best achievers are those students who are at the proper age for grade 8 studies (13 to 14 years old, scoring 53 and 52 respectively). The higher the age is – meaning that the students have either started much later than they should have, or they have doubled the classes – the weaker the results are. The achievement level is remarkable lower when the students are of age 17 or higher (42). Correlation between the variables is -0.248 ($p < 0.001$) indicating moderate effect size ($d = 0.50$). The ANOVA hints that the age (that is, the prolonged studies) explains the achievement level (10% of explaining) more probably than the achievement level the prolonged studies (5% of explaining). Another side of the matter is that it is good that these "over-aged" students are at school to learn; although they should have been identified at a much earlier age for give extra tuition or support.

Dataset gives a signal that the highest performance is with those students studying with their normal age group, that is, at the age of 13 and 14 years. Otherwise the achievement decreases as the age increases.

3.3.3.3.5 Help in study and student achievement

The relation between the help in 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 helpers are given in Table 3.3.22.

Table 3.3.22 Descriptive statistics of helpers of the students

Who helps	N	Mean	SD
Tuition	1326	51	15.7
Brother/Sister	3241	50	15.1
Mother	289	49	16.9
Father	939	49	17.4
No one	9756	49	15.1
Teacher	409	48	14.1

It seems that help is necessary for the students to gain better than average marks on the test. There is about 2% score difference between those who don't get any kind of help and those who receive (private) tuition. Brother's and sister's help seems to raise achievement score (50), which is more than mother's or father's (49) help. It is, though, noteworthy that there were a small number of the students (n = 289) who reported the help of their mothers. It is interesting to note that compared with Mathematics (Section 3.1.3.3.5) and the Nepali language (Section 3.2.3.3.5), most students do not get or want any help for their studies (61% or the students). This may indicate that Social Studies is not taken as the main subject in school. The (private) tuition brings the most advance for the student achievement in Social Studies (51). However, the advance is not at all at the same level as when comparing the illiterate and literate families, for example. It is also unknown whether the time spent on the tuition has something to do with the matter?

The dataset gives a slight signal that tuition given by a brother or sister raises the achievement level more than any other family members help. The highest achieving group are those who receive private tuition. It is possible that those with the private tuition also spent more time on their homework, which may explain the high score.

3.3.3.3.6 Availability of textbook and student achievement

The data shows that there are still some students who don't have Social Studies text book up to the last academic session. Table 3.3.23 shows the descriptive statistics of viability of the textbook of Social Studies and the achievement (mean).

Table 3.3.23 Availability of textbook of Social Studies and the achievement (mean)

Do you have textbook of Social Studies?	N	Mean	SD
Yes	15244	50	15.2
No	600	47	15.8
Total	15844	49	15.3

Out of 15,844 students who responded to the question, 3.8% of the students don't have a textbook available at school. The relation between textbook and achievement is significant ($p < 0.001$) though the effect size is small due to the small group size in one group. The difference in achievement is also not that big (3 percent points).

Data shows that 3.8% of the students lack the proper text book in Social Studies. The achievement level of these students is significantly lower than those who have access to the text book.

3.3.3.3.7 Homework given/checked and achievement

Homework is one of the methods to enhance teaching; it can be used as drill-, exercise-, and as an evaluation tool. When homework is systematically checked it may boost achievement levels. Now, the results are based on students' reports; it may be possible that within the same classroom some students have given a slightly deviant response than the other students from the same class. However, in the wide scope, the results are credible. Statistics related to homework given and checked is condensed in Table 3.3.24 and visualized in Figure 3.3.28.

Table 3.3.24 Homework given and the achievement

Status of homeworks	N	Mean	SD
Not given	100	44	12.9
Given Some day - not checked	118	44	15.0
Given Every day - not checked	218	43	13.9
Given Some day - Checked some day	1514	48	14.7
Given Some day - Checked every day	1324	49	15.8
Given Every day - Checked some day	2147	49	15.6
Given Every day - Checked every day	10249	50	15.2
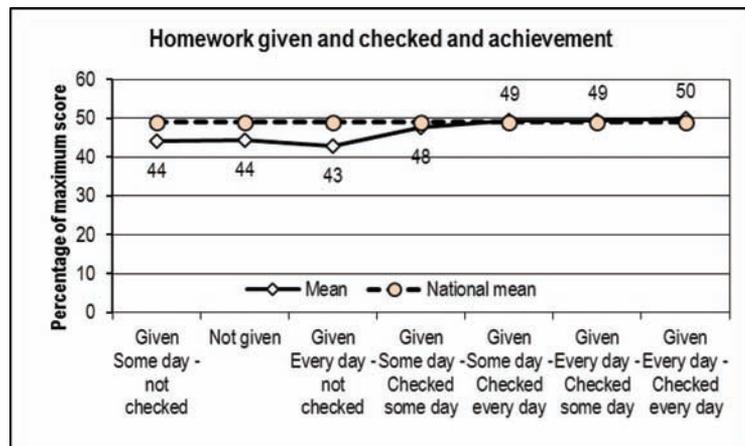
Figure 3.3.28 Connection of homeworks given and checked and achievement in Social Studies

215

It is evident on the basis of the dataset that if the teachers give homework every day to the students and it is checked – even only now and then, then student achievement is higher (50) than when the homeworks is issued non-systematically (43–49) or homework is not given or they are not checked at all (44). The differences are statistically significant ($p < 0.001$). Those groups with no checking or homework are, however, very small and hence, the effect size is small ($\eta^2 = 0.003$).

Dataset gives a solid signal that if the teacher issues homework daily and checks it, the achievement is higher than without checking or issuing of homework. By giving homework daily and by checking it, even not every day, the teacher may have raised the scores up to 6-7 percent point.

3.3.3.3.8 Activities in the school and student achievement

The activities of the students and 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 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' impression 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 learning environment for the students. International Studies like TIMSS and PISA give a specific emphasis to identify such indicators. In the NASA 2011 student questionnaire, five questions indicate the varieties of bullying that may happen in the school. All the questions were stemmed by the phrase "*which of the following activities happened happen in your school in the last month?*" The student responses are condensed in Table 3.3.25 and visualized in Figure 3.3.29. 'No (%)' indicates the percentage of the students response that describes there is 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. Alone the fact that 22% of the students mention that, during the last month, something of their own was stolen is an alarming sign of the system.

Table 3.3.25 Bullying and the achievement

Type of Bullying	No (%)	Yes (%)
Something of mine was stolen	78.2	21.8
I was made fun of or called names	84.9	15.1
I was hit or hurt by other student(s)	89.1	10.9
I was made to do things I didn't want to do by other students	85.2	14.8
Fellow students kept outside without involving me in activities	90.1	9.9
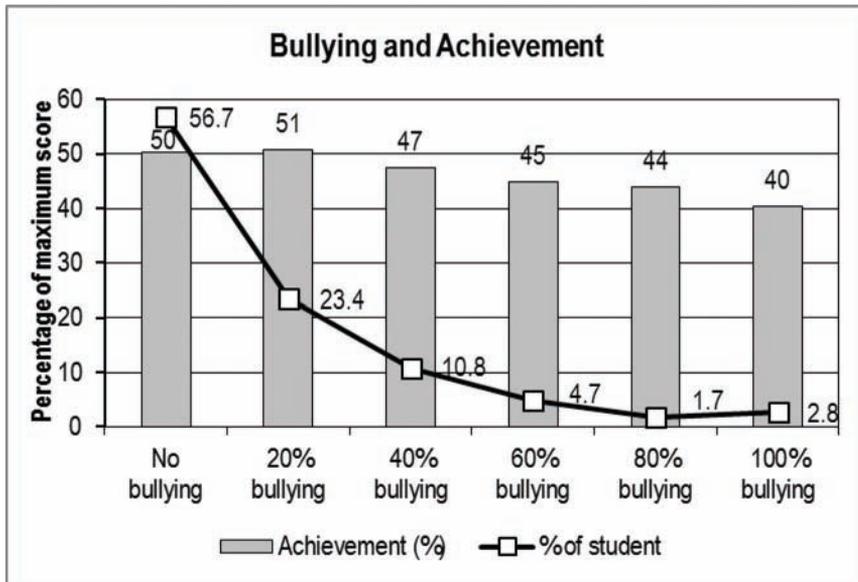
Figure 3.3.29 Relation of bullying and achievement in Social Studies

The sum of all five items is done as an indicator of bullying. Figure 3.3.29 shows the extent of bullying with the percentage of the students and achievement of the students in each category of bullying is shown in the line graph. 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 56.7% of the students did not encounter any bullying during the last month, one can infer that the remaining 43.3% *did* encounter at least one type of bullying. This is a remarkable number of the students. About 4% of the students are experiencing a severe kind of bullying (the sum of 80%- and 100% bullying). The trend is clear: the more bullying the less achievement. It seems, however, that learning outcomes are remarkably lower only with those 2.8% of the students who have encountered

extreme bullying including all five types harassments (40). Students who don't experience bullying and students who encountered extremely bullying of all five kinds have 10% achievement gap though very few number of the students who reported all kinds of bullying (n = 440). However, the difference is statistically significant ($p = 0.001$) though the effect size is small ($f = 0.15$). 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 the schools.

The dataset gives a signal that an alarmingly high number of the students (43.4%) have encountered bullying in schools within the last month. Though the phenomenon does not affect much the result except in the group of extremely bullied students, all possible efforts have to be done to root the phenomenon from the schools.

Positive activities in school

The activities that can boost the learning and achievement of the students are categorized as positive activities. Such positive activities about the school were asked from the students in two sets of questions collected in Table 3.3.26. The table shows the responses of the students in all four categories; the responses are in the 4-point rating scale anchored to fully agree and fully disagree.

Table 3.3.26 Students' response towards teacher- and school-related activities in the schools

Teachers- and Students activities	Respondents in %			
	Fully agree	Partially agree	Partially disagree	Fully disagree
Students get along well with most teachers	79.4	16.8	2.5	1.2
Most teachers are interested in student's well-being	85.5	10.8	2.1	1.6
Most of the teachers really listen to what I have to say	56.0	34.6	6.6	2.8
If I need extra help. I will receive it from my teacher	78.6	17.3	2.4	1.7
Most of my teachers treat me fairly	56.7	22.6	8.1	12.6
I like come and stay in school	91.1	6.3	1.2	1.5
Students in my school try to do their best	76.4	20.3	2.2	1.0
Teacher in the school care about the students	81.2	14.7	2.9	1.1
Teacher wants the students to do their best	90.5	6.5	1.3	1.8
Average	67.2	15.9	3.1	2.6

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. The overall result is that the feeling of the positive actions in the school relates positively with the student achievement. The correlation between the sum of nine positive activities and achievement is positive ($r = 0.19$, $p < 0.001$).

The percentage of maximum score of the positive activities are further categorized into quintile to do further analysis. The descriptive statistics are condensed in Table 3.3.27 and illustrated in Figure 3.3.30.

Table 3.3.27 Students' response towards teacher- and school-related activities in the schools

Attitude (%)	N	Mean	SD
20	3266	28	6.98
20-40	3072	41	3.21
40-60	3606	49	3.48
60-80	3110	58	3.04
80-100	3245	70	6.87
Total	16299	49	15.32
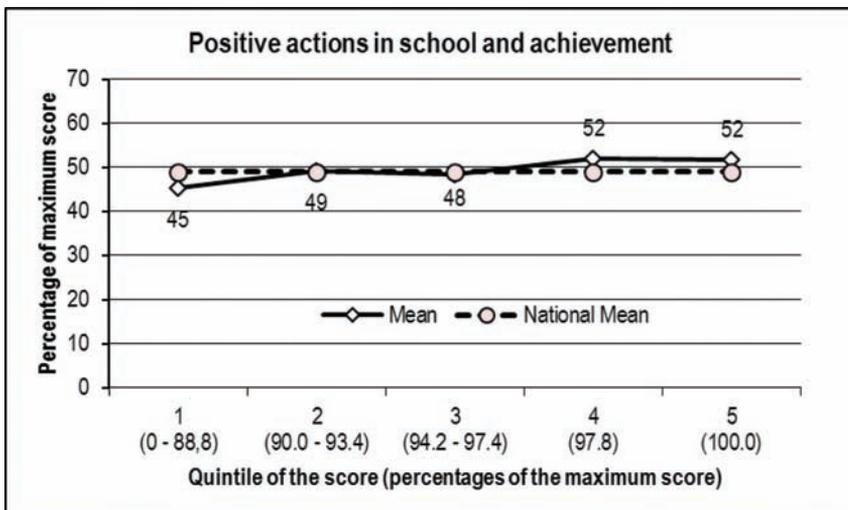
Figure 3.3.30 Relation of positive actions in school and achievement in Social Studies

219

The data shows that there is a positive connection between the students' feeling of the school activities and the achievement. The increase in achievement is directly proportional to the increase in intensity of such activities. After dividing the indicator into five groups using the quintiles, the differences between the groups are statistically significant ($p < 0.001$) however, the effect size is low or moderate ($f = 0.16$). The difference in achievement between the most positive group and the ultimately negative group is 7 percent points. Students' learning achievement is higher than average when students have experienced practically all positive activities.

Dataset gives a signal that when the students think that the teachers and schools' actions are very good the results are better than average (52). At the other extreme of feeling that such actions are ultimately negative, the results are below the average (45).

3.3.4 Discussion

The main findings of grade 8 Social Studies achievement are condensed as follows:

Basic results

- The achievement in Social Studies at grade 8 is Normally distributed.
- The learning outcomes are notably better than average in Economics (58) and Geography (57) and weaker than the average in politics (46).
- Students' ability to solve complex problems is low; only 34% of the maximum scores were reached. Students are much better in comprehensive type of questions (66), which may be explained by the educational system, which seems to be grasping the meaning of given things than solving novel problems.
- Students are good in recognizing the correct answer and in very fundamental knowledge/contents, true and false, matching texts, puzzles, selection of words and fill in the gaps. They are much weaker in reasoning, problem solving, and constructing the information.
- On the basis of the dataset it is not possible to say anything firm of the development in Social Studies between the years 2008 and 2011.
- From the international comparison viewpoint, the average achievement of Geography in Nepal is better than the international average.

Equity indicators

- There is a wide difference between the districts when it comes to the equal opportunities of children to reach the preset goals in Social Studies. The results in Bhojpur (38), Jajarkot (40), Sankhuwasabha (44) Kalikot, Morang and Dhading (45 each) are so low that raising the standard in these districts would raise the standard in the whole country.

- Except the Valley region, there is no difference between the Ecological zones. This is a good sign from the equity viewpoint. Students in the Kathmandu Valley (59) outperform the other students (47–49).

- Except the students in Valley (59), there are no practical differences between the Developmental regions (46–49). From the equity viewpoint, this is a good sign. However, the difference between the lowest performing area (Eastern region, 46) and the highest performing region (Kathmandu Valley, 59) is remarkable – around 13 percent points.

- Students in the institutional schools perform very well and the students in the community schools form two kinds groups of schools: high-performing schools and low-performing schools. The variety between the community schools is remarkable. Most probably this deviance can be explained by the student selection, however, this cannot be known on the basis of the dataset.

- Though the results are somewhat better in cities than in the rural area, the difference is not remarkably high. From the equity point, this is a positive thing.

- There are remarkable differences between the language groups in the achievement of Social Studies. The Dataset arises two difficult questions: What has been done differently – or have not done at all – in Central Tarai area where Gurung, Rai-, Tamang-, and even Sherpa students get very low scores? Why are the results in Tarai area lower than in other regions?

- In Social Studies, Madhesi- (46), Dalit- (47), and Alpasankhyak (47) students' performance is lower than the results of the other castes; the highest achievement is in the group of Brahman/Cheetri students (52). Dalits' performance is especially low in the Eastern Hill- (41), Mid-western Mountain and Hill- (42) and Far-Western Tarai (42) area. A positive signal is that Dalits' performance is high in Eastern- (52) and Central Mountain- (55) and Mid-Western Tarai areas (54).

- All in all, in Social Studies, the differences between the boys and girls are small. Boys are, though, slightly out-performing girls in all content areas of Social Studies. Differences are not wide though they are significant. Sex explains only 0.2% of the student variation; from the equity point of view the signal is positive. Differences are the widest in the Far-Western region; in the Central region there is no difference between the genders.

Selected explanatory factors

- The educational level of the parents strongly predicts the children's future achievement level in Social Studies. Especially harmful for the achievement of the Social Studies seems to be the situation that either the father or both parents are illiterate.

- If the father or mother or both were coming from an agricultural or related occupation, the students' achievement in Social Studies is significantly lower than with the other occupational groups.

- When the children have access to only a few home possessions (like their own table or room or dictionary) – 0 to 4 out of the 12 – the achievement level is statistically lower (<46) than if there are more than four (> 48). With nine to twelve possessions, the average score is very high (> 53) compared with the national average. The result is the same with home accessories (like mobiles, cars, bathrooms): With less than two accessory indicator met, the results are very poor (45–46) and when there are two or more met, the results are remarkably higher (> 49).With four or five indicators met the results are the best (57–59).

- The socioeconomic status (SES) plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (23 percent points). This means that if the social economic standard of the lowest performing students can be raised to a decent level, that is, in practice, that the problems in four out of seven indicators would be solved, the results in these groups may raise remarkably. Especially challenging is the situation in the families where both parents are illiterate or they both work in the agricultural field.

- Working outside of school effectively reduces the achievement of the student in Social Studies. The phenomenon is most probably connected with the poor economic situation in the family. Especially when the children need to work more than 2 hours per day either in a paid capacity or unpaid on household work, the achievement level is remarkably lower

than in the families with no need fot this much support from the 8th grader student.

- The more positive the attitude is towards Social Studies the higher is the achievement. Data also support the fact that positive achievement influences the positive attitude (rather than other way round).

- Achievement decreases as the age increases. The highest performance is with those students studying within their normal age group, that is, at the age of 13 and 14 years.

- Tuition given by the brother or sister raises the achievement in Social Studies more than when given by any other family members. The highest achieving group are those who receive private tuition though the differences are not wide. It is possible that those with the private tuition also spent more time on their homework, which may explain the high score.

- 3.8% of the students lack the proper text book in Social Studies. The achievement level of these students is significantly lower than those who have the access to the text book.

- If the teacher gives the homework daily and checks it, the achievement is higher than if this does not occur. By giving daily homework and by checking it, even not every day, the teacher may have raised the scores up to 6-7 percent point.

- An alarmingly high number of the students (43.4%) has encountered bullying in schools within the last month. Though the phenomenon does not much affect the result except in the group of extremely bullied students, all possible efforts has to be done to root the phenomenon from the schools.

- When the students think that the teachers and schools actions are very good, the results are better than average (52). At the other extreme, if feeling is negative, the results are below the average (45).

References for Section 3.3

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

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

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

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

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

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

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

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

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

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

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

Fullbright (2008). *A final report on national assessment of grade-V students*. Fullbright Consultancy & Cooperative Hands in restoration (CHIRAG). Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education.

House, J. D. & Telese, J. A. (2008). Relationships between Student and Instructional Factors and Algebra Achievement of the students in the United States and Japan: An Analysis of TIMSS 2003 Data. *Educational Research and Evaluation*, **14**(1), 101−112.

Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on Mathematics attitude. *The Teaching of Mathematics*, **9**(2), 41−51. Retrieved from http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf.

Kadijevich, Dj. (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.

Metfessel, N., Michael, W. B., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the Writing of behavioral Objectives. *Psychology in the Schools*. **6**, 227–231

Metsämuuronen, J. (2012a). Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies*, 4(3), September.

Metsämuuronen, J. (2012b). Comparison of Mental Structures of Eighth-Graders in Different Countries on the basis of Fennema-Sherman test**.** *International Journal of Psychological Studies,* 4(4), December.

Papanastasiou, C. (2000). Effects of Attitudes and Beliefs on Mathematics Achievement. *Studies in Educational Evaluation*, **26**(1), 27−42.

Papanastasiou, E. (2002). Factors That Differentiate Mathematics Students in Cyprus, Hong Kong, and the USA. *Educational Research and Evaluation: An International Journal on Theory and Practice,* **8**(1), 129−146.

PEDP (1998). Primary Education Development Project: An assessment. 1998. Kathmandu.

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.

225

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.

TIMSS (2009). TIMSS 2007 User Guide for the international Database. Released Items, Mathematics – Eight grade. IEA, TIMSS & PIRLS International Study Center. Lynch School of Education, Boston College.

Wilkins, J. L. M. (2004). Mathematics and science Self-Efficacy: An international investigation. *The Journal of Experimental Education*, **72**(4), 331−346.

Section **4**

Deepening Analysis of NASA 2011

4.1 Diversity and Educational Equity in Nepal

Dr. Sushan Acharya[1],
Ms. Sunita Shakya[2] &
Dr. Jari Metsämuuronen[3]

1. Tribhuwan University, Kathmandu, Nepal
2. NASA Unit, Educational Review Office (ERO), Ministry of Education, Nepal
3. Helsinki University, Faculty of Behavioral Sciences, Finland

Abstract

This article deepens the knowledge of NASA 2011 by analyzing the dataset from the viewpoint of diversity and equity related variables such as gender, geographic regions, caste, ethnicity, home language, and parents' occupation. The national data shows that students belonging to certain social groups are performing better than others. Presumably by virtue of their social background they are able to access and attain a higher standard of education. In a society which is traditionally stratified on the basis of social roles and functions some groups require education more than others. It can be explained by the fact that those who belong into such group can take education as social and cultural capital; the educational mechanism which is largely developed and run by the same, higher stratum, group suits more children from this group than from other groups. This resulted in a disparity across groups. This article, while exploring the magnitude of disparity in student achievement, reveals that although some groups continued to be high achievers, the gender disparity is in favor of boys across all the variables. The micro level analysis across all the independent variables showed a wide variation in the student scores. The results thus indicate that any programme intended to address the disparity and inequity in the education system demands a micro level understanding about the issues and contextual intervention because students from the same social group or the same sex may be a high achiever in one circumstance but a low achiever in another.

Keywords: Diversity, Equity, Gender, Socioeconomic status

4.1.1 Context

Nepal is a country of diversity where people belong to different geographic and ecological areas as well as social-, economic-, cultural-, religious-, and linguistic groups and sub groups. The sex ratio, as estimated by population Census 2011, is 94.41 male per hundred female (CBS, 2012). According to population Census 2011 (CBS, 2012, p. 4), there are 125 caste- and ethnic groups located in different Ecological zones. Religious groups also have a wide variety of representations. Although the majority of people speak Nepali it is estimated that 125 to 200 languages are spoken in Nepal (Yonjan-Tamang, 2009). Cultural and linguistic diversities are found even within the same ethnic and/or caste group. To a large extent, such diversities are determined by their place of residence. Importantly, very few places are dominated by one or other of the ethnic or caste groups. Therefore, school catchment areas are also mostly heterogeneous leading to a diverse student population.

In the above context, this section presents and analyzes the student achievement results from an equity and diversity perspective. It has been observed that often obtaining an education has been quite challenging for girls and for those who come from a minority group, or are of low economic status, or from deprived areas or who belong to linguistic backgrounds different from the one used as the medium of instruction. However, it is universally accepted that differences in gender-, social-, economic-, caste-, ethnic- and linguistic backgrounds should not be an obstacle in fulfilling educational potential (OCED, 2008, p. 2).

In general, equity is related to justice or fairness. In educational terms, it means that irrespective of a students' background they are entitled to equitable access to a conducive environment to the best utilize the education services in their existing situations. OCED (2008, p. 2) has defined three key policy areas which can affect equity in education: "*a) design of education systems/curriculum structure b) practices in and out of school, and c) resource allocation*". These policy areas are also applicable in the Nepali context. Student learning could improve from an effective school/home relationship, but weak support at home can hold back children from deprived backgrounds (OCED, 2008, p. 3). This observation indicates that parental education, especially mother's who are expected to spend more time with their children; and the household's socioeconomic background all play a key role in children's learning.

Six types of diversities have been considered (Ecological zone, Developmental region, school location, sex, caste, and ethnicity and language) to analyze student achievement. Of these, sex is broadly categorized into two groups – girls (N = 24,818) and boys (N = 23,743) and it is handled throughout the text as one key diversity. Ecological zones are represented by Mountain, Hill and Tarai,

but in this study, the Kathmandu Valley is treated as a separate region so that there are a total of 4 Ecological zones in total.[17] Another geographical division in Nepal is done on the basis of Developmental regions: five areas cover the country from the Eastern region to the Far-Western region. Rural and urban areas are also included under location. For the purpose of this report, caste and ethnicity are broadly categorized into five major groups namely Brahman/Cheetri Janjati, Dalit, Madhesi and Alpasankhyak (minority). Among the language groups, 11 major languages are addressed with separately whilst the rest are categorized as 'Other'.

4.1.2 Methodological solutions

The methodological solutions are mainly described in Section 2. Here the Sample and Methods used in this Section are highlighted.

4.1.2.1 Sample

A total of 25 districts from different Ecological zones were selected randomly in this study – Mountain 5, Hill 8, Valley 3 and Tarai 9 districts. Altogether 1200 schools with the same number of head teachers (HT) and teachers were selected randomly from different Ecological zones, location, and caste, ethnicity and language groups.[18] Out of all sample schools, 73 percent were community schools and 26 percent institutional (private) schools. Among the total sample of the students 83 percent came from community schools (girls 52 and boys 48 percent) and 17 percent from institutional schools (girls 45.6 and boys 54.4 percent). The subject-wise number of the students, schools and teachers is presented in the Table 4.1.1.

Table 4.1.1 Basic statistics of the sample (number of cases)

Subject	Students	Schools	HTs	Teachers
Nepali	16,350	419	415	415
Mathematics	16,033	421	407	407
Social Studies	16,299	413	402	402

---

[17] Throughout the book, the ecologically divided areas (Mountain, Hill, Tarai, Valley) are called *zones* ("Ecological zones") and the Developmentally divided areas (from Eastern to Far-Western) are called *regions* ("Developmental regions") to avoid confusion.

[18] The total number of schools in the analysis was, however, 1224 because in some of the schools all three subject were tested and hence there came three teachers' background questionnaires and three subsets of student papers.

Altogether 48,561 students participated in this assessment.[19]   Among them, 51 percent were girls and 49 percent boys. The highest proportion of the students who participated in the assessment came from the Hill areas i.e. 43 percent (43.7 girls and 41.4 boys), followed by the Tarai 38 percent (37.6 girls and 39.3 boys), 12.5 percent, the Valley (12.5 girls and 12.6 boys) and the Mountain areas 6.5 percent (6.2 girls and 6.7 boys). Most of the students came from rural areas (74 percent). Caste and ethnicity-wise, the highest participation came from Janjati (40.5 percent) followed by Brahman/Cheetri (38.4 percent), Dalit (9.9 percent), Madhesi (9.7 percent) and Alpasankhyak (1.5 percent). Language wise, the highest number of selected students (71 percent) spoke Nepali as their first language. Similarly, 2.6 percent spoke Magar, 2.4 percent spoke Tharu, 2.9 percent spoke Tamang, 1.1 percent spoke Newari, and 17 percent student spoke 'Other' types of languages as their first language. Less than 1 percent of the sample population spoke Urdu, Maithili, Rai, Gurung Sherpa and Limbu as their mother tongue.

4.1.2.2 Methods and tools

An equated mean of the total score is used to analyze data to see the diversities and gender equity aspects of information (see Section 2). The scores are presented as a percentage of the maximum score, that is, 100 would indicate that all the test items would have been successfully answered.

The main analytical tool is Univariate Analysis of Variance (ANOVA). While generating the tables, the valid number or percent is used for this analysis. Therefore the total number of the students may vary depending on the variables.

4.1.3 Results

4.1.3.1 Ecological zones, sex and student achievement

Among the total sample, the boys' score is slightly higher (48) than that of girls (46). Figure 4.1.1 shows the mean achievement scores of the students across the Ecological zones.

---

[19]   Missing values are excluded from this total number of student and number of girls and boys.
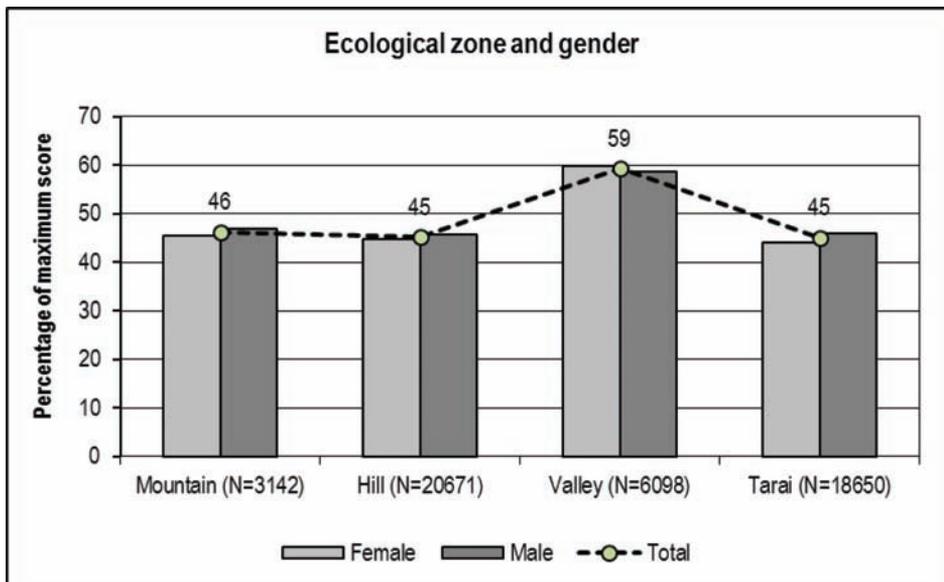
Figure 4.1.1 Mean achievement score of the students across the Ecological zones.

As shown in Figure 4.1.1, in total, students from the Valley performed better than those from the other Ecological zones. It is still the case that many students, particularly from the Dalit and Janjati communities in rural areas, are the first generation of their families to go to school. Therefore, education has not yet been established as part of their household culture. The absence of education as cultural and/or social capital significantly affects children's learning environment both at home and at school resulting in low achievement. More girls than boys are affected because the number of 'first generation schooler' girls is more than boys, which can be understood by the lower educational attainment rate of Nepali women.

With the exception of the Valley, girls' achievement is slightly below that of the boys in all Ecological zones. However, a micro analysis of the scores shows a wide variation in the achievement scores of girls and boys, and between girls and boys. This indicates that contextual understanding and measures, and qualitative investigation are required for the purpose of ensuring social and gender equity in education.

Figure 4.1.2 shows the subject-wise mean score of the students in different Ecological zones.
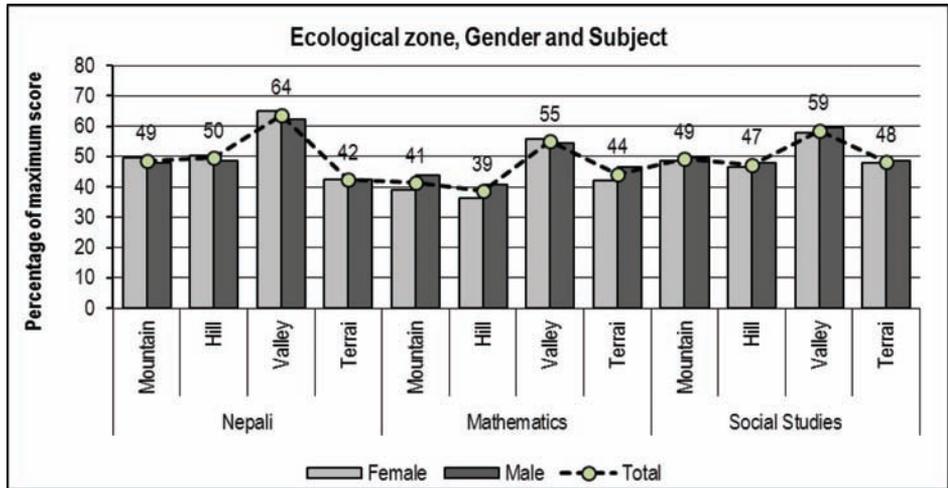
Figure 4.1.2 Gender-wise students' achievement across Ecological zones

A disparity is apparent across Ecological zones and between sexes. The achievement differs significantly ($p < 0.001$) in the Ecological zone and between sexes. The effect size $f = 0.25$ in the Ecological zone shows a moderate difference between the extreme zones; the difference between the boys and girls is small, though.

As the cumulative result of the Valley students is better so are their subject-wise scores. Valley girls did better in all subjects except in Social Studies. In Mathematics, there is 4 percent point difference between the scores of girls and boys. The effect size $d = 0.5$ shows a moderate difference between them in Mathematics. Except in the Valley, boys are ahead of girls in Mathematics. The effect size is $d = 0.09$ indicating a small difference between Valley girls and boys, which is statically also not significant ($p = 0.160$).

In Nepali, except in Tarai, girls have a higher mean score than that of boys. In the Mountain and Hill areas the difference is small ($f = 0.05$) and in Tarai there is no difference in the Nepali subject between girls and boys. In Social Studies, boys are ahead in all the regions. In general, girls' achievement score is higher in Nepali language whereas boys' score is higher in Mathematics and Social Studies. This has been the situation in Nepal. Girls are often found behind in Mathematics, but better in language (Parajuli & Acharya, 2008; Koirala & Acharya, 2005 & CERID/BPEP, 1999 cited by Koirala & Acharya 2005). Both in school and out of school factors including teacher attitude and expectation, bullying by the students in Mathematics classrooms, social mobility that made girls determine their future

professional ceiling, etc. contributed to girls' low achievement in subjects like Mathematics and science particularly in rural areas (Parajuli & Acharya, 2008; Koirala & Acharya, 2005; Lamichhane, 2003). Regarding ecological variation, factors including the presence of more educated parents and academic support positively influenced Valley students' score, which is discussed later in the appropriate section of this report.

4.1.3.2 Developmental region and student achievement

It is somewhat interesting as well as sad, that there are remarkable differences between the Developmental regions when it comes to student achievement as a whole. Remarkably, the lowest scores come from the students of the Eastern region (38) compared with the other regions (46 for boys and 45 for girls, Table 4.1.2).[20]

Table 4.1.2 Differences in achievement in developmental regions

Developmental	Mean		Std. Deviation		N	
region	Male	Female	Male	Female	Male	Female
Eastern	38.1	38.0	15.61	15.82	4492	5568
Central	47.8	47.6	19.58	18.94	4811	4846
Western	46.1	45.2	16.48	16.57	4409	4861
Mid-Western	41.5	41.3	16.85	16.65	2436	2791
Far-Western	43.5	40.0	16.21	16.40	1990	1664
Valley	50.7	52.9	16.40	16.59	1115	1332

The Eastern region was detected also in the teacher-wise- (see Section 4.3) and school-wise analysis (see Section 4.4) to be a specifically low-performing area. The question arises: what kinds of differences are there between the Eastern region compared with the other regions; what can explain why the results are so much lower in this particular region? In Tables 4.1.3a and 4.1.3b, two sets of variables are given to describe some indicators of the Eastern region compared with the other regions. Table 4.1.3a comprises the school-wise differences from the head teachers' questionnaire; Table 4.1.3b comprises the teacher-wise differences from the teachers' questionnaire. All variables in the tables show a statistically significant difference between the regions.

---

[20] The figures are based on comparing the Community schools only; the institutional schools are, practically speaking, somewhere else than in the Eastern region. Hence there is no idea of taking the students from the institutional schools on comparison.

Table 4.1.3a School-wise Differences between the Eastern- and other regions

Variables[1]	Eastern	other	Sig.[2] (2-tailed)	Cohen's d	
Seriousness: school leaving (number of periods, hour)	2.0	1.8	< 0.001	0.71	neg[3]
Seriousness: late in school	2.1	1.9	0.002	0.69	neg
Seriousness: absence (without reason)	2.2	2.0	< 0.001	0.68	neg
Frequency: absence (without reason)	2.7	2.4	0.015	0.66	neg
Average present	128.3	148.2	0.001	0.59	neg
Frequency: school leaving (number of periods, hour)	2.3	2.0	< 0.001	0.57	neg
Shortage: audiovisual material for Nepali teaching	1.5	1.7	0.038	0.54	pos
Frequency: late in school	2.6	2.3	0.001	0.45	neg
Grade 8 total Female teacher	1.1	1.5	0.009	0.45	neg
Shortage: audiovisual material for Social Studies teaching	1.7	1.8	0.001	0.28	pos

1) The variables are ordered on the basis of Cohen's d: the most remarkable differences are at the top of the list

2) Significance is based on t-test.

3) Because the scales in the variables were sometimes a positive one (a higher value indicates the more positive reality) and sometimes a negative one (a higher value indicates the more negative reality), this column tell which is the reality in the Eastern region compared with the other regions; neg = more negative reality, pos = more positive reality.

The head teachers in the Eastern region seem to understand the educational challenge: they inform us that the absence of the students is more frequent and a more serious problem, as are students arriving late for school. Practically then, the head teachers see a problem with student behavior/attendance.

Table 4.1.3b Teacher-wise differences between the eastern- and other regions

Variables[1]	Eastern	other	Sig. (2-tailed)[2]	Cohen's d	
Ext. class observation: Other; get advises to improve	1.34	0.77	0.116	1.63	neg[3]
Ext. class observation: Other; felt benefitted.	1.16	0.73	< 0.001	1.34	neg
Ext. class observation: School supervisor; felt benefitted.	1.31	1.00	< 0.001	1.08	neg
Ext. class observation: School supervisor; get advises to improve	1.47	1.17	0.023	0.96	neg
Teacher: Age of the teacher	39.1	35.1	< 0.001	0.85	neg
Teacher Training: academic I.Ed./B.Ed/M.Ed	8.9	3.3	0.269	0.74	pos
Ext. class observation: Resource person; get advises to improve	1.3	1.1	< 0.001	0.69	neg
Limit teaching, Students: Uninterested students	2.5	2.8	0.033	0.65	neg
Ext. class observation: Resource person; felt benefitted.	1.1	1.0	0.001	0.59	neg
Sch: Parents' activity in school activities	3.7	3.4	0.012	0.52	neg
Interaction with Teacher staff of improving teaching-learning; how many times during the year	5.9	6.8	0.001	0.47	neg
Sch: Parents' support for student achievement	3.4	3.2	0.205	0.44	neg
External class observation: Head Teacher; get advises to improve	1.1	1.0	0.007	0.42	neg
Sch: Students' desire to do well in school	2.9	2.7	0.027	0.40	neg
Teacher: Have used the Teacher's guide	1.2	1.3	0.049	0.39	neg
Sch: Students' regard for school properties	3.5	3.3	0.002	0.38	neg
Teaching: asks students to do: explaining the answers	3.3	3.5	0.019	0.37	neg
Teacher: Have you got the teacher's guide of the subject you teach	1.3	1.4	0.005	0.33	pos
Sch: Teachers' understanding of school's curricular goals	2.6	2.4	0.923	0.33	neg
External class observation: Head Teacher; felt benefited	1.0	1.0	0.023	0.31	neg
Sch: Teachers' degree in success in implementing the schl's curric.	2.6	2.5	0.059	0.22	neg

1) The variables are ordered on the basis of Cohen's d: the most remarkable differences are at the top of the list
2) Significance is based on t-test.
3) Because the scales in the variables were sometimes a positive one (a higher value indicates the more positive reality) and sometimes a negative one (a higher value indicates the more negative reality), this column tells which is the reality in the Eastern region compared with the other regions; neg = more negative reality, pos = more positive reality.

The teachers in the Eastern region seem to have remarkably less external classroom observation from supervisors, instructors, and resource persons or head teachers. They also seem to benefit less and receive less advice to improve upon their situation. The teachers in the Eastern region are also older. From the teachers' viewpoint, the students seem to be uninterested and parents are less active in educational matters. It seems also that there are less interactions within the school between the teaching staff, and the teachers seem to have more difficulties in understanding and implementing the curricular goals compared with the other regions.

4.1.3.3 School location and student achievement

A location-wise comparison shows 7 percent points difference between the urban and rural students with the urban students achieving the higher score (52 with girls' 51 and boys' 52) and the rural the lower (45 with girls 45 and boys 46). The differences are seen in Figure 4.1.3.
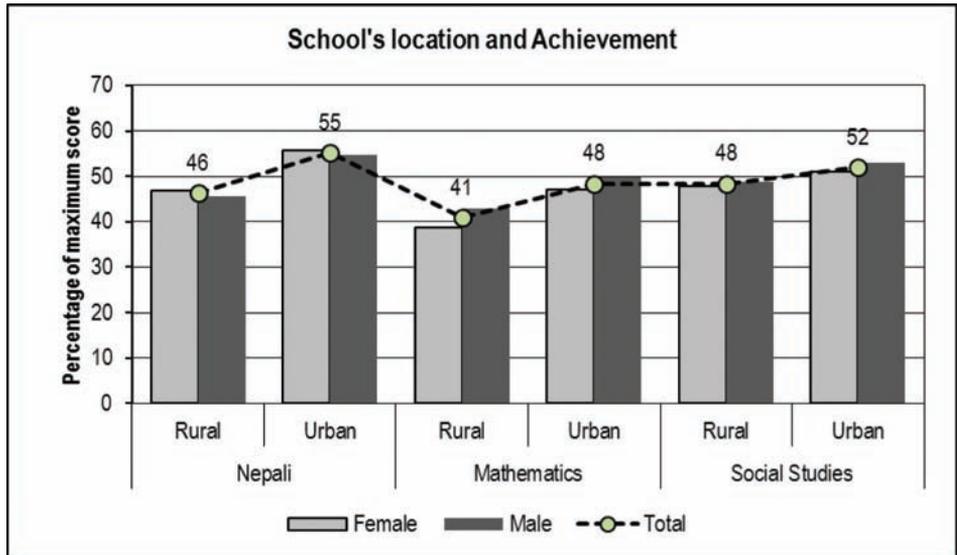
Figure 4.1.3 Subject-wise achievement across location

In Nepali, the girls scored higher than boys in both rural and urban schools, which is reverse of the scores in Mathematics and Social Studies.

4.1.3.4 Caste and ethnicity and student achievement

A social group wise comparison shows that Brahman/Cheetri students are ahead of others (see Figure 4.1.4). Gender disparity in the mean score (girls' 49 and boys' 50) is not prominent in this group. There is also no gender disparity in the mean score among Janjati students. Dalit and Madhesi boys did better than their female counterparts. Within the Alpasankhyak, girls are better than boys. Regarding Madhesi girls, due to limited academic support, restricted mobility, very few female teachers as role models and son's education being given presidency over daughters, gender disparity in score gradually widens in upper grades (Parajuli & Acharya, 2008).
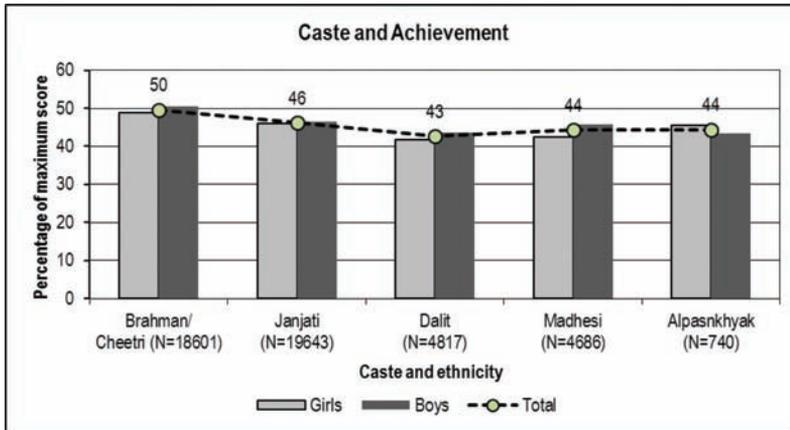
Figure 4.1.4 Caste- and Ethnic group-wise achievement

Figures 4.1.5a, 4.1.5b, and 4.1.5c show the subject-wise achievements across the castes and ethnicities. Subject wise comparison shows that in Nepali, after the Brahman/Cheetri group, the Janjati obtained the highest score (Figure 4.1.5 a). In the same subject girls' mean score is higher than the national average but for boys it is slightly lower than the national average. Dalit and Alpasankhyak are at the same level and Madhesi obtained the lowest. Among Madhesis, the achievement of the girls' is lower than that of the boys. These findings challenge the general perception that Janjatis are behind Nepali speaking children because of the language barrier. This, however, requires deeper examination before drawing any conclusion because many non-Nepali speaking children drop out in early grades (Acharya & Giri, 2009).
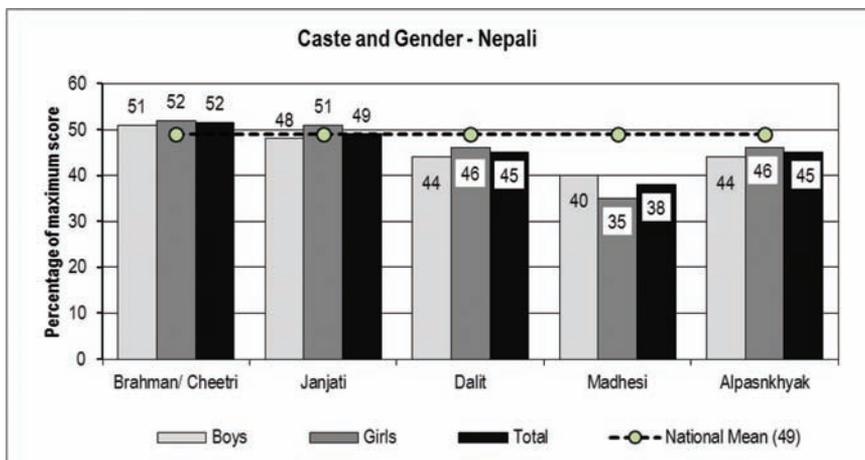
Figure 4.1.5a Students' achievement score in Nepali across the castes and ethnic groups

In Mathematics, boys from all groups except Alpasankhyak are ahead of girls. Madhesi and Brahman/Cheetri students' score is above the national mean score (Figure 4.1.5b). Madhesi students (both girls and boys) are far ahead among the other groups. Mathematics score of Janjati girls, Alpasankhyak boys and girls, and Dalit boys and girls is lower than the national mean score.
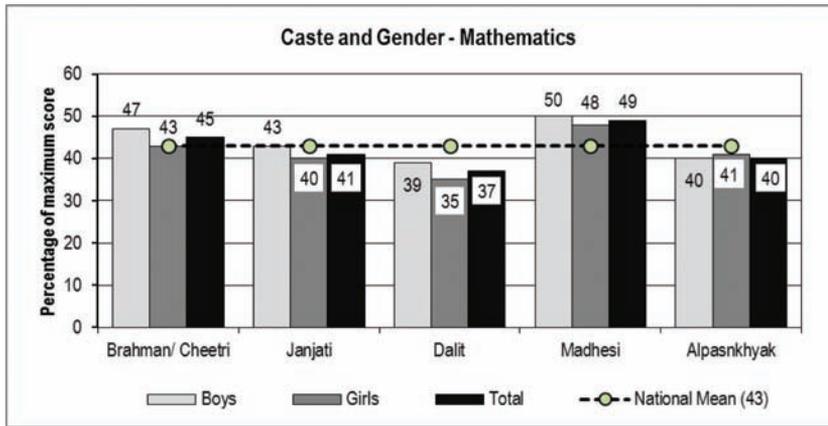
Figure 4.1.5b Mathematics achievement across the castes and ethnic groups

In Social Studies, except for Brahman/Cheetri, all the groups scored below the national average (Figure 4.1.5c). The difference in the average score is significant ($p < 0.001$) however the effect size is moderate ($f = 0.14$) and the caste and ethnic groups explains only 2% of the student variation ($\eta^2 = 0.020$). Among Alpasankhyak, there is no difference between boys and girls ($p > 0.050$).
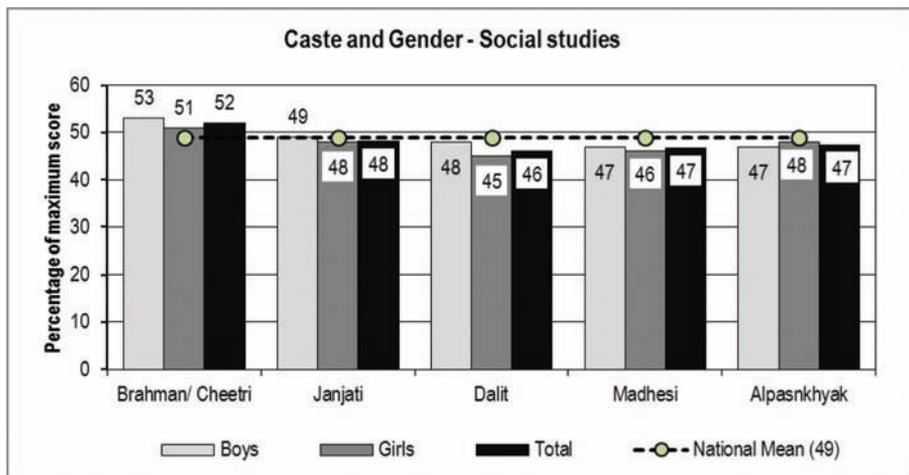
Figure 4.1.5c Social Studies achievement across the castes and ethnic groups

4.1.3.5 Home language and student achievement

Figure 4.1.6a shows that students with Urdu as their mother tongue obtained the highest score in Nepali. Newari, Tamang and Nepali speakers respectively followed them. The data gives a signal that the Nepali speakers do not necessarily outscore others in Nepali studies. Among Sherpa language speakers boys are above the national average but girls are just below it. Both boys and girls from Magar, Tharu, Rai, Gurung, Limbu and other language groups are below the national average. Maithili speakers obtained a significantly lower score than that of the other language speakers and of the national average. This can be explained by the fact that Maithili is one of the most highly developed languages of Nepal. Therefore the practice has been to use Maithili in the classroom transaction where this language is used and understood by many students and teachers in Tarai (MLE Research & Review team, 2009). Other subjects can be transacted in Maithili but not Nepali as a subject. This perhaps provided less opportunity to practice the Nepali language and thereby, contributed positively to the scores of other subjects but not Nepali as a subject.
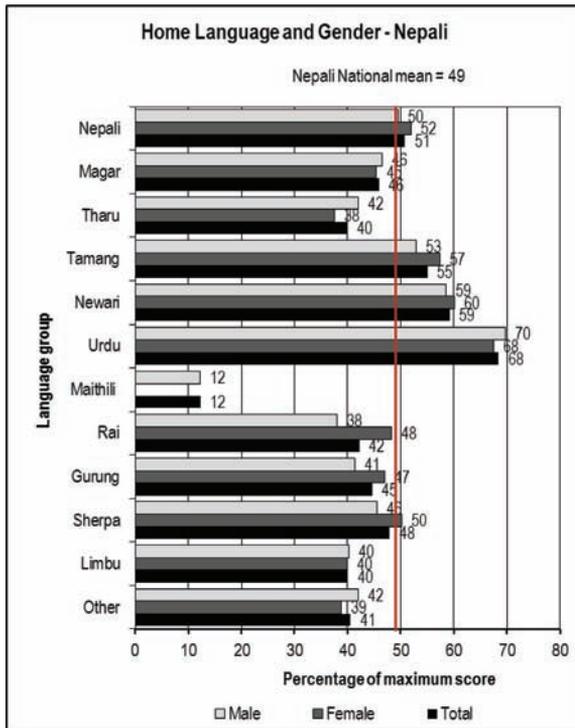
Figure 4.1.6a Language group-wise student achievement in Nepali

Figure 4.1.6b shows that, in Mathematics, among the language groups, Sherpa and Newari speakers achieved the highest scores. In these groups, gender disparity is in favor of girls. Among all language groups except Limbu the situation is reversed; in the Limbu group, the mean score of girls and boys is equal. The Mathematics score of Magar, Tamang, Urdu, Rai and Limbu speakers is below the national average. The achievement score of boys in all language groups is higher than the national average. But for girls from all language groups except Sherpa and Newari, the scores are below the national average.
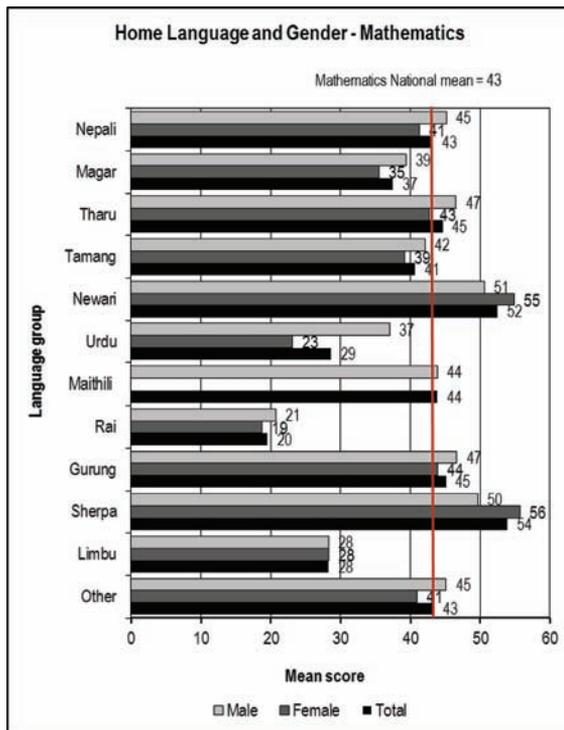
Figure 4.1.6b Mathematics achievement across language group

Figure 4.1.6c shows that in Social Studies, the Maithili speakers' score is far ahead of the national average. As mentioned above, this may be the effect in the use of the mother tongue as the language of instruction. Sherpa, Newari, Gurung and Nepali speakers follow Maithili speakers and are all still above the national average. Urdu speaking boys' score is also above the national average. However scores of other language groups are below the national average in this subject. Comparison between sexes shows that girls are better than boys in Social Studies

241

in Sherpa, Newari, Tamang and Magar language groups. This research thus warns to be cautious when generalizing the general assumption that girls are usually better in subjects like language and social studies.
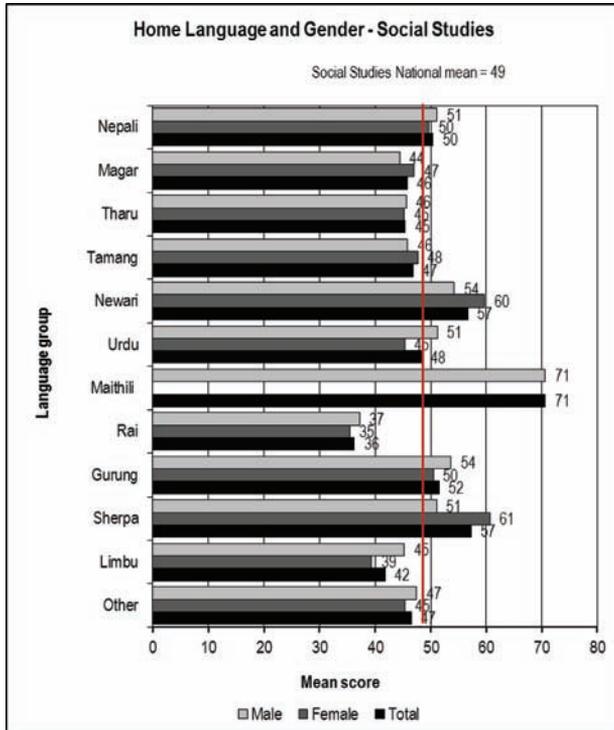
Figure 4.1.6c Social Studies achievement across language groups

4.1.3.6 Diversities, other related factors and achievement

Equity and diversities in the society are usually connected with some other related factors affecting the students' achievement. Some of these related factors, such as parents' education and occupation, household obligations and working in a paid capacity while studying, and bullying in the school, are tackled briefly in this Section. These matters are handled in detail in Section 4.5.

4.1.3.6.1 Family related factors and students' achievement

Parents' occupation and children's achievement

Parental occupation is considered as an independent variable to analyze the effect of occupational background in children's achievement. Depending on the mother's and father's occupation, children's achievement scores are analyzed against different diversity factors (sex, ecology, location, caste/ethnicity and language).

Figure 4.1.7 shows that most (70.1 percent) mothers are involved in agriculture, very few (2.5 percent) are in the teaching profession. Six percent of them are engaged in the service sector and 10 percent in business. Although a large percentage of fathers are involved in agriculture, father's ratio is lower (44 percent) than that of mothers in this occupation. With high male migration and low educational attainment among women leaves more women at home to take the charge of the farm work alone. The higher involvement of fathers in the service and business sectors, opportunities for which are available more in the urban areas supports this explanation. A large segment (74 percent) of the students who participated in this research belonged to rural areas, and Hill students outnumbered (43 percent) students from other regions. The main occupation in both places is agriculture therefore it is understandable to have more parents working in this sector.

Among caste and ethnic groups Dalit parents are mostly in agriculture compared with the others.

Fathers, when treated as one single group, tend to be involved in the services, business and teaching sectors. As with mothers, fewer fathers (4 percent) are involved in teaching. After agriculture, 28 percent of fathers are involved in the service sector and 15 percent in business.
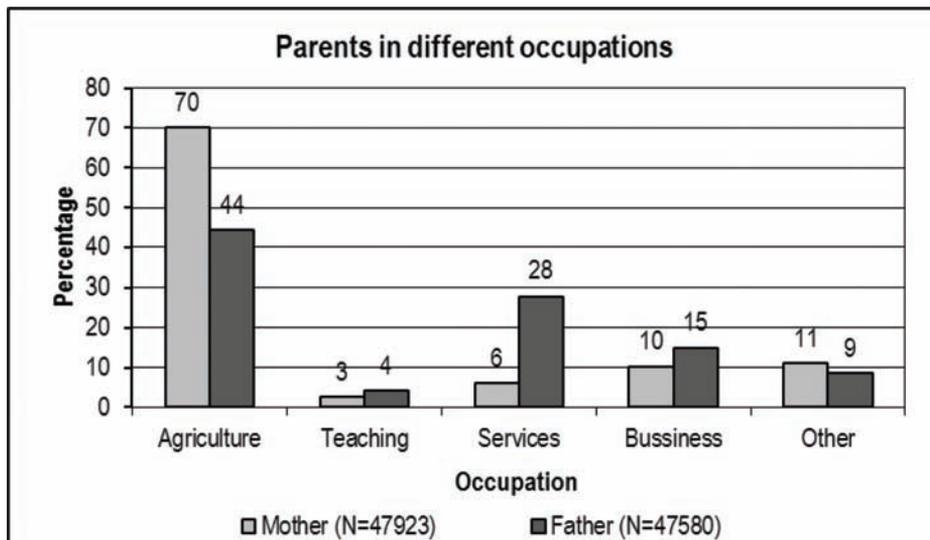
Figure 4.1.7 Mother's and father's occupation

In the Valley, 27 percent of the mothers and 13 of the fathers are involved in agriculture. Business is the second most popular occupation that mothers in all Ecological zones except in Valley are involved in. Valley mothers are noticeably more employed in the service sector, which is the fourth most popular occupation among mothers in other regions. However, among Valley fathers, the service sector is the second priority and business the third. Location-wise distribution also shows, in rural areas, there are 28 percent points more mothers in agriculture than fathers. However, in the non-agriculture sector it is the fathers who outnumber mothers by 28 percent points. In Urban areas though fewer parents are involved in agriculture mothers are still 19 percent points higher than fathers. In this area involvement of both mothers and fathers is higher in the non-agriculture sector with 59 percent and 78 percent respectively.

In language groups, only Newari mothers are ahead in teaching, service and other types of professions, whereas Urdu speaking mothers and fathers are ahead in business. It is because within Newari speakers particularly one group is more concentrated on agriculture (i.e., Jyaapu) and others are traditionally either service holders or traders. Among fathers, most of Maithili (75 percent), Limbu (70.5 percent), Tharu (68.4 percent), and Rai (68 percent) speakers are involved in agriculture. Among the language groups the highest percentage of Newari, Urdu and Gurung speaking fathers work in the service sector, i.e. around 37 percent in each group.

The Univariate ANOVA is used to see whether parents' occupations and students' achievement scores correlate. The significance of $p < 0.001$ of mother's and father's occupations, as the main effect of the independent variable, shows that there is a significant difference. Hence, it seems that certain types of parent's occupations have more effect on children's education than others. When mothers have an 'Other' type of job, students have the highest mean score (55) and those students whose fathers are in the teaching profession have the highest score (51). The differences are seen in Figures 4.1.8a and 4.1.8b.
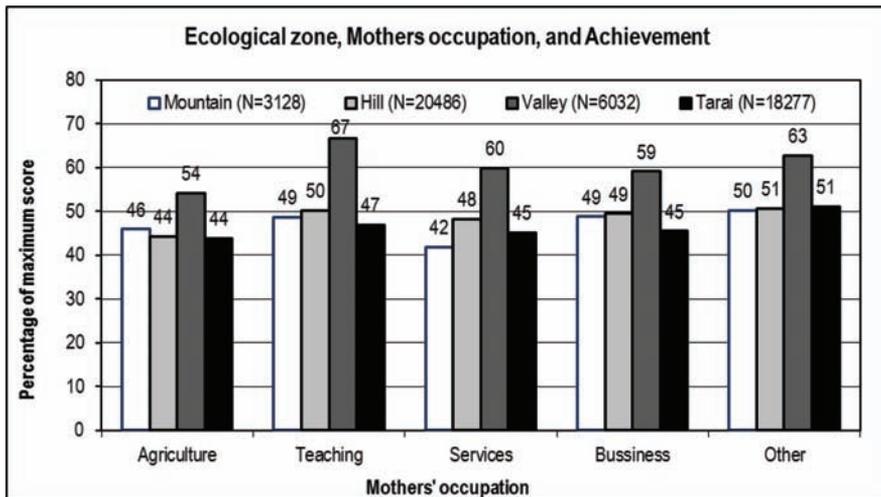
Figure 4.1.8a Mean achievement score of the students by mother's occupation
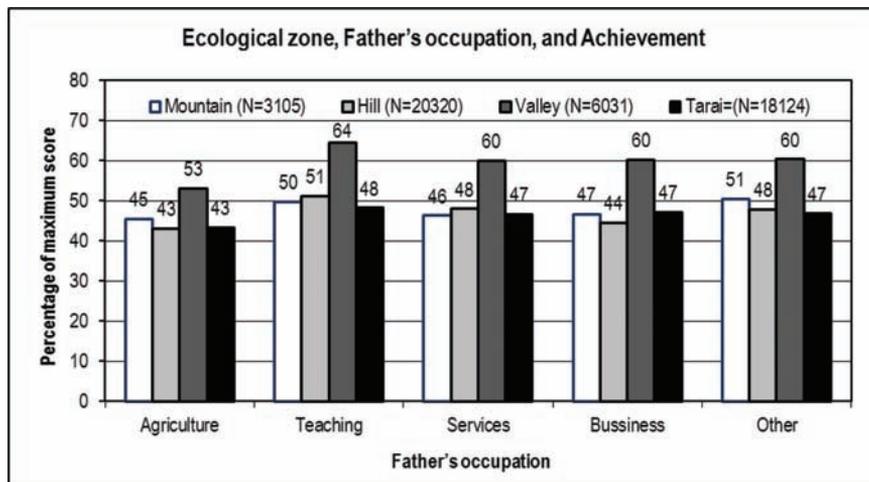
Figure 4.1.8b Mean achievement score of the students by father's occupation

Figures 4.1.8a and 4.1.8b show that, generally speaking, the mean score of the students whose parents are engaged in agriculture is the lowest with the exception of the Mountain areas where the mean score of children whose mothers work in the service sector is lower (42). In general, the mean score of the students whose mothers are engaged in non-agricultural work is higher (55). Whereas the mean score of the students whose fathers are in teaching is higher (51) than those whose fathers are involved in other occupations including agriculture. Nepali households particularly those engaged in agriculture (either as farmers or as farm laborers) rely heavily on children's help. This is true in the Hill, Tarai and Rural areas alike unless the family are large scale farmers. High absenteeism particularly among rural students during plantation and harvesting seasons in schools makes it difficult for these students to catch up their lessons. Students in this situation are most likely to achieve low scores.

On the other hand, the Valley students, irrespective of their parents' occupation, achieved higher scores than those from other regions. Here the mean score of the children of teachers (both mothers and father) is the highest than the children of parents involved in other occupations. Teachers do not require children's support at home like those involved in agriculture and business. Moreover, since the parents themselves are engaged in the education sector they are more likely to prioritize their children's education. In the Valley and the Tarai the mean scores of children whose fathers hold occupations other than agriculture do not vary. However, the variation in the mean score is noticed in the Mountain and Hill regions. In all Ecological zones, the difference is statistically significant ($p < 0.001$).

Across the castes and ethnic groups, Brahman/Cheetri scored the highest and Dalits the lowest. Dalit parents' participation in agriculture is the highest and hence there is a high chance that Dalit students have less time available for study. In all the castes and ethnic groups, the achievement scores of the students with mothers and father's having a similar type of occupation within their respective groups does not vary much. For example, the students whose parents are involved in agriculture obtained the lowest mean score within their groups except among the Dalits. Likewise, within the Dalits, the students whose parents are teachers obtained the lowest achievement score. This type of response requires further examination because in this assessment 'teacher' category was not further broken down. Therefore she or he could be a teacher of early childhood care and education (ECED) or high level. Capability to provide academic support to the children can vary extensively between different levels of teachers.

When the total score within the same language group is analyzed, it was found that, irrespective of their mother's and father's occupations, students' mean achievement score is the same. Compared across the language groups, the children from Sherpa mother tongue with mothers involved in teaching obtained the highest (69) mean score. Likewise, the children from Rai mother tongue with mothers involved in agriculture obtained the lowest score (26). With reference to father's occupation, the highest mean score (73) is obtained by Urdu speaking children with fathers having undefined occupation, whereas Maithili speaking children with fathers involved in agriculture obtained the lowest (23).

The significance level of the different language speaker, with respect to their mother and father's occupation, is lower than $p = 0.050$ level for Nepali, Magar, Tamang, Newari, Rai, Limbu and other language speaking groups. In these groups, parents' occupations seem to be connected to their children's achievement. In the case of Urdu, Maithili, Gurung and Sherpa speakers, there are no statistical differences in student achievement between the occupational groups.

The findings above show that mother's and father's occupation is more likely to determine students' achievement score than their language or caste and ethnic group. Further, the occupation is determined more by the place of residence than any other factors. However, micro analysis shows a variation which needs to be recognized and further explored. For example, students belonging to non-Nepali speaking mother tongue groups and whose help is required at home or on the farm by virtue of having parents involved in agriculture are less likely to have been exposed to the language of instruction in their schools. In this situation understanding the classroom transaction and curriculum can be more challenging.

Availability of support and student achievement

Teachers are the first choice from whom students seek support (Figure 4.1.9). This practice is more popular in Mountain with 52 percent doing so. Many students study independently without anybody's help. However, the ratio ranges, for example, from 17 percent in the Valley to 24 percent in the Hill. The rest either seek support from a family member or receive private tuition. Interestingly, in the Valley, siblings (34.2 percent) are relied on more than teachers (26 percent) or private tuition for academic support. Here 10.2 percent (10% girls and 11% boys) seek private tuition. Private tuition is more popular among Madhesi boys (15%) than Madhesi girls (7%) and any other groups. It is more popular in the Urban areas than in rural areas with 10- and 6 percent respectively accessing this service.

247

Help in students' learning in percent
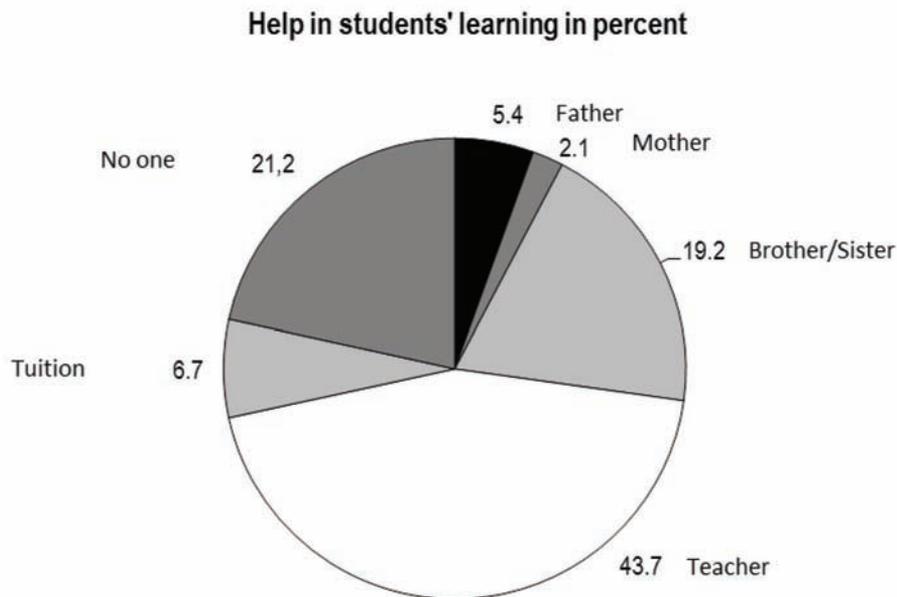
Figure 4.1.9 "Who gives you help when you do not understand what you have read?"

Generally speaking, the parents' role in their children's learning is reportedly low (7.5 percent). Those who do seek support approach fathers (5%) more than mothers (2%). A major reason must be the difference in the levels of education that mothers and fathers attained, or it could be the gendered perception about the roles of men and women in the household. This observation is supported by the fact that in the Valley, where more mothers are educated, more students are supported by their mothers. However, in the Valley, the mean score of those who get support from fathers is higher (64) than those who get support from other sources, including mothers. In the community schools, though, the score of those helped by the fathers was somewhat lower (59) than those helped by mothers (59.8), whereas in the private schools it was the opposite. In private schools, those helped by fathers obtained higher (65.5) scores than those helped by their mother (62). Interestingly, in the Valley, the students who do not get support from anyone score better (60 mean score) than those who get help from mother, siblings, teachers or tuition centers.

Particularly in the Valley and urban areas more parents are involved in the non-agriculture sector, which usually demands more time outside of the home. In this case, it is more likely for them to seek outside help in supporting their children in their studies. Abundant tuition centers also encourage parents to send their

248

children to private tuition centers on additional support. Additionally, heightened concern about their children's achievement scores and a competitive culture among Valley and Urban dwellers might also tempt parents to send their children to the private tuition centers though help from the tuition center may not necessarily increase the student's mean score. In Tarai as well, where private tuition is popular, students' achievement score is equal whether they join the tuition center or they studied independently. However, in Mountain and Hill areas achievement scores of those who joined a tuition center is higher. In rural areas there was a positive impact on students' scores for those who received support from siblings or teachers.

Findings show that, in the Valley area, parents' support plays a vital role in their children's education. Among those who obtained support from parents, the gender disparity is in favor of girls. The mean difference is higher (6 percent points) when students get support from mothers. A significant difference is found in all the Ecological zones at the significant level $p < 0.001$ when students got support from any source. Likewise, in urban areas, students who get support from their parents and enrolled in a tuition center achieved the highest mean score. However, students' mean score does not vary much when they obtain support from their brothers/sister, teachers or study independently.

In Brahman/Cheetri, Janjati and Alpasankhyak groups, girls, supported by mothers, obtained the highest scores, showing the positive effect of mother's support. This applies more to Alpasankhyak with both girls and boys supported by mothers achieving the highest score. Brahman/Cheetri and Alpasankhyak students as well as Dalit girls getting support from teachers obtained the lowest score within their respective groups. It is possible that those students' source of support was the teacher only, which was not enough for them. This finding however also makes one think about ongoing issue related to the quality of teaching learning and different effects of Nepali classroom practices on children with multiple disadvantages.

For Janjati students and Dalit boys who are supported by their fathers obtained the lowest in their respective groups. However, mean scores of Madhesi girls and boys who are helped by either parent are the lowest within their group. This finding shows that, irrespective of caste or ethnicity, academic support, wherever it comes from, plays a significant role in girls' and boys' achievement and the ANOVA also shows it statistically significant at $p < 0.001$ level.

Within language groups, the mean score of boys from Nepali, Magar, Tharu, Newar, Urdu, Gurung, Sherpa, Limbu and 'Other' language groups who are enrolled in tuition centers is the highest. Teachers' support does not have a significant effect in achieving the highest mean score except among Tamang boys. Particularly among Nepali, Magar, Tamang and Limbu speakers the mean score of girls who

receive support from their mother is the highest. These findings also make one think about the adequacy of teaching learning support that children get from Nepali schools.

4.1.3.6.2 Household obligations, paid work, and students' achievement

Household obligations

Among the sample more girls are involved in household chores. Out of the total sample, 74 percent girls and 69 percent boys said that they are engaged in household chores to some degree. In all regions girls work more hours than boys, however it is only in the Valley that boys who do not work is much higher (48%) than girls (35%) who do not work. Valley students work comparatively fewer hours than their counterparts from other regions although it is commonly known that many children who go to community schools in the Valley are domestic helpers. However, more than fifty percent of the sample schools in the Valley were private; parents, who send their children to private schools irrespective of the nature of their occupation, tend to comply with the strict academic rules of those schools. In order to meet the academic demands of the school children are given more time and support to study at home. This also explains why Valley students' mean score is better than the scores of the students from other regions. Figure 4.1.10a and 4.1.10b shows mean scores of the students as per their involvement in household work.
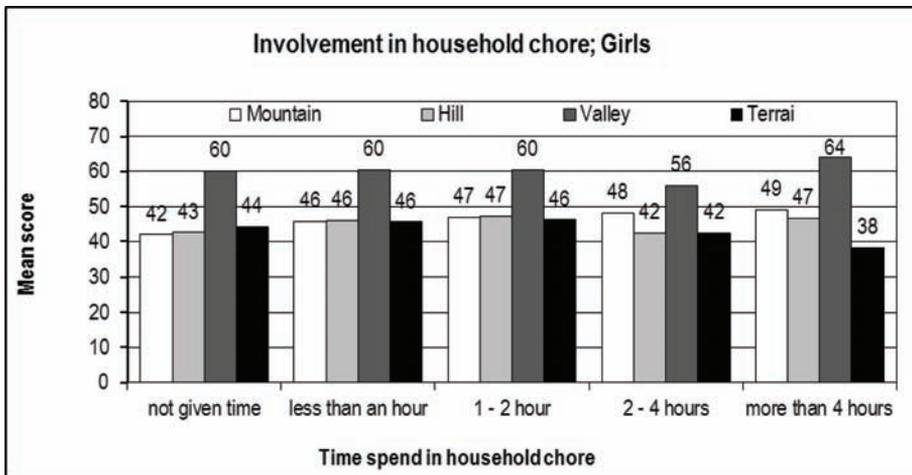
Figure 4.1.10a Ecological zonewise mean score of girls in different groups of time spent in the household work
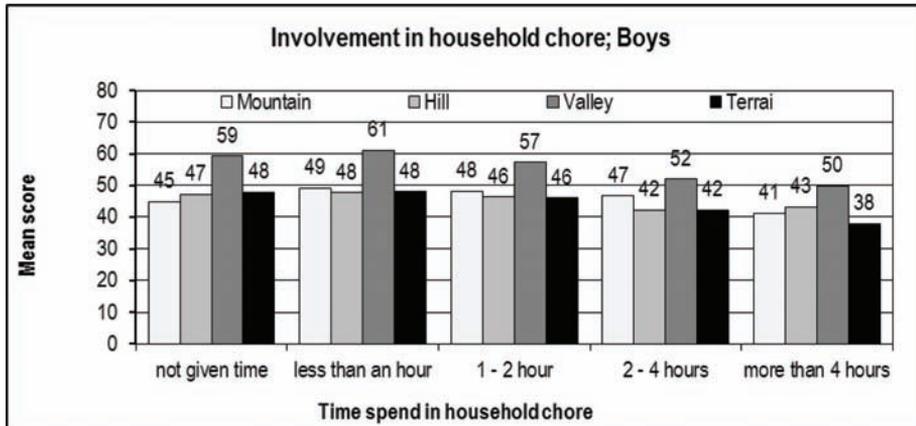
Figure 4.1.10b Eco-zone wise mean score of boys in different groups of time spent in the household work

According to the Figures 4.1.10a and 4.1.10b, there are some differences between the mean scores of boys and girls who are involved in household chores. Generally speaking, the students' achievement goes down when the duration of their household chores increases. This worsens in the case of boys than in the case of girls. In other words, the highest achieving boys are either not doing any household chore or engaged in it for less than an hour to 2 hours. This is true across all the independent variables. Whereas in girls' case the result varies- the highest achieving girls could be engaged in more than 4 hours or less than an hour in household chores. The mean score of girls who are involved in more than 4 hours of household chores is far better than those involved in fewer hours. This is clearly visible when compared across zones. For example, in the Mountain zones, the mean score of the girls who are engaged in household chores for more hours is higher and they are better than girls from the Hill and Tarai areas who also work more hours. In the Hill area, girls' mean score of those who work 'up to 2 hours' and 'more than 4 hours' is the same. Viewed from gender perspective it appears that girls' coping strategies seems to work in the case of those who participated in this study. It has been found that since girls have more household responsibilities than boys they tend to manage their study in between their chores. For example, they are found keeping books on the side while doing kitchen work and study whenever they get time or they stay a little longer after school and finish home work (Parajuli & Acharya, 2008). For boys, most of whom are culturally entitled to either study, play or go for tuition after school (*ibid.*) they are less prepared to manage both household chores and to study at the same time.

251

In Tarai, among those who are involved for more hours in household chores, the gender gap in achievement is in favor of the girls. However, in the Valley among girls the mean score is almost constant whether they get involved in household chores or not.

Among the castes and ethnic groups, the mean score of Brahman/Cheetri and Janjati girls who are involved less than an hour and 1−2 hours in household work is the highest. This is true in the case of Nepali speaking girls as well. Nepali speaking girls who either do not give time or give more than 4 hours in household chores have slightly lower score. Dalit and Madhesi girls who are involved 1−2 hours have the highest score within their respective groups. As in the case of Dalits and Madhesis, Tamang, Magar and Limbu speaking girls who spend 1−2 hours in house activities obtained the highest score. However, among Gurung speakers, girls who spend up to 4 hours and boys who spend less than an hour achieved the highest scores. This also clearly indicates that individual motivation is more prevalent among girls. The reason underneath the motivation could be their expectation that they can achieve a more comfortable life in the future with education. Like the Gurung speaking girls, also the Newari speaking girls are better even if they spend more than four hours on household chores. However, there is a huge gap between minimum and maximum score within the same group in the category of 2-4 hours. Nevertheless, in the case of Newari speaking girls, the prevalence of education as cultural and social capital among most Newars perhaps influences their achievement. This observation is justified by the fact that the literacy rate (62%) and graduation among the literates (4%) among Newar women are higher than almost all caste and ethnic groups (CBS, 2002). Among Tharus and Tamangs both girls' and boys' achievement score goes down as the duration of household chores increases. In general among the language groups mean achievement score of Newari speakers (both girls and boys) are better and Rai the worst. Among Rai speaking girls who spend more than 4 hours and boys who spend less than an hour in household chores obtained the highest scores. The findings above indicate that there is need for exploring the influence of personal/individual motivation as well as learning difficulties in girls' and boys' achievements.

The ANOVA shows a statistical significance in achievement at $p < 0.001$ level between the different groups of household chores. This means students' involvement in household chores to some extent affects their achievement scores. The significant level ($p < 0.001$) of all groups except Alpasankhyak also shows the difference in their achievement score when they involve in household chore. However, the p-value 0.639 (ANOVA; the main effect) of Alpasankhyak girls shows that there is no significant difference in achievement between the different household chore groups. Nevertheless, there is a small main effect noticed at the significant level of $p < 0.03$ in Alpasankhyak boys.

There is no difference of achievement in Newari, Urdu, Rai, Gurung, Sherpa speaker girls and Magar, Tharu, Urdu, Yadav, Gurung, Sherpa and Limbu speaker boys in different household chores groups ($p > 0.05$).

Paid work and student achievement

Students' involvement in paid work before and after school is almost similar to their involvement in household chores, which is higher in other regions and lower in the Valley. In the rural Mountain zone, girls' involvement in paid work is higher (26 percent) than that of boys (24 percent), whereas in urban area boys' involvement is slightly higher (28 percent) than that of girls (27 percent). In rural Hill areas, more boys are involved in paid work. In the rural Tarai region, more boys are involved in paid work whereas in the urban Tarai there is no significant difference between girls' and boys' involvement in paid work. Location wise analysis found that, in general, in rural areas girls' involvement is higher but in urban areas boys' involvement is higher in paid work. Rural areas, where agriculture has been the main occupation people, including children, are engaged in paid labor exchange during farming seasons. Moreover, it is not considered safe to send girls to faraway places and in other sectors to work. Therefore it can be assumed that boys in this study work more in the urban areas and girls in rural areas. Figure 4.1.11 presents achievement scores of girls and boys when paid work before or after school is considered as independent variable.
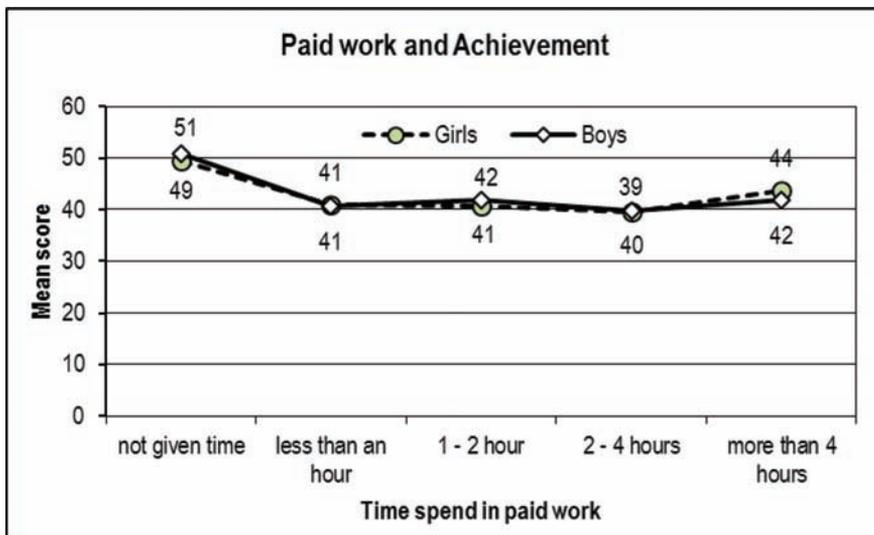
Figure 4.1.11 Students' achievement score: Paid work before or after school

In general, the achievement score of those who do not work on the paid basis is better than those who do work paid, which is close to the scenario found among those who are involved in household chores as opposed to those not involved in household chores. Still, boys' mean score is slightly better than that of girls in this category. The data also shows no significant gender disparity in scores among those involved in paid work for less than an hour to 4 hours. However, when they are involved in paid work their mean score is below that of the average. Unlike the household chore, a definite pattern is observed under this variable. Students' achievement score is the highest among those who do not work for pay irrespective of their place of belonging and social and language backgrounds.

Comparison across the Ecological zones shows that in the Mountain and Hill areas both girls and boys who do not work on the paid basis obtained the same mean score within their group. Similarly, in the Mountain area there is no difference between girls' and boys' score who work for more than 4 hours. As in the case of those involved in household chores, of those involved in 'paid work', Valley students have the highest mean score among the regions. Still, in the Valley and Hill areas, girls' score is lower than that of boys who work equal hours. In Tarai unlike those involved in household chores, boys obtained slightly higher mean score than girls among those who work on the paid basis but among those who work for more than four hours on the paid basis, girls obtained higher.

Among Brahman/Cheetri girls who are involved in paid work for less than an hour and 1−2 hours and for boys who are involved for less than an hour in paid work obtained the highest mean score. Brahman/Cheetri students, although involve in paid work, mostly belong to Nepali mother tongue and 22 percent of whom go to private school give good reason for their high mean score. The mean scores of Brahman/Cheetri, Janjati and Madhesi girls who are involved for 2−4 hours are the same. The same trend is observed among Dalit boys. Janjati girls' mean score is the highest among those who work for more than four hours on the paid basis. Comparison across caste and ethnic groups shows that among boys who work more than four hours for pay Janjati and Dalits are better.

Across language groups mean scores of Nepali, Magar, Tharu, Gurung, Sherpa and Other language speaking girls and Nepali, Magar, Tharu Newari, Urdu, Gurung, Limbu and Other language speaking boys who do not work on the paid basis is the highest in their respective group.

The above deliberations show that paid work affects students' study. Generally, those who are involved no hours of paid work obtained higher scores than those who worked for some hours.

4.1.3.6.3 Bulling and student achievement

According to the data in the sample schools, 43 percent of the students (47% boy and 39% girl) reported that they face some kind of bullying in school. Bullying in school is recorded as being higher in Tarai and the Valley and less in the Hill areas. Location-wise, urban students face bulling more than their rural counterparts. Madhesi students face bullying more than students from other caste and ethnic groups. Bullying seems to occur less among Janjati students.

Research (Rana, 2006) shows that school bullying has an adverse effect on students' psychological and social spheres, which is the most likely to negatively affect their learning. With this understanding an attempt is made to see the connection between the mean scores and incidence of bullying. As said, the achievement score of the students who do not encounter any kind of bullying and those who face fewer incidents of bullying is higher than that of others who experience more bullying. Likewise, the achievement score of the students who face bullying more by their peers is the lowest. Figure 4.1.12 depicts the situation more clearly.
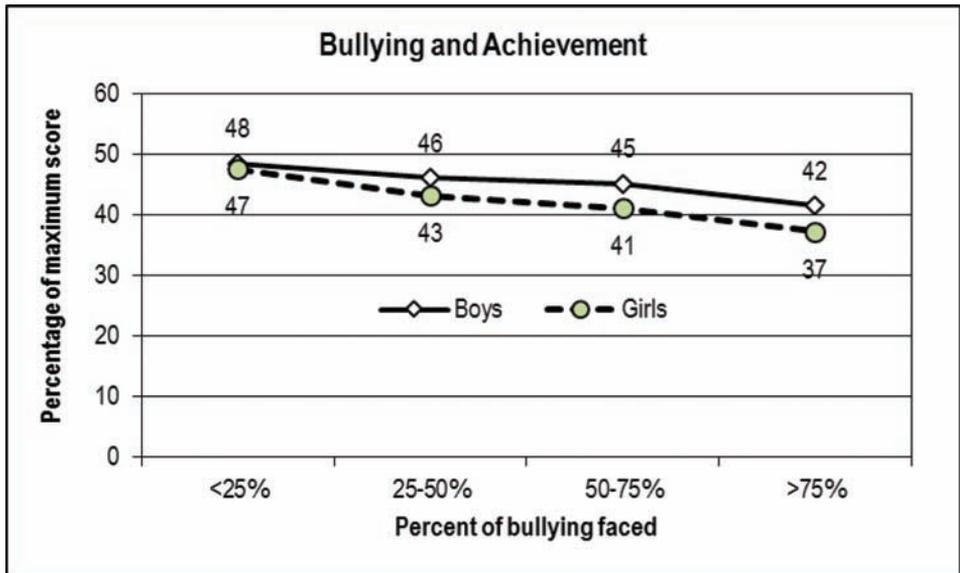
Figure 4.1.12 Bullying in school and achievement

Students who have faced bullying may feel unsafe in school. However, the correlations, though are statistically significant ($p < 0.001$), are mild because of the small number of the students in the highly (50–75% of the incidents faced)

255

extremely (75% or more incidents faced) bullied groups.[21] The effect of bullying is higher within the girls ($r = -0.13$) than boys ($r = -0.09$). The difference in achievement between the extreme groups (less than 25% of bullying and more than 75% of bullying) is 10 percent point in girls and 6 percent points in boys; the difference is statistically significant ($p < 0.001$) and notable and the effect size in girls' group is moderate ($f = 0.16$). It can be safely assumed that bullying and unsafe environment among many other factors can have adverse effect on the student learning and achievement. Thus it requires further investigation.

Even with the same ratio of bullying, girls' achievement score is comparatively lower than that of boys. The similar pattern is found in all Ecological zones, locations, and caste, ethnic and language groups. However, the p-value of 0.190 of Alpasankhyak shows that there is no effect of bullying on student achievement score. Likewise no effect of bullying is found in Newari, Urdu, Rai and Sherpa language groups. Their *p*-value is >0.05 level, i.e. 0.774, 0.130, 0.943 and 0.091 respectively.

4.1.4 Conclusion

Among the sample students, boys scored slightly higher than the girls, however in the Valley where students' score is far better than other Ecological zones, there is no gender disparity in the average score. When cross-verified through other variables, Valley students were less involved in household chores and in paid work. Therefore, it can be assumed that they could afford to spend more time on their studies. On the one hand, in-house as well as outside academic support were more available and, on the other, due to proportionate sampling, the higher number of private schools were selected in the Valley. The combination of all these facts resulted in higher scores among the Valley students. However, micro level analysis across caste and ethnicity; home language; location; involvement in work, etc. showed a wide variation in the student scores. For example, students with Nepali mother tongue were found to be lagging behind students from other mother tongues in Nepali as a subject and the girls spending more than four hours on household chores did better than those working less in some cases. Similarly, in Mathematics Madhesi students (both girls and boys) are ahead among the other groups. This indicated that contextual understanding and measures are required for the purpose of ensuring social and gender equity in education.

---

[21] Students were asked five types of bullying with alternatives 'yes' and 'no': stealing, hitting or hurting, making to do things unwillingly, calling by names, and keeping out of social actions. 100% bullying means that the students had faced all five types of bullying within one month. In the dataset, 687 students faced this kind of severe bullying.

Generally speaking, students achieved higher in the groups where education is a cultural capital. For example, Brahman/Cheetri students' cumulative score was higher than that of others groups and gender disparity in the mean score was also less prominent in this group. This indicated that household's priority, and family member's occupation and education status do have an effect on children's achievement. In this situation, in-school intervention alone may not be the only appropriate measure to raise student achievement level.

Therefore, it can be concluded that, in general, the findings of this assessment support outcome of previous researches. However, when devising interventions to improve learning achievement, the micro level and/or contextual differences and issues need to be considered. For example, neither all girls do require additional support in Mathematics nor can remedial/academic support be the only measure to help them achieve higher scores.

References for Section 4.1

Acharya, S. & Giri, D. (2009). *Measuring learning achievement of Nepali and non Nepali speaking students of Dang district.* Kathmandu: Save the Children Japan.

CBS (2002). *National population and housing census 2001.* Kathmandu: CBS/National Planning Commission.

CBS (2012). *National population and housing census 2011.* Kathmandu: CBS/National Planning Commission.

Koirala, B. N. & Acharya, S. (2005). Girls in science and technology education: A study on access, participation, and performance of girls in Nepal. UNESCO: Kathmandu.

Lamichhane, S. (2003). *Role of classroom management in girls' Mathematics learning.* [Unpublished MEd Thesis]. Tribhuvan University, Central Department of Education, Kirtipur, Kathmandu. (In Nepali).

MLE Research & Review team (2009). Report on MLE policy and strategy based on the Analysis of Mother tongue Speaking Children's Learning Environment in Kanchanpur, Palpa, Rasuwa, Dhankuta, Sunsari and Jhapa. Sanothimi: Inclusive Education Section/Department of Education.

OECD (2008, January). Ten steps to equity. *Policy Brief.* Retrieved from http://www.oecd.org/publications/Policybriefs

Parajuli, M & Acharya, S. (2008). Measuring gender equality in education. A Micro-study of learning environment at home and school: Through the perspective of gender equality. UNESCO: Kathmandu.

Rana, N. (2006). The Experience of Bing Victims of School Bullying: A Phenomenological Study [Unpublished Ph.D Thesis]. Kathmandu: Kathmandu University

Yonjan-Tamang, A. (2009). *Linguistic Heritage of Nepal. Multilingualism for Social Justice, Harmony and Peace.* [Presentation] Kathmandu University, Kathmandu. MPhil/PhD in Education.

4.2 A Good Teacher – A Comparison of Nepalese and Finnish Teachers´ Perceptions of what Constitutes a Good Teacher

Ms. T. Maria Metsämuuronen[1]

&

Dr. Jari Metsämuuronen[1]

1. Helsinki University, Faculty of Behavioral Sciences, Finland

Abstract

This article compares the Finnish and Nepalese teachers' perceptions of what constitutes a good teaching and how teaching can or should be organized for optimum positive impact. The perceptions of the Finnish teachers are done mainly by reviewing Finnish research literature. The perceptions of the Nepalese teachers were acquired by a focus group interview of 14 English-, Nepalese language- and Mathematics teachers in Kathmandu and by using the information on the teachers' background information questionnaires collected from 1,161 teachers in NASA 2011.

In the Finnish literature, the idea of good teaching skills can be condensed into the concepts of 'substance knowledge', 'pedagogical skills' and 'teacher's personality'. The Nepalese material brings up, in addition, the 'classroom managerial skills' as one of the main characteristics of a good teacher. The interviews showed that the teachers in Kathmandu seem to have, in general, quite the similar views of the modern constructive psychological point of view for education as in Finland. However, in Nepal, the modern thinking seemed to be more concerned with theory than practice; the teachers in Kathmandu pointed out that in order to show a good way of (modern) teaching or a sign of a good (modern) teacher, it may not be possible with the resources they have. For example, the huge class sizes (up to 160 students per class in extreme cases) rarely make it possible to use individual learning and teaching strategies in the classroom. On the basis of the questionnaires, the teachers under the age of 26 felt themselves statistically more confident than the older teachers. Also the average learning outcomes in the schools with younger teachers (27 years or less) were statistically better than those of older teachers (< 48 years). Mathematics teachers were more confident than the Nepali language- and Social Study teachers.

Keyword: Teacher, teaching, confidence, Focus group interview

4.2.1 Introduction

Finnish teachers' basic quality seems to be at the top of any world scale – at least when assessed on the basis of the students' results in the international comparisons of Programme for Student Achievement (PISA, 2001; 2003; 2007; 2009) and Trends in Mathematics and Science Studies (TIMSS, 1999; 2011). In Finland, in particular, the differences between schools are among the lowest in the world scale (Schleicher, 2006, 13). It can therefore be argued that in Finland educators have no great deviations when it comes to teaching skills. Instead, in countries where the differences between schools are large, for example, in Nepal, the differences between the qualifications and competencies of teachers may be one reason for the divergent results.

All Finnish teachers need to enroll for a minimum of 4 years of university studies to obtain the qualification for being a classroom teacher and 6"7 years' studies to reach the qualifications for being a subject teacher. In provinces in Nepal, teachers may not have longer than ten months training for the profession. Hence, most probably, there are substantial differences between the teachers of these countries. This article focuses on the perceptions and practices of the teachers in these two countries; in Section 1.1, the Finnish teachers' perceptions are sieved through the relevant Finnish research literature of good teaching and what constitutes a good teacher. In the empirical part in Sections 2 and 3, the Nepalese teachers' perceptions are in focus.

4.2.1.1 Finnish point of view to the good teaching and a good teacher

In Finland, pedagogy and teaching has been studied widely – including both teaching and learning as well as other areas of the discipline. Finland has done well in the PISA survey, and assumingly the long and theoretical-practical teachers' training plays a role in this (see Kansanen, 2003; Niemi, 2011; 2010; Niemi & Jakku-Sihvonen, 2011; 2006; Sahlberg 2011a, 2011b; Schleicher, 2011). Learning to be a teacher begins in the first few grades of school and continues throughout the study period of up to workplace socialization (Lauriala, 2000, 88–90.) Teacher's role is twofold - on the one hand a good teacher must be a reformer but on the other a guardian of traditions. School traditions are strong, and to change them is a difficult task. In addition, teacher's work requires a wide range of content information, substance knowledge, and ethical maturity in a demanding profession. The modern teacher in the 2000s is a reflective, respectful and ethical individual as well as a professional whose aim is to expand and support the people, whose identity is still taking shape. (Heikkinen, 2000, 16–17.)

A contemporary teacher requires broad-based dual qualifications: education needs both theoretical expertise of teaching and learning and in-depth knowledge

about the subject (Patrikainen, 2000, 27–29). According to the current constructivist view of learning, learning takes place not only at school but for life in all those environments where a person works (e.g. Bruner, 1996). In constructivist learning philosophy, learning takes place like a spiral form, and learning is the learner's active thinking activity where the learner builds knowledge upon knowledge. (Bruner, 1996; Heikkinen, 2000, 8; Tynjälä, 2002, 21–22, 37–39; Patrikainen, 2000, 21–23; Hakkarainen, Lonka & Lipponen, 2001). The learner has to learn to apply, integrate and deliver information rather than to adopt it; the teacher is the director of this learning process. The teacher's role is to consider, what the appropriate methods of knowledge construction of learners are (Heikkinen, 2000, 9).

In Finland, the teachers are taken to be high-level experts of the teaching profession; according to a great consensus in Finland, the teachers are thought to be reflective and independent professionals in their own field. This kind of teaching needs a new kind of expertise, which goes beyond the traditional teacher professionalism. Central to this new professionalism are various forms of cooperation. The teaching profession is no longer just about creativity combined with a technical skill; the modern Finnish teacher works through the personality. The neo-professionalism is said to depend on how well teachers are able to structure themselves, their world, and the value of human knowledge and conception of learning (Patrikainen, 2000, 27–29). Hence, according to Partikainen, it is not possible to build a one absolute and objective truth of a model teacher or one proper way to teach. According to him (2000, 27–29; also Heikkinen, 2000, 10–11), the teacher and teacher's personality are shaped by how each teacher observes and constructs the environment and information. Although the teacher would be at the crest of a wave of the neo-professionalism and an expert in his/her subject area, (s)he must also pay attention to the social and emotional side of education (Lauriala, 2000, 88–90).

In a contemporary Western reality, the teacher must be present in environments that are more problematic, restless, and complicated than before. Teaching in a modern world is an interactive process and the starting point for the development of needed skills is the teacher's own value base and the conception of humankind (Patrikainen, 2000, 21–23). The skills of a teacher are developed specifically in those situations which (s)he faces in her/his work; each teacher's decisions are based on either conscious or unconscious values. Also, the teacher must act in such a way what is the best for a student. Lauriala (2000, 88–90) notes that only learnt and developed teachers can guarantee that in today's rapidly changing world students will grow motivated and capable of continuous learning. For these reasons, teachers' qualifications in the neo-professionalism include an ethical approach on the top of reflective theoretical knowledge, practical information, and wide-ranging competence (Lauriala, 2000, 88–90; Syrjäläinen, Jyrhämä & Haverinen, 2009). Heikkinen (2000, 16–17) and Patrikainen (2000, 27–29) note that to be a good teacher in a modern world, the crucial characteristics of the

261

teachers personality includes professionalism, reflectivity, introspectiveness, and appreciation of individuality.

The pedagogical thinking of a teacher can be reflected in the solutions made in everyday life. The pedagogical thinking manifests itself not only in actions but also in what the teachers tell about their work and how they justify their solutions (Jyrhämä, 2002, 18−19; Syrjäläinen, Jyrhämä & Haverinen, 2009; Kansanen, 1995a, 14−16; 1995b, 33). Theoretically, teachers' pedagogical thinking should focus on the teaching-studying-learning process as a whole and it should be highlighted as the aim and interactivity of the learning actions. However, though a teacher should be able to reflect on classroom activities from the theoretical viewpoint, the actual work is still more or less practical in nature through activities. Syrjäläinen, Jyrhämä and Haverinen (2009) remind us that the basis for the teachers' didactic skills lies in the discipline in the contents and the standards set by the teaching profession. Studying the learning- and teaching theories helps students to develop the aims of teaching, learning situations, interaction design, and the evaluation of learning.

To summarize the Finnish literature of good teaching, the emphasis seems to be on substance knowledge, pedagogical skills, and the teacher's personality (see Figure 4.2.1). Each of these areas includes its own detailed contents.
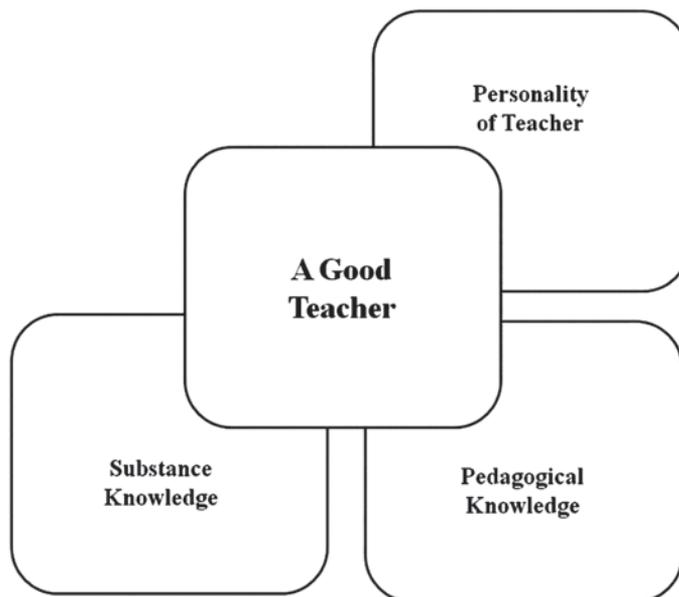
Figure 4.2.1 A good teacher based on the Finnish literature

4.2.1.2 Purpose of the empirical study

The purpose of the empirical part is to investigate the context of the Nepalese teacher and Nepalese teacher's perceptions about what kind of teacher is a good teacher in Nepal on the basis of what is known of the Finnish teachers. Another purpose of the study is to find out how teachers experience their work in Nepal and how education could be organized more optimally. For these purposes, several teachers and specialists were interviewed and 2011 teachers' questionnaires were used to assess teachers' competence in their own subject area.

4.2.2 Methods

The main method for data collection was the focus group interview (see Metsämuuronen, 2013, 283; Hirsjärvi & Hurme, 2004, 41−53). Focus group interviews are a form of research interviews where a number of people are interviewed at the same time. The purpose of the interview is to activate a group discussion about the subject, in this case, the concept of good teaching and what constitutes a good teacher. Two focus groups were set: one for Mathematics teachers and another for Nepalese language teachers. In addition, English language teachers were placed in the groups so that the answers were easier to mediate in English. By doing so, in order to gather the answers as naturally as possible, all the focus group members had the opportunity to respond by using their mother tongue and the English language teachers interpreted the contents to the interviewer (T. Maria Metsämuuronen). Hence, in what follows with the cited references, an English teacher may seem to use a lot of words instead of Mathematics- or Nepali language teachers. The interviews were recorded and transcribed in order to facilitate the analysis of interviews.

Another set of information was collected during the focus group sessions. While the other group was in the interview, the other groups completed two questionnaires: subject-wise background questionnaires (the same as was used in 2011 NASA teachers' questionnaire) and an achievement test, which was the same student test as which was used in Himalaya region in the districts of Mustang and Kalikot (see Section 2). The questionnaire provides wide information about phenomena related to teaching-learning processes and teachers' competencies. The second set of tests was built as a real competence test.

In addition, in order to get a larger picture of the local situation in education, there were interviews on the 16th of December and the 23rd of December with the Joint Secretary Dr. Lava Deo Aswath, and the process expert, teacher and trainer Shyam Acharya and the Head of Unit Hari Aryal, as well as on the 22nd of December with Dr. Jari Metsämuuronen who was working in the Ministry of Education at the time. Later the experts Acharya, Aryal and Metsämuuronen

gave information to clarify the general education system (internet interviews on the 8th of March to the 13th of March 2012).

4.2.2.1 Interviews in Kathmandu

The teachers' interview took place in Kathmandu, on the 23rd of December 2011. Before December, teachers in an item writers' seminar were asked about their interest in participating in Finnish research about good teacher professionalism. Of the list, a total of 14 teachers were contacted, and they all took part in both the interview and the questionnaires. Of the teachers, six were Mathematics teachers and eight language teachers, of which five were Nepali language teachers and three were English language teachers.

The interviews were conducted in Keshar Mahal, Kathmandu, where the Ministry of Education was located at the time of interviews in the NASA unit. The interview proceeded so that initially the interviewer presented the research topic. After this, the interviewees discussed the following topics: what makes a good teacher and what constitutes good teaching.

In the Results section, the interviewees' answers are used as direct quotations (in English). These quotes are marked with a code to maintain anonymity. The first letter indicates which group the person in question has been (M / N / E), the second letter tells the sex of the interviewee (M / F), and the last number indicates the number of that person's serial number (1"14).

4.2.2.2 National survey of teachers

Another material used in this article was the teachers' questionnaires completed during the final testing at the beginning of March 2012. Altogether 1,161 teachers completed the questionnaire. In each questionnaire (Mathematics, Nepali language, and Social Studies) there were a set of questions for the teachers to assess their own competence in selected substance areas related to the curriculum.

Altogether 1,161 teachers from 1,201 schools (97%) answered the questions concerning their self-assessment. The set of questions – borrowed from the TIMSS 2003 teachers' background questionnaire (TIMSS 2006) and modified to fit the Nepalese context – differed between different subjects because of different content areas; 17 areas were covered in Mathematics, 21 areas in Nepali language, and 4 in Social Studies. The basic question was, however, the same in all subjects: "*Considering your training and experience in both Mathematics/Nepali/Social Studies content and instruction, how ready do you feel you are to teach these topics at the eighth grade*?" and the scales were identical: simple (1), somewhat simple (2), not simple (3), not at all simple (4). After reversing the scale, shifting the scale to start from zero, and rationed by the maximum score, the percentage

of maximum score was calculated for each teacher in all subjects. Hence, the self-assessment is made comparable over the different subjects. 100(%) would mean that the teacher felt all the subject areas 'simple' to teach and 0(%) would mean that the teacher felt all the subject areas 'not at all simple' to teach. Technically speaking, the transformation was done in the SPSS software by the following syntaxes:

COMPUTE P_Math_Competence_%=(3-(mean(TQ17_1 to TQ17_4)-1))/3*100

COMPUTE P_Soc_Competence_%=(3-(mean(TQ17_1 to TQ17_17)-1))/3*100

COMPUTE P_Nepali_Competence_%=(3-(mean(TQ18_1 to TQ18_21)-1))/3*100

4.2.2.3 Statistical methods

The study combines the qualitative information from the interviews and quantitative information from the background questionnaires and the achievement tests of teachers. In the statistical part, basic statistical methods, such as Analysis of Variance is used. Additionally, the Decision Tree Analysis is used to find the best cut-offs for the independent variables for explaining the differences in self-experienced competence. The methods are handled in detail in Section 2.

4.2.3 Results

4.2.3.1 General perceptions of good teaching and its challenges in Nepal

The interviewees' answers about teaching and learning were generally in accordance with the modern constructivist view. However, what the teachers said to be a good way of teaching, or a sign of a good teacher, may not have been practically possible with the resources that were available to (average) teachers. The teachers generally think that they did not have the time or resources to develop their own teaching materials. Also the training was said to be too short. Teachers say that it would be good to have teachers' joint meetings, where they could share ideas and teaching materials. One related fact raised up by the teachers is that the final year of the general education, grade 10, is used, in practice, almost entirely to make sure that the students would learn by heart as much content as possible in order to be a success in the School Leave Certificate Examination (SLC). The SLC exam, as the main gateway examination, practically determines the student's further education.[22]

---

[22] Outside of the interviews, it seems evident that, in many cases, the SLC examination has got too important a role in children's lives. One sad example of this is the set of 12 girl suicides in the spring of 2012 reported to be connected to poor success in the SLC examination.

In contemporary Finnish schools both students and teachers are self-reflecting. When asked what the interviewees think about reflection and the reflection of their own teaching, one of the active interviewees responded:

> *Yes, it's very important. Whenever we are teaching, sometimes what we do, we just give whatever we are learnt ourselves. Afterwards, just then [we] see the result. Result would not be good. So we have to think: "I have done this much, it would be like that. I must change [it] in this way." By thinking that we change many more activities according to need time, classroom situation, by interest of our learners. MM1[23]*

This teacher recognizes that it is important for the teachers to reflect on their own teaching in order to assess afterwards whether it had the desired result or not. One needs to think about what has been done and how it should be changed – if needed. However, the interviewees continued, that the number of the students in the private schools (A-grade boarding schools) and governmental public schools are different. In the public schools, the resources for planning and teaching were considered being less than in the private schools, and teachers' general endurance is a test because of the large class sizes, and long days of non-stop teaching. Teachers raised the question whether a teacher can develop their full potential when during the school day there are several hours of lessons without a break.

> *... So that they [teachers in the private schools] can teach in a proper way. But in ours [public school] it's not. Another thing is the allocation of teacher, like the teacher is given six or seven periods of a day. That means she has to go in the class regularly. She does not have breaks in the middle. She cannot be mentally prepared there. Is not it? Because she just teach the first period and there is no second period leisure or rest then what she can think of matter or subject matter, she cannot construct teaching materials. She has to take the whole classes of the day. That is, how a teacher can prepare? So I want to say that here is no proper ratio of teacher and student. EF3*

4.2.3.2 Mathematic teachers' perceptions of a good teacher

When asked what makes a good Mathematics teacher, the teachers agreed that it is important that the teacher is hard-working, that the teacher should possess a sharp mathematical mind and intelligence, and he should be innovative and practical.

> *Yes, that's why I have said math teacher must have mathematical intelligence with him or her. So that he should know the different kinds of techniques, different ideas of techniques and methods. EF3*

---

[23] The interviews are not modified. Hence also all the grammatical mistakes are genuine.

In addition, a good Mathematics teacher should know about teaching methods, as the interviewees agreed that Mathematics is not an easy subject for students.

> *I my job, mathematical teaching is very difficult. All students don't have interest to learn the math. Some students want to learn English language, Nepali language, Social Studies like that and other subjects but, average students don't want to learn math, are not interested to learn math and science also. So I feel difficult to teach. MM1*

In Finland, variety of games and helpful methods for mathematical thinking, such as the Varga-Neményi method (Tikkanen, 2008; Lampinen & Korhonen, 2010; the Varga-Neményi Association, 2012), are considered good for learning Mathematics; these kinds of tools aim to enhance the understanding of the abstract matters by using practical examples, such as everyday life objects and color rods. However, the interviewees did not identify these kinds of new methods, with traditional classroom teaching expressed to be the predominant form of teaching. As noted above, teachers at the community schools do not have time or money to use a variety of methods, such as applied Mathematics games in their teaching, even though it would probably be desired.

> *Yes, we can use Mathematics games and other ideas of Mathematics also. But in Nepal, teachers we have no time. We cannot give such time, more time to collect the materials because in the context of Nepal, teachers have little salary. We cannot be based on that salary. We have to do other works in our house and field/farm. So we cannot provide sufficient time to collect such things. MM1*

It seems that ideas of good teaching do not meet the reality at the general nor subject-wise level. The way teaching should be organized was known by the teachers, but they were not empowered enough to cope with the large amounts of the students and small resources. As a solution to keep the teaching meaningful, the interviewees in the Mathematics focus group proposed teacher meetings to be a good way to share teaching materials and new ideas in a collegial forum.

4.2.3.3 Nepali- and English language teachers' perceptions of a good teacher

Teachers in the focus group of languages pointed out that the qualities of a good language teacher include, among others, linguistic skills, good grammar skills in the Nepali language, clear pronunciation, the skill to tell stories, and a clear voice. Additionally, on teaching it is important to be student-centered and the teacher should be able to control the class and to keep order in the class; the teacher should also keep focus on the curriculum and act in such a way that the students can participate in lessons.

Teaching English was said to be different to teaching Nepali and Mathematics. Mathematics was kept as practical and place-oriented whereas English language teaching could be also fun.

Math is little bit practical subject than English. But in English we can teach in a funny way, in a play way method. But, in math, we have to sit in a particular area, they have to do many more steps, they have to remember it and write it. BUT IN OUR ENGLISH, what we can do, either teach in class, or go in the garden and just teach. And if we are teaching the vocabulary and action words, we can just teach students by showing the acting, by giving them play a game, play way method, role play method, I can use different student centered method. So that they can learn easily. EF1

Language teachers noted the same as the Mathematics teachers that the amount of teachers and students varies remarkably in schools and so the activities at schools and in the classroom differ. One of the interviewees stated that in other schools in one class there can be 65 students and just one teacher, whereas in other schools there might be five students and ten teachers.

Another great impact is that the ratio of the teacher of teacher and students in not matching. Not kept in the proper way. Somewhere there are more students 65 students in a class, single teacher, she has to do everything in 45 minutes. It's very hard to manage there. In some schools, there are 5 students and 10 teachers. So the ratio of the students and teachers is not properly managed. EF3

It is understandable that in these situations there are not the same possibilities for all students to learn. It is also challenging for a teacher to be able to teach in these circumstances where there might be over 50 students in one classroom. It is important that the teachers themselves are aware of their limits and that they are ready to find answers to the challenges they are facing in their profession.

4.2.3.4 A Good teacher is a competent teacher with managerial skills

The interviews revealed the general issues described above as well as specific characteristics of good teaching, such as a good design of lessons, the preparation of a work plan, punctuality, and working according to the curriculum. Both Mathematics- and language teachers saw the pedagogical skills as well as the substance knowledge important when it comes to the characteristics of a good teacher. A good teacher is a competent teacher. One of the interviewees described a good teacher as follows:

A teacher is a good actor. He has to have good character, good personality, dutiful, punctual, politeness, social, helpful, must use teaching techniques. (EF3)

From the educational policy perspective, the quality of the teaching-learning process is essential. If the learner learns, the education can be seen as high-quality. As Joint Secretary Dr. Lava Awasthi pointed out, each school day should be meaningful to the student and the teachers have the key role to ensure that the students will learn (interview on the 23th of December 2011). This is the reason why school reforms (such as SSRP) have a strong focus on teachers' professional development. However, interviewees strongly suggested that if classroom management does not work, it is difficult to teach students and the teacher's own resources are overloaded.

In conclusion, good teaching seems not to work in Nepal as presented on the basis of the Finnish literature in Figure 1. In Nepal, the classroom managerial skills seem to be more important than the personal factors on the side of substance knowledge and pedagogical skills. According to the interviews, although the teacher would have good personality, substance knowledge, and pedagogical skills it is impossible to achieve good learning results, when the class sizes are very large and if the classroom management skills remain incomplete. Hence, a better model for a good teacher in Nepal can be re-formed as a combination of four elements instead of three (Figure 4.2.2).
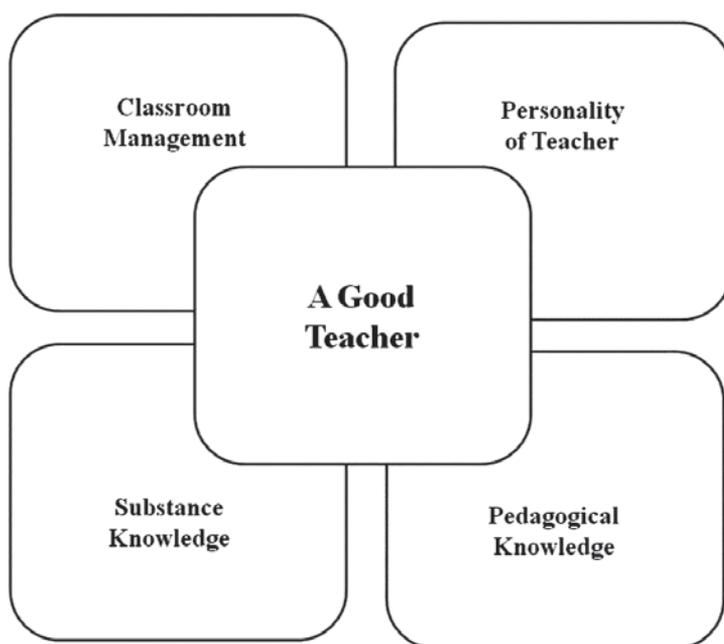
Figure 4.2.2 Qualities of a good teacher on the basis of Nepali teachers' interviews

4.2.3.5 A competent teacher on the basis of the survey material

Because competence seems to be an important characteristic of a good teacher, some characteristics of competent teachers are addressed deeper on the basis of the teachers' background questionnaire collected from the Mathematics-, Nepali language-, and Social Studies teachers.

The Mathematics teachers felt themselves more confident (mean 73% of the maximum score, Std. Dev. 17.8) than the Nepali language- (70, 14.2) and Social Studies teachers (70, 18.9). The difference is statistically significant ($F_{(2,1158)} = 5.42$, $p = 0.005$, $\eta^2 = 0.009$, see Table 4.2.1) but the effect size is small ($f = 0.10$). This means that the difference is real but it is small compared with the variance in the datasets.

Table 4.2.1   ANOVA table of difference between the confidences within three subjects

Dependent Variable: P_confidence_%

Source Type III	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3155.155[a]	2	1577.577	5.425	.005	.009
Intercept	5827957.371	1	5827957.371	20041.751	.000	.945
Subject	3155.155	2	1577.577	5.425	.005	.009
Error	336735.771	1158	290.791			
Total	6168998.575	1161				
Corrected Total	339890.926	1160				

a. R Squared = .009 (Adjusted R Squared = .008)

The Decision Tree Analysis (DTA) – a general tool for data mining – with CHAID algorithm (Kass, 1980) reveals that a more important factor in discriminating the teachers than the subject is the *age* of the teachers. The most confident teachers are those who are 25 years old or younger (75.7% confidence) and the least confident are those who are over 35 years of age (67.5%) (Figure 4.2.3). This may show some kind of wisdom of the older and experienced teachers not to be over confident of their skills as a teacher.
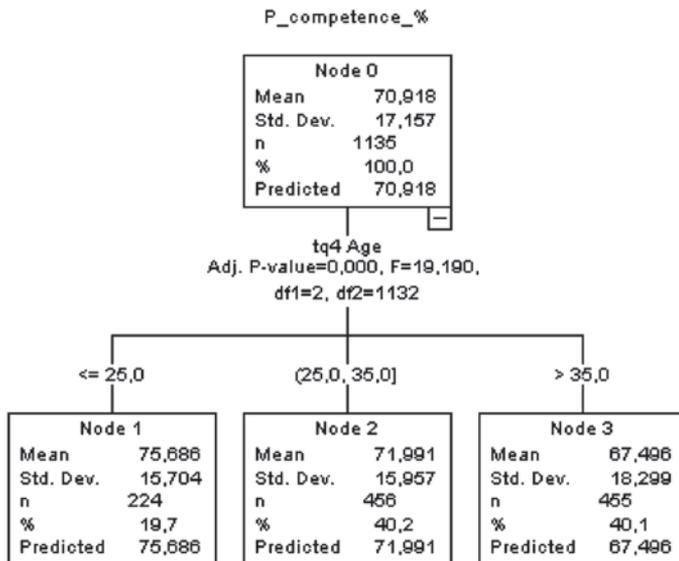
Figure 4.2.3 DTA of Confidence explained by the age of the teacher
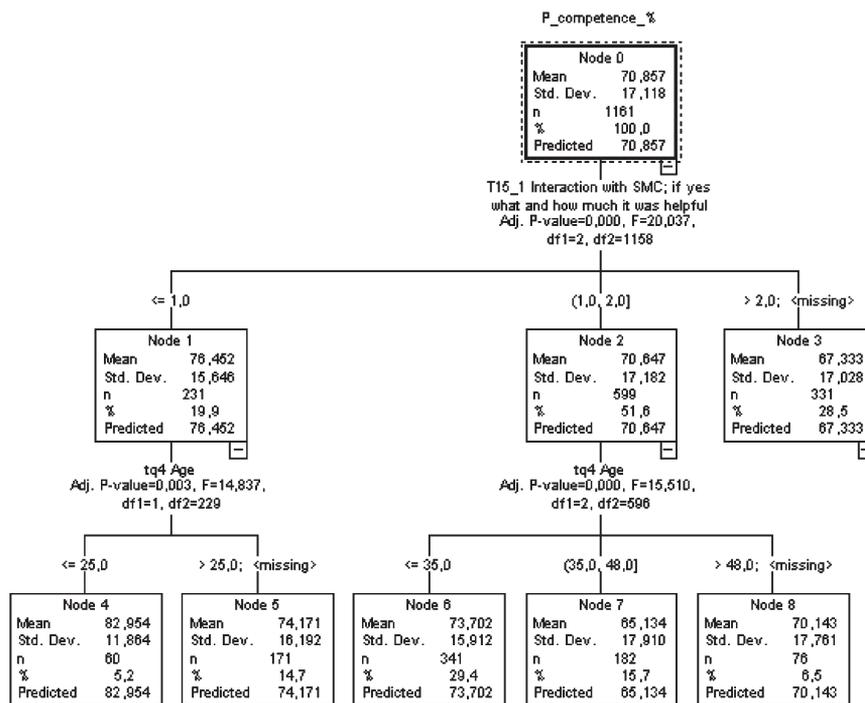
Figure 4.2.4 DTA of Competence explained by more factors

271

The DTA is sensitive to the variables selected in the modeling. During the analysis, different relevant sets of variables were selected in the analysis. When adding variables to the model, an interesting specific group of young teachers with very high confidence are found (Figure 4.2.4). Namely, if there has been *active interaction with School Management Committee (SMC) of improving teaching-learning processes* in the school and the (young) teacher of age lower than 26 years has experienced strong help in his/her work, the self-expressed competence is very high (83% confidence). If taking the confidence in the substance matter as a good thing, it seems that the positive interaction with SMCs seems to especially help young teachers to find their confidence.

It is noteworthy that the teachers' education, school type (community/institutional), school location (rural/urban), Ecological zone (Mountain, Hill, Terrai, Valley), Developmental region (Eastern, Central, Western, Mid-Western, Far-Western), or working experience (in years) does not explain differences in confidence levels. However, it seems that the teachers' age is not only related to confidence, it seems to be related with the better learning achievements in the school; the DTA reveals that, in the higher- performing schools, the teachers' age was, on average, 27 years or lower (with minimum 3 years of experience) and in the lower-performing schools the teachers' age was 48 or higher.

4.2.4 Conclusions and discussion

4.2.4.1 Classroom management as Nepalese characteristics of teacher

In the Finnish literature, the teacher qualification consists of three elements: personality, substance knowledge and pedagogical skills. It is noteworthy that the Finnish literature does not specifically raise the issue of classroom management as one of the basic skills of the teacher. In Nepal, however, it seems to be one of the main characteristics of a good teacher. Though the Finnish literature does not emphasize classroom management, in the international literature it has been addressed quite often. Here some ideas brought by Muijs and Reynolds (2005, 75−84) and Landau (2009, 739−755) are brought up.

Muijs and Reynolds (2005, 75−84) note that because teaching time is limited, it is important to invest in the quality of both teaching and behavior management. The concept of 'classroom management' is connected with the common rules and principles and compliance with them (Landau, 2009, 743). This makes teaching more meaningful and the class is more meaningful to work with. Landau (2009, 747−751) points out that the purpose of the classroom governing practices is to create a meaningful learning environment for students focusing on successful studies and to emphasize fairness toward others and public tranquility. Class management practices are meant to increase students' sense of responsibility.

Muijs and Reynolds (2009, 75) argue that it is classroom management which distinguishes the effective and ineffective teacher from each other.

According to Muijs and Reynolds (2005, 76−83), the effective means of control for the class include the following:

1)   the start of lessons must take place on time and the work instructions should be given before starting to do an assignment,

2)   the seating arrangements must be thought out in advance so that children's temperaments are taken into account,

3)   interruptions (external disturbing factor) during a lesson should be avoided,

4)   transitions between classes will be smooth - thus avoiding the delays,

5)   homework should be administered immediately after making the task, as the students' attention is still in the task, and

6)   ending of lessons should be done in such a way that the students have the opportunity to collect their belongings and leave the classroom.

Donald D. Quinn (Rhem, 2010) has summarized, ironically, teaching and class management as follows:

If a doctor, lawyer, or dentist had 40 people in his office at one time, all of whom had different needs, and some of whom didn't want to be there and were causing trouble, and the doctor, lawyer, or dentist, without assistance, had to treat them all with professional excellence for nine months, then he might have some conception of the classroom teacher's job.

4.2.4.2 Essentials perceptions of Nepalese data

The interviews revealed that teachers in Nepal think in much the same way as teachers in Finland. However, the feasibility of teaching methods is limited by the resources of schools and teachers' large workload. Classroom management skills are needed because of the high quantity of the students and hence, a need to be able to control the masses.

In Nepal there is just one text book that is used, which contains most of the content of the curriculum, and teachers will use much of this while teaching. The interviewees stated on a number of occasions that the use of different materials is meaningful but the resources were seen as an obstacle to their own production of material.

The quantitative analysis of the self-assessment of 1,161 teachers showed that mathematic teachers were statistically significantly more confident than the Nepali language and Social Studies teachers. Most confident teachers are the young ones – below 26 years – especially if they have had good experiences of

School Management Committees visits with helpful comments to improve their teaching-learning processes.

4.2.4.3 Differences between the Nepalese and Finnish teachers – what are they?

The realities in Finland and Nepal are quite different. The last estimate (Central Bureau of Statistics, 2011) of the population in Nepal in 2011 was 28.6 million, 5.5-times higher in comparison with the population in Finland. Nepalese will have an average of three children, and 37% of the population is under 15 years of age – in Finland, the figures are two and 16% (Statistics Finland, 2011). However, the infant mortality rate is still high in Nepal. The life expectancy in Nepal is 67.5 years – in Finland, the life expectancy of males is 76.7 and of females 83.2. In recent years, urbanization in Nepal has been strong, but still most of the population lives in rural areas. Kathmandu Valley, it is estimated, has 1.5 million inhabitants, but the city is home to many unregistered residents, which makes accurate population calculations impossible. (Embassy of Finland, 2012; UNDP 2010.)

In Nepal, around 93% of children start school between the ages of 5 and 9 – in Finland, practically all the children, including deeply retarded children, start school at the age of 7. Approximately 78% of those starting school in Nepal, get to the fifth grade, and only a small portion graduate after 10th grade – in Finland, practically all the school starters also pass the final class (9th grade) of compulsory education. In Nepal, there is still widespread use of child labor, and it is estimated that about 30% 5 to 14 year olds have some kind of work. From the student background questionnaire of the Nepali language of over 16,000 students one knows that 19.5% of the 8th graders work at least one hour per week in a paid job. In Finland, it is practically impossible to find a child of this age working on a paid basis.

There are nine colleges and universities in Nepal, and around 90% of all university students are studying under Tribhuwan University colleges in Kathmandu and different part of the country. The number of study places in tertiary education is obviously very limited in comparison with the number of the potential students. In Finland, around 97% of the students find a study place in some institution – either in vocational schools or at colleges after which 47% of the students receive a tertiary education. Schooling in the community-based schools in Nepal is free of charge as it is in Finland, but the tests are charged for. In Kathmandu Valley, however, the majority - nearly 80% - of the schools are Institutional-based (private) schools. (Embassy of Finland, 2011; UNDP 2010.)

Nepal has tens of thousands of schools delivering compulsory education. On the basis of the latest official records used in the sampling of the NASA 2011 study, there were more than 33 000 schools giving teaching for classes 1 to 6 – in Finland, the corresponding figure in 2011 was 2020 (Statistics Finland, 2012a). In

2007, there were more than 165 000 teachers and more than 6 500 000 students in Nepal (Nepal in Figures, 2008). This means that there were on average 39.4 students per teacher – in Finland the figures were (nearly) 40 000 teachers (Kumpulainen 2011, 38) and 542 000 pupils (Statistics Finland, 2012b) and hence the student/teacher ratio was 13.5 students per teacher. The large number of schools, pupils, and teachers challenge the Ministry of Education in Nepal to evolve and develop teachers' capacity and competence.

Findings from the head teachers' reports of the classroom sizes suggest that the average Nepalese students' classroom size is 5.5 m x 4.1 m and, after subtracting the teacher's space (1 m) and the middle isle (0.5 m), there is practically 4.5 m x 3.6 m H" 16 square meters of floor space for an average 40 students. Hence, in theory and the observations support this, in each square meter of the average classroom there are an average of 2.5 students. In Finland, according to Palonen *et al*. (2009) in a median school for the 1st–6th graders there are 60 m^2 for 25 students which equals somewhat 2.5 m^2 for each student to work and move in. Practically speaking, all the Finnish students have their own desk with half a meter of space all around the work area.

Even though the realities in Finland and Nepal deviate radically, the Finnish and Nepalese teachers find their working ideals quite the same. In the Finnish literature, the attention is given to the teacher personality, substance knowledge and pedagogical skills – the same topics were raised in the interviews in Nepal, too. In Finland, however, the teachers seem to have greater opportunities to carry through these teaching ideals and instruction compared with the Nepalese teachers. In Nepal, the class sizes in compulsory education are easily double (average 40 students) of those in Finnish classes (maximum 24 students). Hence, in the reality of the Nepalese teachers, there is also a strong need for classroom management.

The teachers' salaries in Nepal are low and training (minimum ten months) is short in comparison with the Finnish four- to six-year teacher training period. As teachers pointed out in the interviews, this is not enough for them. The strength of the long education is that within five years of training it is possible to receive a wide training in the theoretical studies but also a deep practical rehearsal in the real schools.

The equipment in teaching can be very rudimentary in Nepal because of lack of resources. However, the discussions in Finnish educational publications also talk about teachers' fatigue and strain, in the same way as the Nepalese teachers feel powerless. From this perspective, the teachers perhaps share some general and global challenges in Nepal and Finland. The responsibility of a teacher in Nepal - as well as in Finland - is great. Teachers are responsible for the entire young people's training and opportunities for a good life. Teachers' professional development projects and training centers are a good opportunity for teachers to get additional resources for education.

A specific feature of the Finnish teacher training is that the teacher profession is highly appreciated (see Kansanen 2003) – it is possible to select the students for the teachers training from the best sequence of the applying students. The high quality of the teachers has been seen as one of the main reasons behind the good results of PISA and TIMSS studies (see Niemi, 2010; 2011; 2012; Niemi & Jakku-Sihvonen, 2011; 2006; Sahlberg 2011a; 2011b; Schleicher, 2011). In Nepal, the situation may be best described in an interview by a teachers' educator and statistician Shyam Acharya (16th December 2011): **"***The teaching profession may not be valued on the same scale as in Finland, but it is one of the few professions that Nepal has to offer*.**"

References for Section 4.2

Acharya, S. (2011). Interview 16.12.2011

Acharya, S. (2012). Email communication 13.3.2012

Aryal, H. (2012). Email communication 13.3.2012

Awasthi, L. (2011). Interview 23.12.2011

Bruner, J.S. (1996). *The Culture of Education.Cambridge*, London: Harvard University Press.

Central Bureau of Statistics (2011). Population. Retrieved from http://www.cbs.gov.np/population_projection_nepal.php.

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

Embassy of Finland, Kathmandu (2012). Information on Nepal. Retrieved from http://www.finland.org.np/Public/default.aspx?nodeid=44214&culture=fi-FI&contentlan=1&displayall=1.

Embassy of Finland, Kathmandu (2011). Unpublished background documents of School Sector Reform Project.

Hakkarainen, K., Lonka, K. & Lipponen, L. (2001). *Tutkiva oppiminen. Älykkään toiminnan rajat ja niiden ylittäminen.* Porvoo: WSOY. [In Finnish]

Heikkinen, H.L.T. (2000). Opettajan ammatin olemusta etsimässä. In **K. Harra** (Ed.), *Opettajan professiosta*. Artikkelisarja. OKKA -säätiö. 8−19. Saarijärvi: Saarijärvi Offset Oy. [In Finnish]

Hirsjärvi, S. & Hurme, H. (2004). *Tutkimushaastattelu: teemahaastattelun teoria ja käytäntö.* Helsinki : Helsinki University Press. [In Finnish]

Jyrhämä, R. (2002). Ei kysyvä tieltä eksy. Pedagogisen ajattelun kehitysvaiheita. In **P. Kansanen & K. Uusikylä** (Eds.), *Luovuutta, motivaatiota, tunteita. Opetuksen tutkimuksen uusia suuntia* (p. 73−95). Jyväskylä: PS-kustannus. [In Finnish]

Kansanen, P. (1995a). Kuorolaulu, lausunta, kansantanhut ja tanssi. Entä pedagoginen ajattelu. In **A. Vaahtokari & A. Vähäpassi** (Eds.) *Tutki, vertaile ja arvioi. Näkökulmia opetuksen suunnitteluun ja tutkimukseen.*(p. 9−23). Helsingin yliopiston Lahden tutkimus- ja koulutuskeskus. [In Finnish]

Kansanen, P. (1995b). Teachers' pedagogical thinking – What is it about? In **C. Stensmo & L. Isberg** (eds.) *Omsorg och engagemang*. Pedagogiska institutionen, Uppsala universitetet.

Kansanen, P. (2003). Teacher education on Finland: current models and new developments. In **M. Moon, L. Vlãsceanu, & C. Barrows** (Eds.), *Institutional Approaches to teacher Education Within Higher Education in Europe: Current Models and New Developments*. Bucharest: Unesco-Cepes, 85–108.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics,***29**(2), 119–127.

Kumpulainen T. (ed.) (2011). *Opettajat Suomessa 2010*. Koulutuksen seurantaraportit 2011:6. Opetushallitus. Retrieved from www.oph.fi/download/131532_Opettajat_Suomessa_2010.pdf. [In Finnish]

Lampinen, A. & Korhonen, H. (2010). Suomessa opitaan matematiikkaa Varga-Neményi-menetelmän mukaan. *Dimensio* **2**/2010, 24–28. [In Finnish]

Landau, B. (2009). Classroom management. In **L.J. Saha & A.G. Dworkin** (Eds.). *International Handbook of Research on Teachers and teaching.* 739 – 753. New York: Springer.

Lauriala, A. 2000. Opettajan ammatillinen uudistuminen: sociokulttuurinen näkökulma opettajan oppimiseen. In **K. Harra** (Ed.), *Opettajan professiosta*. Artikkelisarja. OKKA -säätiö. 8–19. Saarijärvi: Saarijärvi Offset Oy. [In Finnish]

Metsämuuronen, J. (2013). *Handbook of Basics of Researcher Methods in Human Sciences*. 5th Edition. Oy International Methelp Ltd. E-book. [Beta version available at www.Booky.fi]

Muijs, D. & Reynolds, D. (2005). *Effective teaching – evidence and practice*. Gateshead: Athenaeum Press Ltd.

Nepal in Figures (2008). Retrieved from http://www.cbs.gov.np/ Nepal%20in%20figure/Nepal%20in%20Figures%20%20eng%202008.pdf.

Niemi. H. (2012). The societal factors contributing to education and schooling in Finland. In **H. Niemi, A. Toom, & A. Kallioniemi** (Eds.), *The Miracle of Education: The Principles and Practices of Teaching and Learning in Finnish Schools*, Rotterdam: Sense Publishers. 19–38.

Niemi, H. (2011). Educating student teachers to become high quality professionals – A Finnish case. *Center for Educational Policy Studies Journal*, **1**(1), 43–66.

Niemi, H. (2010). Teachers as high level professionals – What does it mean in teacher education? Perspectives from the Finnish teacher education. In **K. G. Karras & C. C. Wolhuter** (Eds.), *International Handbook of Teacher Education: Issues and Challenges* (Vol. l & ll, pp. 237–254). Athens Greece: Atrapos.

Niemi, H., & Jakku-Sihvonen, R. (2011). Teacher education in Finland. In **M. Valenèiè Zuljan & J. Vogrinc** (Eds.), *European Dimensions of Teacher Education: Similarities and Differences* (pp. 33–51). Slovenia: University of Ljubljana & The National School of Leadership in Education.

Niemi, H., & Jakku-Sihvonen, R. (2006). Research-based teacher education in Finland. In **R. Jakku-Sihvonen & H. Niemi** (Eds.), *Research-Based Teacher Education in Finland – Reflections by Finnish Teacher Educators* (pp. 31–51). Turku: Finnish Educational Research Association.

Palonen, J., Kurnitski, J., Haverinen-Shaughnessy, U., & Shaughnessy, R. (2009). Preliminary results from Finnish primary schools' ventilation system performance study. In *Proceedings of Healthy Buildings*, Syracuse, USA.

Patrikainen, R. (2000). Opettajuuden laatu ja opettajan ammatin professiomuutos. In K. Harra (Ed.), *Opettajan professiosta*. Artikkelisarja. OKKA -säätiö. 20–31. Saarijärvi: Saarijärvi Offset Oy. [In Finnish]

PISA (2001). Knowledge and Skills for Life. First results from PISA 2000. Paris: OECD;

PISA (2003). *Education at a Glance*. OECD Indicators 2003. Paris: OECD;

PISA (2007). *PISA 2006 Results*. Retrieved from http://www.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html

PISA (2009). *PISA 2009 Results*. Retrieved from http://stats.oecd.org/PISA2009Profiles/#

Rhem, J. (2010). *Quotations on Teaching, Learning, and Education*. Retrieved from www.ntlf.com/html/lib/quotes.htm.

Sahlberg, P. (2011a). The Professional Educator: Lessons from Finland. *American Educator,* **35**(2), 34–38.

Sahlberg, P. (2011b). Lessons from Finland: Where the Country's Education System Rose to the Top in Just a Couple Decades. *Education Digest*, **77**(3), 18–24.

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.

Schleicher, A. (2011). Is the Sky the Limit to Education Improvement? *Phi Delta Kappan,* **93**(2), 58–63.

Statistics Finland (2011). Internet tables. Retrieved from http://pxweb2.stat.fi/Dialog/varval.asp? ma= 020_vaerak_ tau_101_f i&ti= Kansalaisuus + i%E4n + ja+sukupuolen+mukaan+maakunnittain+1990+%2D+2011&path=../Database/StatFin/vrm/vaerak/&lang=3&multilang=fi. [In Finnish]

Statistics Finland (2012a).[Internet issue]. Retrieved from http://pxweb2.stat.fi/Database/StatFin/kou/kjarj/kjarj_fi.asp

Statistics Finland (2012b) Esi- ja peruskouluopetus [internet issue]. ISSN=1799-3709. Helsinki: Tilastokeskus Retrieved from http://www.stat.fi/til/pop/index.html. [In Finnish]

Syrjäläinen, E., Jyrhämä, R. & Haverinen, L. (2009). *Praktikumikäsikirja*. Helsinki: Studia Pædagogica 33. Original 2004, Reneved 2009. Retrieved from http://www.helsinki.fi/behav/praktikumikasikirja/lukuIII/index.htm. [In Finnish]

Tikkanen, P. (2008). *"Helpompaa ja hauskempaa kuin luulin": matematiikka suomalaisten ja unkarilaisten perusopetuksen neljäsluokkalaisten kokemana.* Jyväskylä studies in education, psychology and social research, 337. Jyväskylä University. Retrieved from http://urn.fi/URN:ISBN:978-951-39-3247-3. [In Finnish]

TIMSS (1999). Mullis, I,V.S., Martin, M.O, Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J. & Smith, T. A. (1999). TIMSS 1999 International Mathematics Report. Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade. Boston College, Lynch School of Education. Retrieved from http://www.timss.org/timss1999i/math_achievement_report.html.

TIMSS (2006). *Questionnaires*. Retrieved from http://nces.ed.gov/timss/questionnaire.asp.

TIMSS (2011). Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Tynjälä, P. (2002). Oppiminen tiedon rakentamisena. Konstruktivistisen oppimiskäsityksen perusteita. Tampere: Tammi. [In Finnish]

UNDP (2010). Nepal Millennium Development Goals. Retrieved from http://www.beta.undp.org/content/dam/aplaws/publication/en/publications/MDG/nepal-mdg-report-2010/Nepal%20MDG%20report%202010.pdf.

Varga-Nemenyi Association (2012). Internet material. Retrieved from http://www.varganemenyi.fi/includes/menetelma.php. [In Finnish]

4.3 Teacher Effect in Learning

Mr. Shyam Prasad Acharya[1],
Dr. Jari Metsämuuronen[2], &
Ms. T. Maria Metsämuuronen[2]

1. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal
2. Helsinki University, Faculty of Behavioral Sciences, Finland

Abstract

This article deepens the knowledge of NASA 2011 by combining all the datasets into one and by explaining the differences in student achievement with *teacher-related* variables. The results are based on 1200 teachers' background questionnaires. The highest performing community schools are compared with the lowest performing schools. The research question is: what has been done differently in the highest performing school to gain the higher results?

The teacher/school effect in Nepal is 68%; this is a very high value. This means that the students within the schools are very much alike and there are wide differences between the schools. The highest student achievement seems to come from the schools with the teachers aged less than 32 years. The younger the age of the teacher also seems to relate to more active behavior in the schools. Mathematics teachers' confidence in their own subject matter was higher in the higher performing schools. In the higher achieving schools, teacher's understanding about the curricular goals is higher, Students' and parents' activity is higher, and the teachers are more actively communicating with the SMCs and other teaching staff to enhance the teaching learning processes. They are also more active in the evaluation of their students. In classroom teaching activities, there are no practical differences between the low- and high-performing schools. However, indirectly, in the active schools and hence, in the higher performing schools, the students are directed to more constructive learning experiences (by giving them more possibilities to work with their own problems and by working in small groups) as well as directed to more complex thinking processes (by letting them explain the answers and decide their own procedures for solving complex problems) and more practical tasks (by giving them possibilities to relate what they have learned to their daily life). In the higher performing schools, the students are evaluated more frequently. The frequent and many-faceted evaluation is also more typical of the active schools. In the higher-performing community schools, there are less unmotivated students and their morale (assessed by the teacher) is ranked higher.

Keywords: Teacher effect, Effective school, Active school, High-performing schools

4.3.1 Introduction

As seen in the rough model of the contextual factors explaining the learning outcomes (See Section 2.3.1), there are various factors that affect the learning achievement of the students. Such general factors such as the home environment, the school environment, the socioeconomic status of the student's family, the student's health and physical conditions may have a fundamental effect on the learning outcomes because they may boost or hinder the educational opportunities of the children. On the other hand, one possibly very influential a set of factors of student performance are the teacher-related variables. This branch of studies is quite wide – alone in the ERIC database one finds 290 publications with the keywords "Teacher effect" or the "effect of Teacher". Teachers' high quality has been seen as of the possible reasons seen behind the high ranking of Finland in PISA studies (see Kansanen, 2003; Niemi, 2011; 2010; Niemi & Jakku-Sihvonen, 2011; 2006; Sahlberg 2011a, 2011b; Schleicher, 2011; and the discussion about other explaining factors in Metsämuuronen, Kuosa & Laukkanen 2013). In Section 4.2, it was shown on the basis of selected Finnish literature and focus group interviews in Nepal that, in the Finnish and Nepalese context, there are at least four angles to good teaching: A good teacher should have adequate *substance knowledge, pedagogical competence, suitable personality* for being a teacher, and *managerial skills*. The last one is a very important factor in the situation where there may be more than 60 students in the same class.

'Teacher effect' as a concept refers to the different learning outcomes produced by the teachers. When the average achievement of the students' taught by different teachers differs radically, it may indicate that the teacher has produced better or poorer results. Hence, it may indicate that one school has more effective teaching when compared with another school. However, from the methodological viewpoint the question of teacher effect is somewhat difficult – in many cases it is difficult to separate the school effect from the teacher effect and in some cases the selection of the students to the school may explain the differences between the schools rather than the teachers' actions on the classroom. International studies – such as PISA and TIMSS studies – have shown remarkable differences between the countries when it comes to the magnitude of the teacher effect. In Finland, for example, where the differences between the schools are very narrow (see, for example, Schleicher 2006, 13), the teacher effect and school effect are together 9–12%, that is, the actions in the school explain only around 10% on the student variation

at the grade 9 (15 years olds).[24] In Nepal, the school/teacher effect may be much higher because of the wide difference between the average achievement levels in the schools (see Section 3.1–3.3). The way the teacher effect is mathematically defined restricts the discussion only to the difference between the teachers *after the children are attending school* – it does not tell of the overall effect of schooling compared with the situation when the children do not attend school at all.

There are few (if any) pieces of research that tell us about the teacher's effect on the learning achievement in Nepal in the same manner of the international studies. However, the Ministry of Education (MOE) in Nepal has carried out various processes to enhance the educational practices and national assessment projects to monitor the learning outcomes. In order to assess the outcomes of the enhancing processes, several National Assessment projects have also been carried out (see BPEP, 1997; EDSC, 1997; BPEP, 1998; PEDP, 1998; EDSC, 1999; CERID, 1999; EDSC, 2001; 2003; CERSOD, 2001; EDSC, 2008; Fullbright, 2008). The results of these projects are not, unfortunately, comparable with each other because of the missing linking procedure between the tests. Though the reports have indicated various conditions which have some effect on achievement and have highlighted various recommendations, as far it is known, there are no straight studies on the teacher effect on the student achievement in Nepal. [25]

---

[24] The school effect (and, in many cases, the teacher effect) in Finland varies somewhat with the school subjects, age group and language groups. For example, at the grade 6 mathematics 2008, the average school effect was 13%, but in the Swedish-speaking minority in the rural area the effect was 34% (Metsämuuronen, 2010, 115). This means that the students within a school were very much like each other and there were some differences between the schools. In the urban areas, the effect was 10% in the Finnish population and 11% in the Swedish population; the low- and high-performing students in the cities were more mixed within the classes. In the later report, Metsämuuronen (2013b) showed that in Mathematics the school's effect in the Finnish population (majority) was round 3.5% in the cities and population density areas whereas in the rural area in the Swedish population (minority) it was 22%. The difference is notable.

[25] Since the New Education System Project 1972, the government of Nepal has started to train teachers to boost student learning. Various short term and long term training programs have been launched by targeting teacher development (SSRP core document, 2009). Teacher Education Project, Primary Education Project (1990), Basic and Primary Education Project (BPEP) (1993), Secondary Education Support Project 2007-10), Secondary Education Development Project, etc. are examples of the investment made by the government in teacher development. But, there has been no significant improvement in student achievement as a result. Teachers started to think that Teacher Training was a burden for them (Action Research, ETC, Bhojpur). As a result, the government of Nepal launched the Teachers' Professional Development (TPD) in 2010, which is now in its third year, and one has to see good signs of improvement in learning achievement as a result. The latest project, the School Sector Reform Plan (SSRP), was developed in 2009 by the MOE and was launched in the fiscal year 2009/2010. NASA 2011 is the first assessment project of the SSRP.

This article shows how strong the teacher effect is on the learning achievement in Nepal when using the same methodology as is used in the international studies. It tries to identify the activities of an effective teacher by comparing the teachers' actions on the highest performing community schools with the lowest performing ones. As an outcome, the final aim is to identify the characteristics of an effective teacher. Research questions are set as follows:

(i)   *How strong is the effect of the teacher on the learning achievement in Nepal?*

(ii)   *What kinds of activities are carried out by the teachers of the highes-performing schools compared with the lowest performing schools?*

(iii) *What are the characteristics of effective teachers?*

4.3.2 Methodological solutions

The analysis is based on the combined data comprising round 1,200 schools and 50,000 students covering the subjects of Mathematics, Nepali, and Social Studies. The sampling, achievement tests, background questionnaires, reliabilities and validities of the tests, as well as the other technical matters are discussed in Section 2; they are not repeated here.

4.3.2.1 Design

The mean achievement level of the schools in Nepal was found to be very interesting from the comparison view point. There are two distinctive populations of schools – lower-performing schools and higher-performing schools (see, for example, Figures 3.1.5 and 3.1.6 in Section 3.1.3). The situation does not change when including all the schools in the analysis (Figure 4.3.1a): most of the institutional schools are performing very well but community schools vary from the very high-performing schools to the very low-performing schools. When simplifying the situation and by focusing on the highest and lowest performing schools, the phenomenon looks as in Figure 4.3.1b.
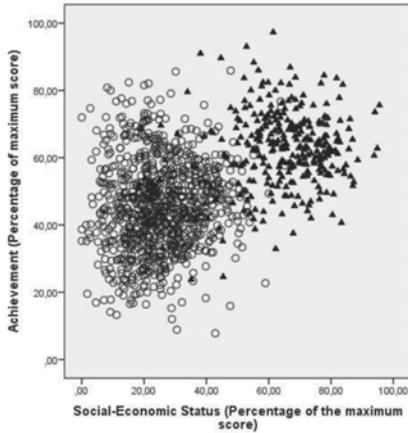
Figure 4.3.1a   Connection of SES and achievement
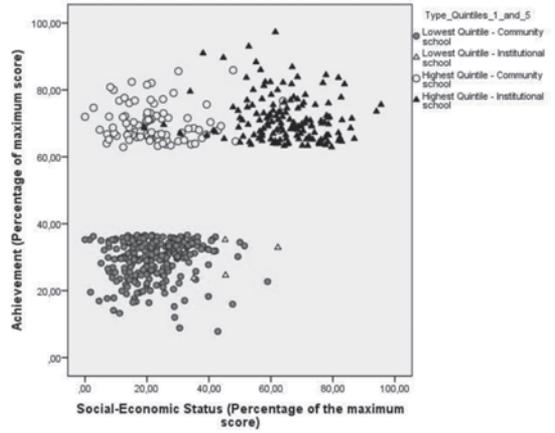
Figure 4.3.1b Connection of SES and achievement in the lowest- and highest quintiles

In the NASA 2011 dataset, there are very few private schools where the learning results are very poor – there are, however, a couple of those – but the institutional schools are mainly in the highest performing quintile. The community schools are divided into two: very high- and low-performing schools. From the methodological point of view, it is possible to form two sets of comparisons as visualized in Figure 4.3.2.
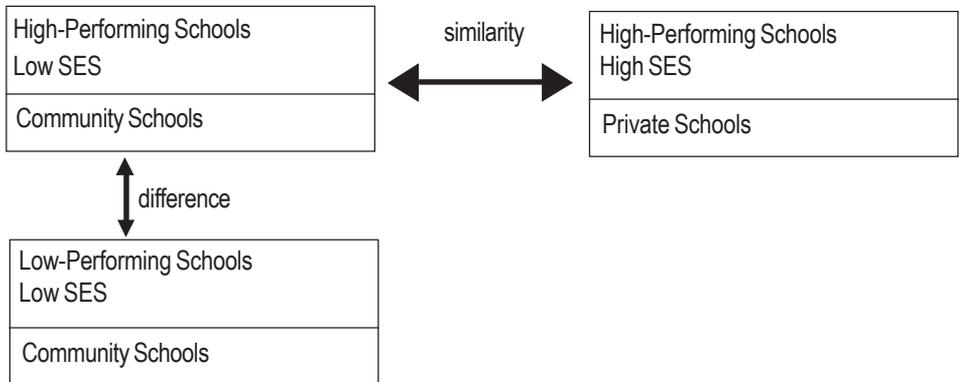
Figure 4.3.2 Research design

The community schools share an equal structure of the socioeconomic status of the students' home, that is, the economic and educational background of the families are the same in the community schools. However, the learning outcomes are

285

radically different. An adequate question, therefore, is what happens differently in the community schools of the highest quintile compared with the community schools in the lowest quintile? Another comparison of note comes from the fact that the results in the highest performing schools with the high educational and economic status (private schools), do not differ from those without economic and educational support (high performing community schools). An adequate question is, what are the common elements shared by the high performing schools? The latter question is not, however, handled here because of space. It seems evident that the community schools in the highest quintile are effective schools: with much lower input they are able to produce the same results as in the private schools which benefits of much more input. This article tries to find the answer to the questions above by concentrating on teachers' actions and characteristics; Section 4.4 focuses on the School factors and Section 4.5 on Student factors**.**

For a more effective statistical analysis the division into quintiles was made again within the community schools; originally out of 250 schools in the highest performing group there were only 102 community schools (the rest were private schools) but in the lowest performing group there were 244 community schools.[26] After the new division, there were the 187 lowest and 187 highest performing community schools to be compared. Some of the schools (6 + 6) did not provide us with the head teacher or teacher information. However, these schools gave the student information and demographical information. The number of schools (n = 374) is sufficient for the most statistical analysis**.**

4.3.2.2 Conceptual framework of the teacher factors

When preparing the teachers' background questionnaire (TBQ), several kinds of factors were identified, which may affect the performance of the teacher. Figure 4.3.3 shows the conceptual framework of the TBQ. Some of these factors are analyzed to detect the differences between the low- and high-performing schools and similarities between the high-performing schools.

---

[1] The situation is somewhat more complicated for the comparison. Namely, from several schools, the students did the test and answered the background questionnaire. More, the students might have been quite good in Mathematics but not in Nepali – or vise versa. In 36 cases, there was no teacher- or head teacher information available. These schools were mainly omitted from the analysis. If the school participated the assessment in 2–3 subjects, the head teachers' information is replicated while the student means are different.
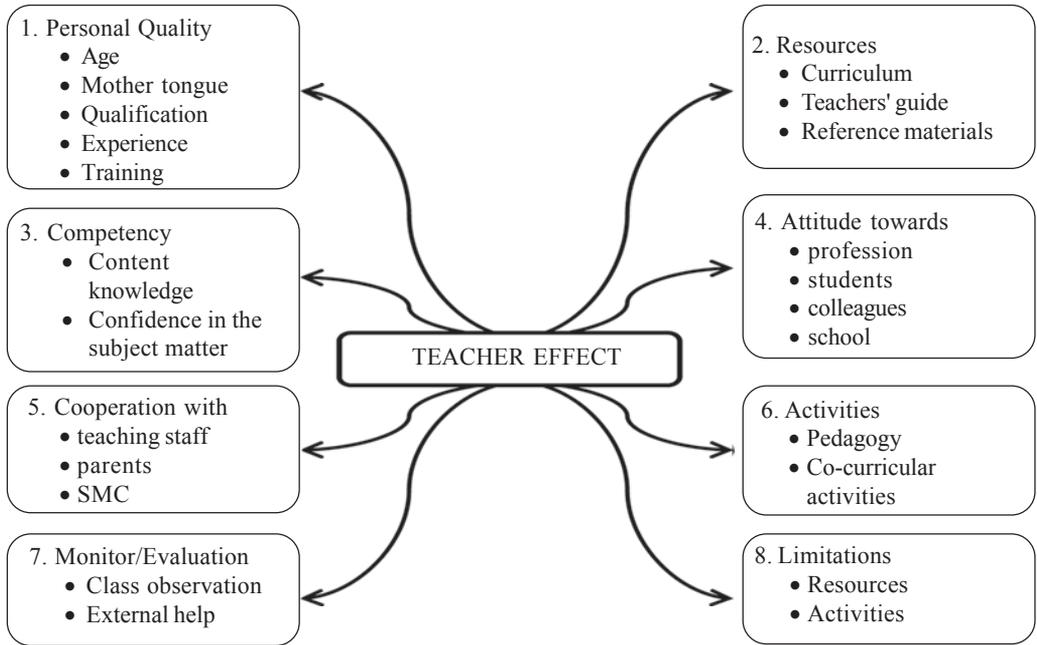
Figure 4.3.3 Conceptual framework of the teacher factors

The framework is not discussed here, it can be criticized and debated, however, a civilized guess is that all the factors in Figure 4.3.3 may be valuable when explaining the differences in students' learning outcomes in the different schools. Note, however, that for all the students in the same class, the teacher is behaving in exactly the same manner but some students just do not learn and some others may learn. In many cases it might be too much to say that these characteristics or actions of the teacher are the real reason for the learning. However, even a small shift in students' achievement levels may be valuable and a secondary factor to found may be a valuable asset in the hands of a wise teacher.

4.3.2.3 Variables used in analysis

In what follows, the teacher effect is estimated in terms of variation of student achievement throughout the sample. Student achievement is estimated as the latent ability *(θ)* by using IRT modeling (see Section 2.5). In the first phase, all the different versions within each subject were equated – that is, their difficulty levels were adjusted to make them comparable – after which all three student datasets were merged. It proved that the Mathematics test was more demanding in the

287

population than the other tests.[27] Hence, without any further transformations, the lowest schools would automatically be those schools with the Mathematics test. That's why the schools' scores were further transformed by using so called 'mean equating', which means that the means of the different subject populations were shifted so that all the distributions had the same mean.[28] Technically, the mean of the means (46.9) was taken as the baseline for the transformation. All the school means were shifted depending on the subject tested in the school; Mathematics schools were shifted up and Nepali and Social Studies were shifted down. In the graphs and statistical outputs one may see a variable label 'P_Eqd_Total_mean_SHIFTED to the mean of 46.9 for all subjects'.

The equated and shifted total score in this full data is used for estimating the school/teacher effect with Multi-level modeling (Goldstein, 1986) or Hierarchical linear modelling (Bryk & Raudenbush, 1987). The large student dataset of approximately 50,000 students was further aggregated to the school level with a total of 1,252 units.[29] Hence, the students' mean in each school forms the estimated "schools proficiency level". It is worth noticing that the individual school mean in the aggregated data no longer carry the information whether the results came from a small school or a large school; information on the school size was stored as an independent variable.

Another variable derived from the previous one is the indicator of the quintiles (Q1 and Q5). Q1 schools were given the code 0 and Q5 schools the code 1. Hence the variable can be used two ways. First, it is a nominal variable which groups the schools into two categories. This way the variable can be used as a fixed factor in ANOVA modeling; as such one asks whether there is a significant difference between the groups with regard to some other variables, such as the attitudes of the teachers. Second, the same variable can be interpreted as an interval variable; so that the mean of the variable tells the proportion of better achieving schools in relation to some grouping variable, such as the language background of the teachers. This way the variable can be used in DTA modeling

---

[27] In the original data of all students, the mean in Mathematics was 42.8, in Nepali 48.6, and in Social Studies 49.3.

[28] In the 'mean equating' one assumes that the distributions are equal but that their locations are different. Now, it is known that the distribution of Mathematics differed from the other subjects (see Section 3.1.3). It was taken as a minor thing compared with the radically lower mean in the population. Other option would have been so called 'linear equating' which is used when there is a difference between the means as well as between the variance.

[29] Note that actually there were 1201 schools but some of the schools were replicated because of the students did several tests and the results varied in many cases remarkably. In one case, for example, the same school was at the highest quintile in Mathematics but in the second lowest in Nepali. Hence, the average of the subjects would have been less accurate than dividing the school into two.

(see the next Section); in this way one can ask whether there is a significant difference between the proportions of the highest performing schools in different classes based on the grouping variable**.**

4.3.2.4 Statistical methods

Four main methodological tools are used in the analysis: Multilevel modeling (MLM), Decision tree analysis (DTA), Analysis of Variance (ANOVA) and both Logistic- and Traditional Linear regression analysis (LRA and RA). MLM is used only in estimating the teacher effect as the proportion of two variance components: teacher-related variance and student-related variance. MLM is run by using the original Theta value and the full student data (around 50,000 students) – the Theta values are comparable over the subjects; zero refers to the average students in each subject. ANOVA, LRA, and RA are basic methods referred in many textbooks (see, for example, Metsämuuronen, 2013a) and hence they are not described here.

DTA is one of the data mining tools in SPSS. It is used when desiring to segment the data into statistically the most significantly deviating groups. Figure 4.3.4 shows an example of a DTA output. In the example, all school data has been analyzed; the equated and shifted total score (*P_Eqd_Total_mean_SHIFTED)* has been used as the *dependent* variable and the teacher's age (tq4_Age) as an *independent* one. The first box, the 'mother node', tells that the mean of all schools in the analysis (n = 1176) is 50.023% of the maximum score with the Standard Deviation 14.979. In the statistical process, the teacher's age has been divided countless ways to find the grouping which produces the statistically most significant difference between the groups in regard to the original dependent variable. Output tells that in regard to maximizing student achievement in the schools, the teacher's age can be divided into three groups by using the following cut-offs: 23 reas or lover (<=23.0), 32 years or lower ((23.0, 32.0), 48 years or lower ((32.0, 48.0]), and over 48 years (> 48.0). The school means are the highest when the teachers are 23 years or younger (55.5% of the maximum scores) and lowest when the teachers are older than 48 years (44.1%). This division of teacher's age to exact these three groups produces the statistically highest test statistics (here $F = 12.81$ with the degrees of freedom of $df_1 = 3$ and $df_2 = 1172$) which produces the lowest p-value (here as adjusted one: Adj. P-value = 0.000 which is notated conventionally as $p < 0.001$). The analysis does not know the reasons for the segmentation; it just gives the cold fact.
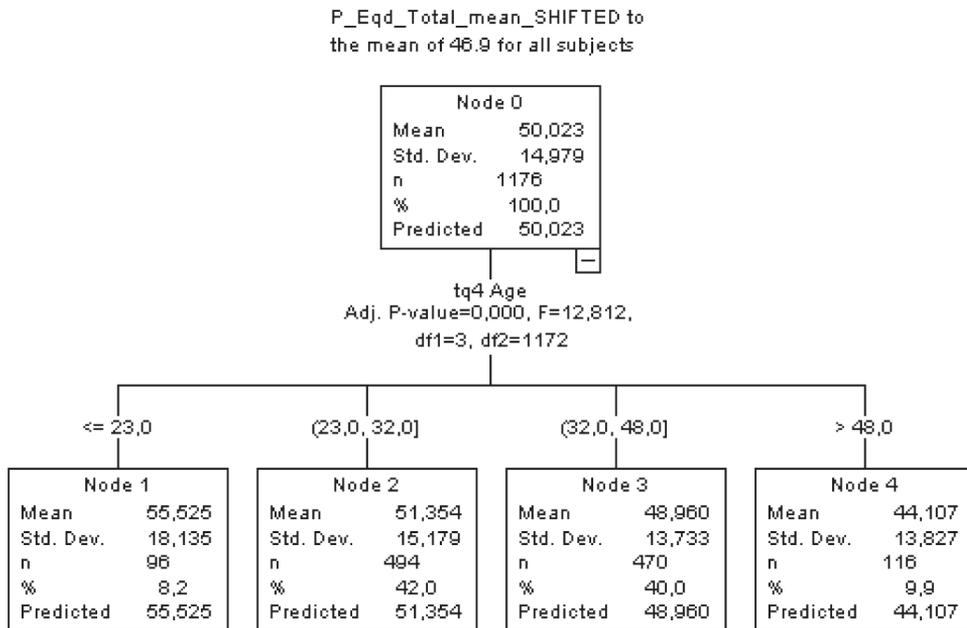
Figure 4.3.4 An example of DTA output

In what follows, DTA is used with the CHAID algorithm (Kass, 1980). When the graphs are wide – as they become easily with several variables, only selected parts of the graph is shown or the results are just explained verbally.

4.3.3 Results

4.3.3.1 Teacher effect

Teacher effect is estimated by using the whole student dataset of 50,000 students and by knowing about which school the students came from. From all the school, there came only one teacher paper and in most cases there actually was only one teacher of Mathematics, Nepali or Social Studies. In larger schools, this is not true and hence, the estimation is somewhat crude when assessing the teacher effect. The figures tell the school effect more accurately than the teachers' effect though they are most probably very close to each

other. The original estimate for the latent ability (Theta) is used as a basis of the calculation.[30]

MLM with Restricted Maximum Likelihood Estimation (RMLE) gives the following results:

Estimates of Covariance Parameters[a]			
Parameter		Estimate	Std. Error
Residual		.386516	.002509
Intercept [subject = New_sch_code]	Variance	.831243	.034640

a. Dependent Variable: Theta_TOTAL.

In the output, the variance related to school/teacher is $\tau_{00} = 0.83$ and the residual variance related to students is $\sigma^2 = 0.39$. On the basis of these figures, the value of the intra-class correlation (Rho) is

$$\rho = 0.831243/(0.831243 + 0.386516) = 0.68$$

In other words, variable *SCHOOL* (or TEACHER) explains 68% of the student variation. This is remarkably much higher than in Finland, for example. The result means that the students within the individual schools are very much alike; value $\rho = 1$ would mean that they were identical within the schools. When knowing that there is a wide difference between the schools and that the schools/teachers effect is as high as 68%, the result reflects a huge inequity between the students and their opportunities to receive equal educational possibilities. Practically all students in low-performing schools get low results and in the high performing schools all get good results. The reality is quite opposite in Finland, for example, where in each class there are very good students as well as very poor students and hence, in all schools the results are practically equal. In Table 4.3.1, the estimated intra-class correlations in the different parts of the country are collected.

---

[30] If using the equated score as a basis for the calculation/estimation, the results would have been 59%. However, the equated score, compared with the original Theta, is cruder; in the equating process, several different values of Theta were transformed to be as equal score. Hence Theta gives a more accurate picture of the students' performance than the equated score. However, the values of Theta are difficult to report as they are standardized values. Thus, the equated scores are used otherwise in the report.

Table 4.3.1 Intra-class correlation in developmental regions and Ecological zones

Developmental region	Ecological zone	Intra-class correlation
	Mountain	0.49
Eastern	Hill	0.69
	Tarai	0.69
	Mountain	0.60
Central	Hill	0.61
	Tarai	0.78
	Mountain	0.09[1]
Western	Hill	0.61
	Tarai	0.66
Mid-Western	Mountain	0.31[1]
	Hill	0.66
	Tarai	0.76
Far-Western	Mountain	0.71
	Hill	0.74
	Tarai	0.64
Valley	Valley	0.60

1) the areas with a low intra-class correlation are highlighted

Three notes of the table may be worth highlighting. First, in Mustang (Western Mountain area), the effect size seems to be at the same level as in Finland – less than 10%. It can be interpreted in a way that regardless of which school the children are in, the results would be equal though the students' achievement may vary within the school. Remarkably lower effect than the average can be found also in Kalikot (30%). Both districts come from the Mountain zones. The reason for the low teacher effect may be that the schools are usually small and hence, the students may be more or less independent actors in the school. Another option for explaining the result is that the test was administered in a slightly different way in those two districts; in Mustang and Kalikot, all three subjects were tested in the same class at the same time. This means that the effect of "social work" in the testing situation was minimized. Second, most unequal areas seem to be Central Tarai (78%), Mid-Western Tarai (76%), Far-Western Hill (74%), and Far-Western Mountain (71%) areas. This may refer either to a strict selection of the students in the schools on the basis of their achievement level or to a high level of "social work".

4.3.3.2 Teachers characteristics related to high- and low-performing schools

In this section, several sets of variables are analyzed to find out what kind of difference can be found between the highest and lowest performing community schools in the highest and in the lowest quintiles. The variables are segmented on the basis of Figure 4.3.2 into eight groups: 1) Personal characteristics, 2) Resources available, 3) Competency, 4) Attitudes towards profession, students or colleagues,

5) Cooperation with staff, parents and school management committees, 6) Pedagogical activities, 7) Monitoring the classroom, and 8) Limitations in teaching, such as in resources.

4.3.3.2.1 General differences between the high- and low-performing schools

Before introducing the discriminative characteristics of teachers in the highest performing- (Q5) and lowest performing schools (Q1), it is good to place the schools in a wider context. There is no statistically significant difference between rural- and urban schools when it comes to the number of Q1 and Q5 schools, that is, from the statistical point of view, the Q1 and Q5 schools are randomly distributed to rural- and urban schools (Table 4.3.2.b).[31] From the equity viewpoint, this is a good signal. However, there is a significant difference between the Developmental regions[32] and Ecological zones[33]. There are more Q1 schools in the Eastern- and Mid-Western region than what is expected in a random case[34] (Table 4.3.2a). On the other hand, in the Valley, Western-, and Central regions, there are more Q5 schools than expected in a random case.[35] The Eastern region and the Valley seem to represent opposite realities; the ratio of Q1- and Q5 schools is 86:14 in the Eastern region and 15:85 in Valley.

Table 4.3.2a Developmental region of the Q1 and Q5 schools

	Developmental region					
	Eastern	Central	Western	Mid-Western	Far- Western	Valley
1 Lowest quintile (Q1)	86,3	37,8	34,1	65.0	46,3	14,8
5 Highest quintile (Q5)	13,7	62,2	65,9	35.0	53,7	85,2
Total	100	100	100	100	100	100
Binomial probability ($p = 0.5$)	< 0.001	0.006	< 0.001	0.002	0.242	< 0.001

The difference between the Ecological zones is much milder than between the Developmental regions; except the Valley, statistically significant difference between the Q1 and Q5 schools can be found only in the Mountain region (Table 4.3.2.b); as there are more high-performing schools in the Mountain region than is expected in a random case.[36]

---

[31] $\chi^2(1) = 0.21$, exact $p = 1.00$
[32] $\chi^2(5) = 71.59$, $p < 0.001$
[33] $\chi^2(3) = 18.58$, $p < 0.001$
[34] Binomial $p < 0.001$ and $p = 0.001$ respectively
[35] Binomial $p < 0.001$, $p = 0.001$, and $p = 0.010$ respectively
[36] Binomial $p = 0.028$

Table 4.3.2b Ecological zone and school location of the Q1 and Q5 schools

	Ecological zone				School location	
	Mountain	Hill	Valley	Tarai	Rural	Urban
1 Lowest quintile (Q1)	40.7	53.3	14.8	57.5	50,2	49,1
5 Highest quintile (Q5)	59.3	46.7	85.2	42.5	49,8	50,9
Total	100	100	100	100	100	100
Binomial probability (p=0.5)	0.028	0.242	<0.001	0.067	0.460	0.460

The Eastern region seems to be interesting from the student achievement viewpoint; the student performance is much lower than in other zones, the lowest performing districts come from the Eastern region, and the number of highest performing schools is radically lower than in the other zones. The characteristics of the Eastern region are deepened in Section 4.1 when addressing the diversity of the country. However, when it comes to teachers' working conditions, it may be worth noting that the average number of schools days in the Eastern region is significantly ($p = 0.001$) lower (138 days) than the average in the other districts (152 days).

4.3.3.2.2 Teacher's personal characteristics

Teachers' personal characteristics, such as Age, Mother tongue, Qualifications, Working experience, or Activity in continuous training, seem not to differ radically between the Q1- and Q5 schools. However, DTA shows that, in the whole schools data (1200 schools), the results are better when the teacher is 32 years old or less, than when the teacher is 48 years or older (see Figure 4.3.4 in Section 4.3.2.4 above). Two other differences between the Q1- and Q5 schools are related to professional training. The first is that the teachers in the Q5 schools have taken significantly more training from the area of *student assessment* ($p = 0.030$) and in *content matters* ($p = 0.035$) (Figure 4.3.5); somewhat double the amount of teachers participated in the education of assessment in Q5 schools (24%) in comparison with the Q1 schools (13%). The other, may be more crucial, difference is that the teachers in the Q5 schools participated in *several kinds* of professional development processes more Q1 school teachers; double the amount of teachers in Q5 schools (33%) attended two or more out of six given possibilities of professional development possibilities compared with the Q1 teachers (16%). Further, more than three times more teachers in Q5 schools (21%) attended three or more types of education courses compared with the Q1 teachers (7%). This

may refer to the better possibilities or stronger motivation of the teachers to take advantage of continuous education opportunities in Q5 schools. It is worth noting that there is no difference between the schools when it comes to teachers' qualifications, their working experience, or other professional training.

Professional development in assessment

Professional development in content matters

Q1_5_NEW2_COMMUNITY_
Based_on_more_schools USE
THIS

```
              Node 0
Mean           0,502
Std. Dev.      0,501
n              231
%              100,0
Predicted      0,502
```

V18e Professional development
last 2 years: Assessment
Adj. P-value=0,030, F=4,748,
df1=1, df2=229

<= 0,0   > 0,0

```
      Node 1                    Node 2
Mean        0,468      Mean        0,651
Std. Dev.   0,500      Std. Dev.   0,482
n           188        n           43
%           81,4       %           18,6
Predicted   0,468      Predicted   0,651
```

Q1_5_NEW2_COMMUNITY_
Based_on_more_schools USE
THIS

```
              Node 0
Mean           0,502
Std. Dev.      0,501
n              231
%              100,0
Predicted      0,502
```

V18a Professional development
last 2 years: Content matters
Adj. P-value=0,035, F=4,484,
df1=1, df2=229

<= 0,0   > 0,0

```
      Node 1                    Node 2
Mean        0,447      Mean        0,589
Std. Dev.   0,499      Std. Dev.   0,495
n           141        n           90
%           61,0       %           39,0
Predicted   0,447      Predicted   0,589
```

Figure 4.3.5 DTA of professional development

4.3.3.2.3 Resources for the work

The resources available for the high quality teaching – Curriculum, Teachers' guide, and Reference materials – does not explain the differences between the lowest and highest performing schools. Actually, the situation is better in the lowest performing schools compared with the highest performing schools (see Table 4.3.3); more often there is material available and it is used.

Table 4.3.3 Resources in use in the Q1 and Q5 schools

Variable	Q1 (%)[1]	Q5 (%)[1]	p[2]
pq12_a Nepali Curriculum, books & Teacher Guidelines (all)	51.1	44.3	0.178
pq12_b Social Studies Curriculum, books & Teacher Guidelines (all)	49.7	42.9	0.135
pq12_c Mathematics Curriculum, books & Teacher Guidelines (all)	47.7	45.3	0.347
pq12_d Other Subject Curriculum, books & Teacher Guidelines (all)	41.0	37.0	0.186
T10 Do NOT have the curriculum of lower secondary level	22.2	22.9	0.544
T10_1 If curriculum yes, have NOT USED the curriculum	13.5	17.0	0.280
T11 Have NOT got the teachers' guide of the subject you teach	35.2	44.3	0.119
T11_1 If yes teachers guide, have NOT USED the guide	22.3	31.4	0.151

1) The valid percent of those who answered the question
2) Significance when testing whether the frequencies are equal

Though there is no difference between the groups (see Table 4.3.3) it is still somewhat interesting that 35–44% of the teachers have not the resource materials. Another interesting fact relates with teacher's understanding of the objectives of the curriculum. Teachers are supposed to understand and teach according to the curricular goals and objectives. Data shows, however, that only 43% of the teachers in Q1 schools and 58% of the teachers in the Q5 schools had high or very high understanding of the curricular goals of grade 8. Also DTA points out that, in the lower-achieving schools, the medium-, low-, and very low understanding was more frequent than in the high-performing schools.

Table 4.3.4 Teachers' understanding of school's curricular goals

	Q1 (%)[1]	Q5 (%)[1]	p[2]
1 very high	9.4	7.3	0.292
2 high	33.1	53.4	0.003[3]
3 medium	50.3	35.4	0.015
4 low	6.1	3.9	0.240
5 very low	1.1	0	0.250

1) The valid percent of those who answered the question
2) Significance when testing whether the frequencies are equal
3) Statistically significant differences are highlighted

4.3.3.2.4 Confidence in the content matters

Teachers' competence was not tested or asked as such. However, their confidence in their own subject matter was asked in several questions in the original teachers' questionnaires. When combining the data, this rather nuanced information cannot be used. Nevertheless, in the merging phase the mean of all variables related to confidence was calculated, it was transformed into the percentages of the maximum score, and added into the common file. Hence, the teachers' competence is comparable though in the original questionnaires the questions were different and the lengths of the entity differed between the subjects. It is worth emphasizing that confidence does not tell strictly how competent the teachers are. Nevertheless, as best, it gives a rough idea of it as a self-experienced report. To some extent, it may though reflect the actual competence. As a whole, the overall confidence does not explain the division into the Q1- and Q5 schools *(p = 0.817)*. The means are actually *identical* as seen in Table 4.3.5.

Table 4.3.5 Overall confidence in substance matters in Q1 and Q5 schools

Group Statistics

	N	Confidence (Mean)	Std. Deviation	Std. Error Mean
Lowest quintile school (Q1)	174	1.9	0.53	0.041
Highest quintile school (Q%)	170	1.9	0.56	0.043

However, when the analysis is done as a two-way ANOVA with the different subject groups, there are significant differences between the subjects[37] (Figure 4.3.6); the Mathematics teachers in the highest performance schools are significantly more confident in their subject than the teachers in the lowest performance schools. The same holds true with the whole dataset: for reason or another, the Mathematics teachers feel themselves more confident than the other teachers.[38] It is notable that this is not seen in the lowest performing schools (see Figure 4.3.6).
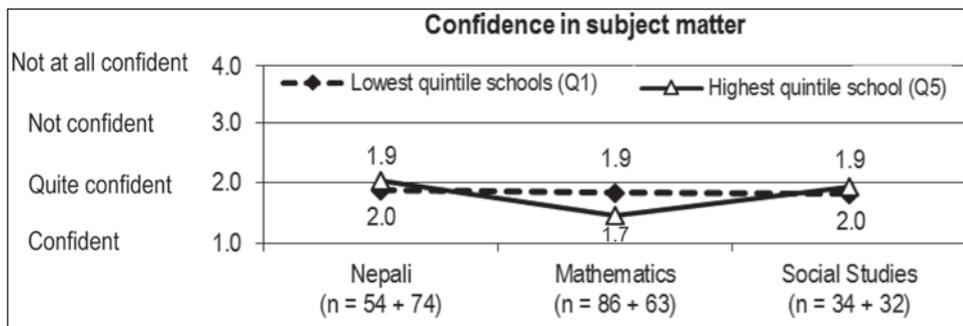
Figure 4.3.6 Difference in the confidence in substance matters between Q1 and Q5 schools

---

[37] With Q1- and Q5 schools: $F(2, 340) = 2.586$, $p = 0.077$ (Two-way ANOVA)
[38] With all schools: $F(2, 1138) = 9.249$, $p < 0.001$ (One-way ANOVA)

4.3.3.2.5 Teachers impression of school's safeness, teachers capacities, parents, and students

Teachers were asked two sets of questions concerning their perceptions of their school's safeness, their fellow teachers' capacities, parents' involvement in the studies, and students' regard for the school property and their will to do well in school. The questions are seen in Figure 4.3.7 below. The first set of questions (T19) was anchored to the values 1 (very much agree) and 4 (not at all agree) and the other set (T20) to the values 1 (very high) and 5 (very low) – note the reversed scales.
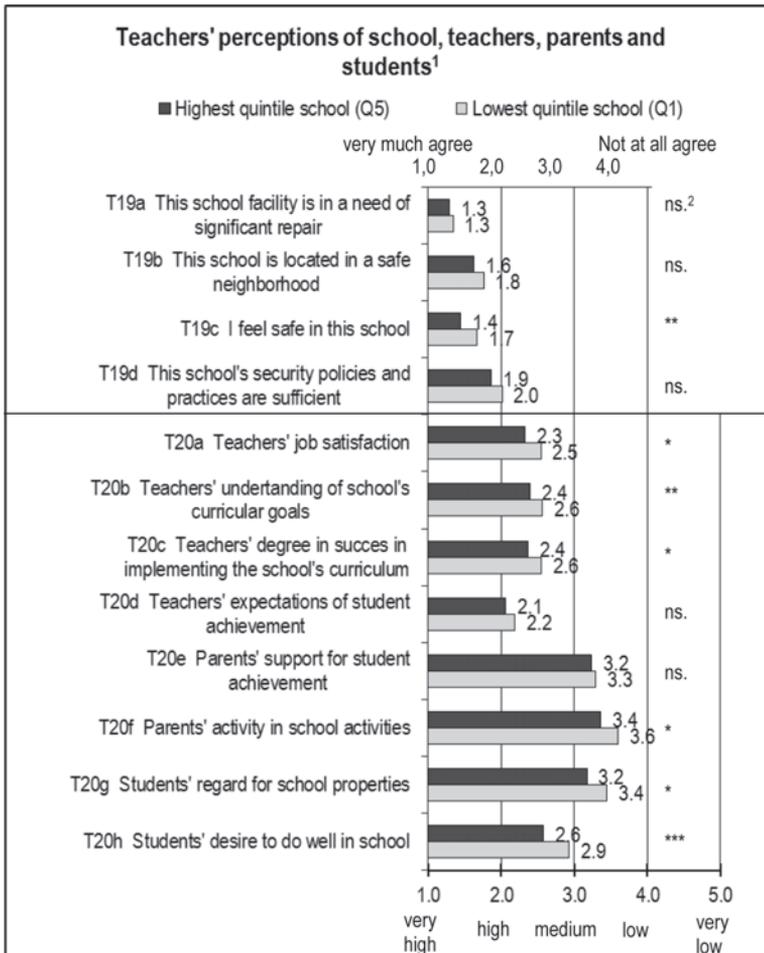
1) Note the difference in scales: the upper variables vary from 1 to 4 and the lower variables vary from 1 to 5.
2) Statistical significance: ns. = non-significant, * = significant at 0.05 risk, ** = significant at 0.01 risk, *** = significant at 0.001 risk

Figure 4.3.7 Differences in teachers perceptions between Q1 and Q5 schools

A number of indicators show statistically significant difference between the Q1- and Q5 schools' teachers. Most significant differences are those related to students' behavior: in the lowest performing schools, the students – from the teachers' perspectives – are not that interested in doing well in school[39] and the students seem to be slightly more concerned with school properties.[40] In the high-performing schools, the teachers feel safer than in the lower level schools.[41] The differences are, though, not remarkable as seen in Figure 4.3.7.

The analysis is intensified by forming three summed variables indicating the teachers' capacity in quality teaching (t20a – t20d), the parents' involvement in education (t20e – t20f), and the students' positive behavior in active learning (t20h – t20i). The summed variables were reversed so that the most positive value is the highest values which are then transformed into the percentages of the maximum score. Hence, the value 100 indicates that in all variables the teacher selected the most positive alternative and 0 means the opposite. The differences are illustrated in Figure 4.3.8.
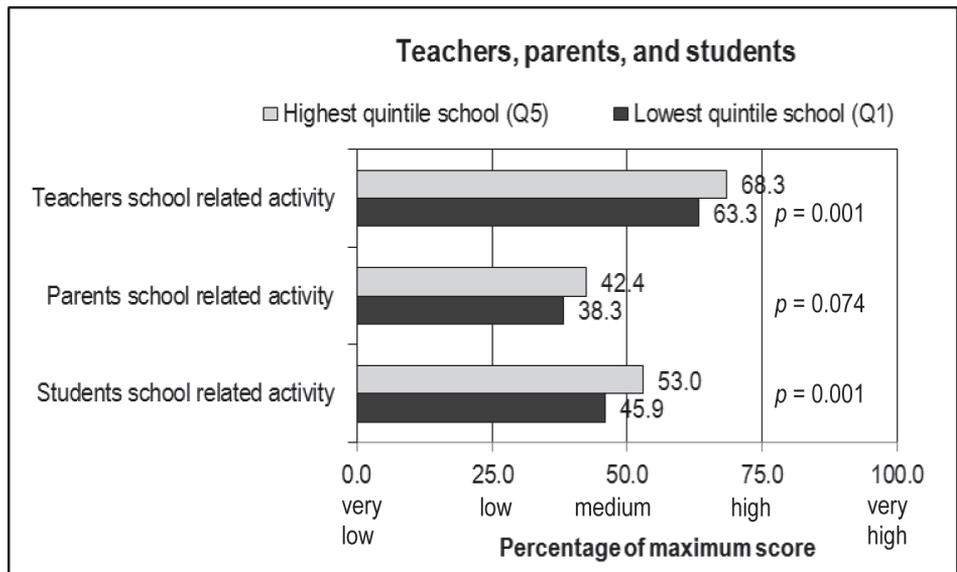
Figure 4.3.8 Differences in teacher-, parent-, and student-related activities for learning between Q1- and Q5 schools

---

[39] t(358) = 3.344, p = 0.001
[40] t(359) = 2.485, p = 0.013
[41] t(357) = 2.778, p = 0.006

299

There is no statistically convincing difference in parents' activity to learning outcomes between Q1- and Q5 schools ($p < 0.074$). After summing up the variables, there seems to be a wider difference between the low- and high-performing schools in teachers'- and students' activity ($p = 0.001$ for both); in the lowest quintile, the teachers' seem to have challenges in job satisfaction, understanding the curriculum, and the degree to which they can implement the curriculum in reality. Maybe the challenge is – as the teachers report: in the Q1 schools, the students seem not to take their studies as serious as in the Q5 schools.

4.3.3.2.6 Cooperation with the teachers staff and SMC

A set of two questions was asked to the teachers, first, whether or not there was interaction between the School Management Committee (SMC) and, second, how many times there was interaction with the teacher staff within a year regarding the improvement of the teaching and learning. DTA detects the interaction between the teaching stuff to be more significant variable ($p < 0.001$) than meeting the SMC ($p = 0.007$) though both appear to be statistically significant predictors for separating the Q1- and Q5 schools. The differences between Q1- and Q5 schools are visualized in Figures 4.3.9 and 4.3.10. The differences are radical when it comes to those teachers who reported more than 7 interactions per year; almost double the amount of these teachers came from the highest performing schools (55%) compared with the lowest performing schools (31%). The same happens with the SMC; from those teachers who reported that they have not met the school management committees, two times more came from the Q1 schools (22%) than from the Q5 schools (11%). This may tell that, in the high-performing schools in comparison with the low-performing schools, both teachers and SMCs are more active in developing the school.
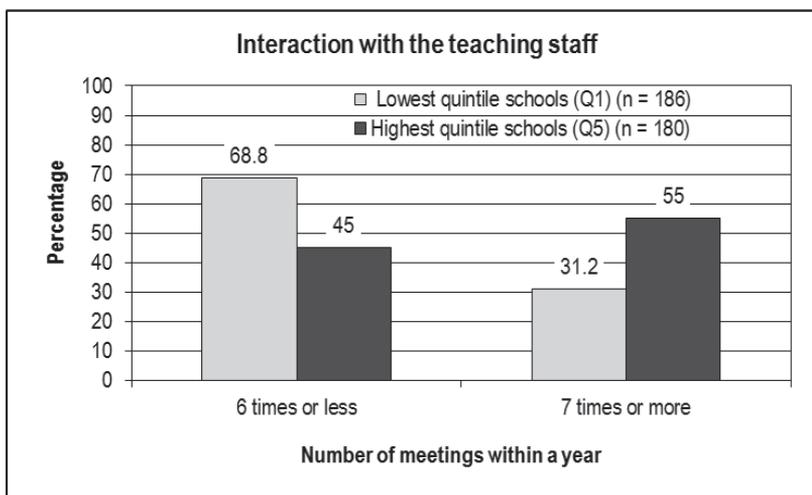
Figure 4.3.9 Differences in the number of teaching staff meetings in Q1- and Q5 schools
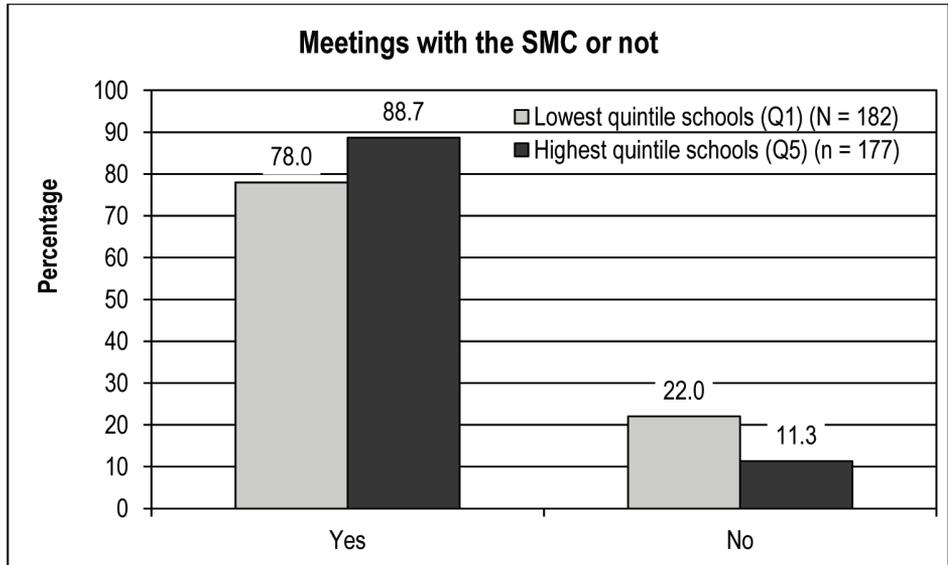
Figure 4.3.10 Differences in the number of teaching staff meetings in Q1- and Q5 schools

4.3.3.2.7 Teacher's activities, pedagogy, and student evaluation

Two sets of questions scanned the pedagogical decisions of the teachers. The one (T21) focused on percentages of *used time in a typical week lesson*. It seems that the teachers in both the low- and high-performing schools act in somewhat the same way: almost half of the time (46% of the time in both groups) is spent with more or less teacher-led methods: lecturing, guiding, and re-teaching the material (Figure 4.3.11). There is no statistical difference between the schools in respect of any of these variables.

Another set of questions (T22) asked *how often* the students did different general things during the lessons. The items are collected and profiled in Figure 4.3.12. Except the item "relate what they have learned in their daily life" ($p = 0.078$) there are no statistical differences in the profiles. On the basis of "relating the daily life experiences" in the classroom, it though seems that in the highest performing schools the teaching is slightly more grounded to the practical experiences than in the lowest performing school. However, the differences are not remarkable.
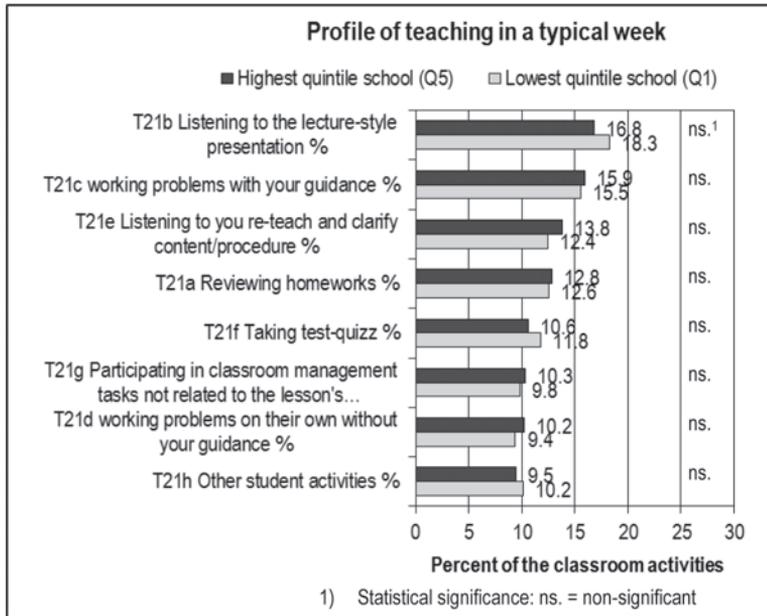
Figure 4.3.11 A typical teaching week in Q1- and Q5 schools
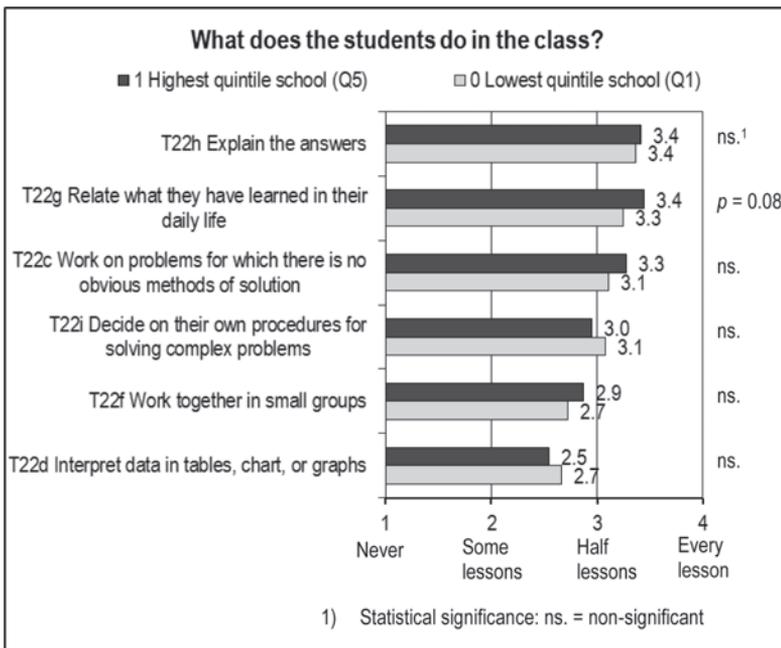
Figure 4.3.12 Students' activities in Q1- and Q5 schools

4.3.3.2.8 Evaluation and monitoring

The practices of student evaluation were asked in two sets of questions: "do you use the following ways of student evaluation" and "how often do you use…" with alternatives of classroom evaluation, homework evaluation, discussion, project work, observation of behaviour, and unit examination. It is notable, that there is no difference between the Q1- and Q5 schools when it comes to "do you use" questions – practically all the teachers (99–100%) use these methods. The difference comes when asking the "how often" questions; the teachers in the lowest performing schools use significantly less classroom evaluation ($p = 0.001$) and homework evaluation ($p = 0.002$) per week. It is, though, notable that this is not seen in their percentages of the normal week (see Chapter 4.3.3.2.7); both groups used 12% of their teaching time for checking homework.

Table 4.3.6 Difference in the amount of valuation between the Q1- and Q5 schools

How often you use the following:	Lowest quintile school (Q1)	Highest quintile school (Q5)	Sig.
T13_1B classwork evaluation (per week)	3.0	3.7	0.001[2]
T13_2B homework evaluation (per week)	3.6	4.2	0.002
T13_3B discussion (per week)	2.5	2.8	0.176
T13_4B project work (per month)[1]	1.6	1.7	0.925
T13_5B observation of behavior (per month)	3.8	3.7	0.898

1) There are very many missing values in the variable.
2) Statistically significant differences are highlighted.

One interesting difference between the teachers in the lowest and highest performing schools, maybe related to the student evaluation, is that the teachers in the highest performance schools seem to be much more aware of their students' real achievement level. From Table 4.3.7, one infers that there is not difference between the teachers in the lowest level schools in comparison with the teachers from the highest level schools (all $p > 0.170$). When knowing, however, that the average achievement of the students in the lowest level schools was extremely low but 66% of the teachers in those schools have high or very high an expectation of their students, one understands that their expectations are without any kind of realism.

Table 4.3.7 Teachers' expectations of their students

	Lowest quintile school (Q1)[1]	Highest quintile school (Q5)[1]	p[2]
1 very high expectations	20.4	25.1	0.220
2 high expectations	45.3	48.0	0.409
3 medium expectations	29.3	24.0	0.179
4 low expectations	5.0	2.2	0.133
5 very low expectations	0.0	0.6	0.500

1) The valid percent of those who answered the question
2) Significance when testing whether the frequencies are equal

The discrepancy between the teachers' expectations becomes clear when dividing the students into five groups on the basis of their actual achievement: 1) below 20% of the maximum marks (very low achievement), 2) 20–35%, 3) 35–50%, 4) 50–65%, and 5) over 65% of the maximum marks (high achievement). These groups are given ranks 1–5 the same way as the quintiles are done – now, however, on the basis of the absolute level of the students, not on the basis of frequencies of the students. When knowing from which school the students came and hence, what the teachers are expecting of their students in those schools, a discrepancy indicator is calculated by substracting the teachers' expectation and the rank of the student. The values of the indicator ranges from -4 (far too *low* expectation, that is, the teachers think that there is not much to expect of their students but the students belong to the most advanced group) to +4 (far too *high* expectation, that is, the teachers expect the highest from their students but the students can hardly read or write). Values -1 to +1 indicate quite realistic expectations. This indicator tells crudely how far the teachers' expectations are from the reality. Figure 4.3.13 reveals the great difference between the Q1- and Q5 schools.
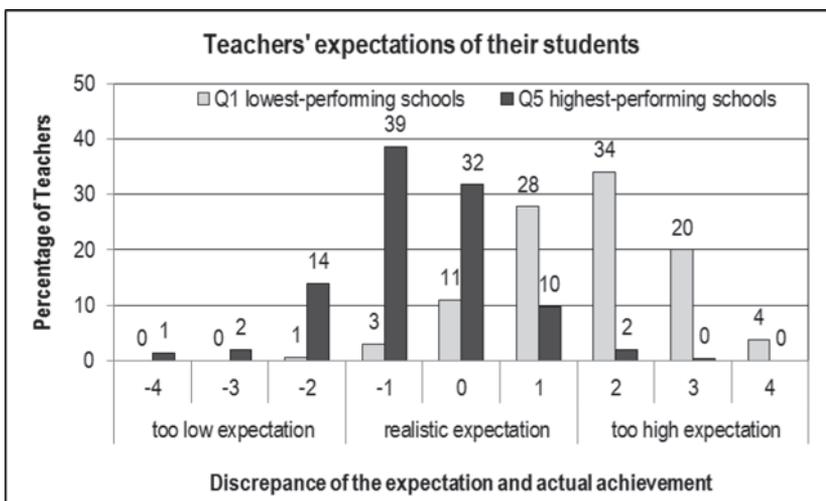
Figure 4.3.13 Teachers' expectations of their students in Q1- and Q5 schools

In the highest performing community schools, 81% of the teachers expressed quite a realistic view of their students. This means that if the students (in fact) were at a high level, the teachers would have expected high or very high of the student and, contrary, if the students were at a very low level, the teachers would have expected low or very low of the student. However, in the lowest level schools, only 42% of the teachers expressed a realistic view of the students and 58% expressed far too high expectations. The difference between the schools is remarkable. The phenomenon is interesting but the reasons are not discussed here. However, it raises the question whether the teachers at the lowest level schools realize at all the low achievement level of their students.

4.3.3.2.9 Limitations of teaching in the school

Two possible directions to find explaining factors for lower achievement are the challenges related to the students or shortages of the elementary facilities in the school. These were scanned by using a set of variables (T23) covering several student- and school resources-related factors (Table 4.3.8).

Table 4.3.8 Difference in the limiting factors of teaching in the Q1- and Q5 schools

Limit teaching, Students:	Lowest quintile school (Q1)	Highest quintile school (Q5)	Sig.
T23a Students with different academic abilities	1.9[1]	1.9[1]	ns.
T23b Students who come from a wide range of backgrounds	2.4	2.3	ns.
T23c Students with specific needs (hearing, vision, etc)	2.7	2.8	ns.
T23d Uninterested students	2.6	2.8	0.002[2]
T23e Low morale among students	2.8	3.0	0.015
T23f Disruptive students	3.2	3.3	ns.
Limit teaching, Shortage of resources:			
T23g computer hardware	1.5	1.6	ns.
T23h computer software	1.5	1.6	ns.
T23i support for using computers	1.6	1.5	ns.
T23j textbooks for student use	3.4	3.5	ns.
T23k other instructional equipment for student use	2.5	2.7	0.024
T23l equipment for use in demonstrations and other exercises	2.4	2.5	ns.

1) scale used: 1 = very much limits - 4 = no limits at all; the higher the value the better situation
2) Statistically significant differences are highlighted.

305

There seems to be some general challenges in teaching with no differences between the Q1- and Q5 schools: students with different academic abilities and shortages in computer hardware and software as well as in support for using computers. Many teachers have seen these matters as "very much limiting" factors for teaching. Statistical difference between the Q1- and Q5 schools appear to be students' low motivation ($p = 0.002$) and morale ($p = 0.015$); there are statistically significant less problems in these areas in the high-performing schools than in the low-performing schools. DTA suggests the cut-offs as follows: for both variables the best cut-off would be "higher than 3" ("not much limits") (Figure 4.3.14).

4.3.3.3 Summary of the teachers characteristics – Modeling the phenomenon

Above, several individual teacher-related factors have been detected which individually explain the difference between the lowest and highest performing community schools. These factors are collected in Table 4.3.9. Some of them may be strongly related to each other and hence they may not add value in explaining why some community schools are performing much better than others. It is also worth remembering that the teacher-related factors are not the only relevant variables in explaining the phenomenon; related school factors are handled in Section 4.4 and student factors in Section 4.5.

4.3.3.3.1 Modeling the highest and lowest quintile community schools with the teacher-related variables

Logistic regression analysis was used to analyse which of the factors are independent predictors for belonging to either the lowest or highest performing schools. When taking at the same time several variables in the analysis, all the cases with even one missing value are taken away from the analysis. Hence, such variables with a remarkable number of missing values were omitted in the analysis. In Table 4.3.9, these variables are marked as a cursive font type. Table 4.3.10 shows the model including all the variables (Enter-selection) and Table 4.3.11 the statistically best model (Conditional-selection). It is notable that actually only a couple of variables really show their own (main) effect on the whole model. These factors are presented below in the tables.

Table 4.3.9 Teacher-related factors differing significantly between the Q1- and Q5 schools

Developmental region Dummy Eastern = 1/ other = 0

Developmental region Dummy Valley = 1/ other = 0

Ecological zone Dummy Mountain = 1/ other = 0

T11 Have you got the teacher's guide of the subject you teach (yes/no)

[1])T11_1 If have the teachers' guide, have used the guide (yes/no) (Missing 80)

T13_1B Evaluation: classwork evaluation, how many times in week Dummy (2 or below = 0, above 2 = 1)

T13_2B Evaluation: homework evaluation, how many times in week Dummy (2 or below = 0, above 2 = 1)

T15 Interaction with SMC improving teaching-learning (yes/no)

T16 Interaction with Teacher staff; how many times during the year, Dummy (7 or below = 0, above 7 =1)

[1])T18a Professional development last 2 years: combined Dummy (6 or below = 0, above 6 = 1) (Missing 138)

[1])T18e Professional development last 2 years: Assessment (Missing 138)

T20b Teachers' understanding of school's curricular goals Dummy (2 or below =1, above 2 = 0)

T20f&g Students' school related activity (0 – 100)

T20a-d Teachers school related activity (0 – 100)

T20f Parents' activity in school activities (1 – 4)

T22g Students do: Relate what they have learned in their daily life (1 – 5)

T23d Limit teaching, Students: Uninterested students Dummy (2 or below = 1, above 2 = 0)

T23e Limit teaching, Students: Low morale among students Dummy (3 or below =1, 4 = 0)

1)   variables are not in the model because of the high number of missing values

307

Table 4.3.10 Logistic regression analysis explaining belonging to the Q1- and Q5 schools

Variables in the equation[1]	B	S.E.	Wald	df	Sig.	Exp(B)
DevRegion Dummy Eastern	-1.957	0.417	22.025	1	< 0.001	0.141
DevRegion Dummy Valley	1.252	0.62	4.073	1	0.044	3.497
EcoRegion Dummy Mountain	0.671	0.395	2.876	1	0.090	1.955
T11 Have you got the teacher's guide	0.189	0.276	0.469	1	0.494	1.208
T13_1B Evaluation: classwork evaluation, how many times in week Dummy (2 or below = 0, above 2 = 1)	0.677	0.292	5.364	1	0.021	1.968
T13_2B Evaluation: homework evaluation, how many times in week Dummy (2 or below = 0, above 2 = 1)	0.696	0.333	4.375	1	0.036	2.005
T15 Interaction with SMC improving teaching-learning	-0.321	0.37	0.753	1	0.386	0.726
T16 Interaction with Teacher staff; how many times during the year, Dummy (7 or below = 0, above 7 =1)	0.643	0.272	5.583	1	0.018	1.902
T20b Teachers' understanding of school's curricular goals Dummy (2 or below =1, above 2 = 0)	0.381	0.332	1.313	1	0.252	1.464
T20f&g Students' school related activity	0.007	0.008	0.636	1	0.425	1.007
T20a-d Teachers school related activity	-0.003	0.013	0.065	1	0.799	0.997
T20f Parents' activity in school activities	-0.01	0.008	1.466	1	0.226	0.99
T22g Students do: Relate what they have learned in their daily life	0.07	0.161	0.189	1	0.664	1.072
T23d Limit teaching, Students: Uninterested students Dummy (2 or below = 1, above 2 = 0)	0.606	0.309	3.838	1	0.05	1.833
T23e Limit teaching, Students: Low morale among students Dummy (3 or below = 1, 4 = 0)	0.491	0.396	1.533	1	0.216	1.634
Constant	-1.565	1.109	1.991	1	0.158	0.209

1) (Method = Enter)

Table 4.3.11 Statistically the best model of Logistic regression analysis explaining belonging to the Q1- and Q5 schools

Variables in the Equation[1]	B	S.E.	Wald	df	Sig.	Exp(B)
DevZoneDummy_Eastern	-2.091	0.401	27.247	1	<0.001	0.124
T13_1B Evaluation: classwork evaluation, how many times in week Dummy (2 or below = 0, above 2 = 1)	0.733	0.281	6.833	1	0.009	2.082
T13_2B Evaluation: homework evaluation, how many times in week Dummy (2 or below = 0, above 2 = 1)	0.812	0.317	6.568	1	0.010	2.253
T16 Interaction with Teacher staff; how many times during the year, Dummy (7 or below = 0, above 7 =1)	0.729	0.257	8.053	1	0.005	2.073
T23d Limit teaching, Students: Uninterested students Dummy (2 or below = 0, 3 or higher = 1)[2]	0.707	0.285	6.168	1	0.013	2.029
Constant	-1.486	0.375	15.697	1	0	0.226

1) (Method = Forward: Conditional)
2) the original scale: 1 = very much problem , 4 = no problem

All the variables in the last model (Table 4.3.11) are statistically significant predictors for a school to belong to either Q1- or Q5 schools. Out of five variables, the most effective predictor is the Eastern Developmental region. The last column on the table, Exp(B), directly tells the "risk" of being in the highest performing school when the value of the Dummy variable is higher. The value 0.124 for the Eastern region indicates that the "risk" of being a Q5 school is minimal. Actually, taking the inverse of the value, $1/0.124 = 8.06$ tells that when knowing that a school comes from the Eastern region, there is 8 times higher risk that it will be one of the *lowest performing* schools in the sample compared with the other Developmental regions. All the other variables show somewhat two times a "risk" of being in the midst of the highest level schools: more than two times a week class room- and homework evaluation, more than seven times a year meetings with the teacher staff for enhancing the teaching-learning processes, and students with a positive attitude towards the school. When taken all the variables at the same time into account, they explain 29.8% of the phenomenon.[42] This means, in practice, that 71% of the schools can be classified correctly though the model is better in detecting the lower-performing schools (81% correct).

---

[42] Nagelkerke R^2 = 0.298

4.3.3.3.2 Modeling the overall achievement on the community schools with teacher-related variables

The methodological decision made to concentrate on profiling only the most extreme schools lead to a situation that actually most of the schools were not in the analysis; usually this is reasonable because nothing can explain the difference between similar schools with the average student achievement. The decision and analysis done above, however, was carried out just to *clarify* the phenomenon; it is expected that the predicting variables found in the analysis above can explain also, to some extent, the general differences of the average achievement level in the schools. Traditional Linear regression analysis with Stepwise regression was used to model the phenomenon; the equated (and shifted) mean score of the school is explained by the same variables are above (see Table 4.3.9). Table 4.3.12 shows the results.

Table 4.3.12 Statistically the best model of linear regression analysis explaining the average of student achievement in the community schools (Method = Stepwise)

	Coefficients[1]				
Model	Unstandardized Coefficients B Std. Error		Standardized Coefficients Beta	t	Sig.
(Constant)	42.211	1.069		39.501	0
DevZoneDummy_**Eastern=1**	-7.593	1.37	-0.227	-5.541	0
T13_1B Dummy Evaluation: **classwork evaluation,** how many times in week (<=2 = 0, >2 = 1)	5.195	1.089	0.193	4.771	0
T16 dummy Intercation with Teacher staff of improving teaching-learning; how many times during the year (=<7 = 0, >7 = 1)	3.185	1.088	0.119	2.928	0.004
DevZoneDummy_Valley=1	7.602	2.33	0.134	3.263	0.001
EcoRegionDummy_Mountain=1	4.734	1.564	0.123	3.026	0.003

1) Dependent Variable: P_Eqd_Total_mean_SHIFTED to the mean of 46.91 for all subjects

The model in Table 4.3.12 can be interpreted as follows: The mean of the community schools is 42.2% of the maximum score (B of the Constant). If the school came from the *Eastern region*, the score was, on average, 7.6 percent points *lower* (note the sign of the coefficient). This means the simple fact that if the school is *not* from the Eastern region, this gives those schools more than 7 percent points' advance in the test compared with the schools from the Eastern region. Additionally, if there

was a *classroom evaluation* three or more times a week, the score was 5 percent units higher than with only two or less evaluations per week; and if there was *interaction with Teacher staff* (TS) of improving teaching-learning, eight or more times a year, the advance was 3 percent points higher compared with the situation of only seven or less interaction sessions. Hence, when the school was *not* from the Eastern region, the teachers were *evaluating the student performance frequently* and there were *frequent teaching staff meetings* to improve the teaching-learning processes, the school would have gained, in theory, *15percent point higher score* than without these. Additional advance comes when the school happen to be situated in the *mountain region* (+4.7) or *Valley* (+7.6). The latter are, however, not the choices of the school (or, of course, the selection of being in the Eastern region either) and hence, these cannot be changed. The differences between the Eastern district and the other districts are deepened in Section 4.1 with the diversity analysis.

It is worth noting that the result does *not* mean that if the school just decided to raise the number of classroom evaluation sessions from 2 to 3 and the number of interaction between teacher staff from 7 per year to 8 per year, the results would change radically. Behind these variables, there lies a set of *other* variables and practices. Just to illustrate the complexity of the phenomenon, Table 4.3.13 shows the differences between the "low-activity schools" (less frequent evaluating of the student performance in the classroom and less frequent teaching staff meetings to improve the teaching-learning processes) and "high-activity schools" (frequent evaluating of the student performance in the classroom and frequent teaching staff meetings to improve the teaching-learning processes) in a rough way.

Table 4.3.13 Comparison of low- and high-activity schools

Teacher-related variables explaining the low- and high activity	Low activity[1]	High activity[2]
t04 Age	older teachers[3]	younger teachers
t05 Mother tongue of the teacher	more non-Nepali	more Nepali
T08 Teaching experience in years	more experience	less experience
T12 Have finished all courses of Gr. 8	less finnished	more finnished
T13_2B Evaluation: homework evaluation, how many times in week	less	more
T13_3B Evaluation: discussion, how many times in week	less	more
T13_4B Evaluation: project work, how many times in month	less	more
T13_5B Evaluation: observation of behavior, how many times in month	less	more
T14b2 External class observation: Instructor; get advises to improve	more	less
T15 Interaction with School Management Committee (SMC) of improving teaching-learning	less	more
T15_1 Interaction with SMC; if yes what and how much it was helpful	less	more
T19c School: I feel safe in this school	less positive	more positive
T19d School: This school's security policies and practices are sufficient	less positive	more positive
T20a School: Teachers' job satisfaction	less positive	more positive
T20b School: Teachers' understanding of school's curricular goals	less positive	more positive
T20c School: Teachers' degree in success in implementing the school's curriculum	less positive	more positive
T20d School: Teachers' expectations of student achievement	less positive	more positive
T20e School: Parents' support for student achievement	less positive	more positive
T20f School: Parents' activity in school activities	less positive	more positive
T20g School: Students' regard for school properties	less positive	more positive
T20h School: Students' desire to do well in school	less positive	more positive
T21d Typical week Lessons: Students are working problems on their own without your guidance %	less	more
T22f Teaching: asks students to do: Students are working together in small groups	less	more
T22g Teaching: asks students to do: Students relate what they have learned in their daily life	less	more
T22h Teaching: asks students to do: explaining the answers	less	more
T22i Teaching: asks students to do: Decide on their own procedures for solving complex problems	less	more
T23m Limit teaching, Resources: Inadequate physical facilities	more	less

1) Less frequent classroom evaluation & teaching staff meetings
2) Frequent classroom evaluation & teaching staff meetings
3) All the differences are statistically significant at least 5% risk (t-test with the individual variables)

On the basis of Table 4.3.13, the situation looks logical: Compared with the low-activity schools (which explains the lower results), in the high-activity schools (which explains the higher results), the teachers are *younger*, they *evaluate the students more frequently*, they are *more active with the SMCs*, they are *more content with their work* (which may be explained by the good results, perhaps), the other *teachers seem to understand the educational goals* better and they *expect students to succeed*, the *parents are more active* involved in the school, and the *students are working constructively to reach their goals*, the students are lead to *more constructive learning experiences* (by giving them more possibilities to work with their own problems and in small groups) as well as lead to *more complex thinking processes* (by letting them explain the answers and decide their own procedures for solving complex problems) and *more practical solutions* (by giving them possibilities to relate what they have learned in their daily life).

4.3.4 Conclusions

Several teacher-related variables were detected which explain, at least to some extent, the differences between the low- and high performing community schools. Some of the variables showed a high percentage of missing values; these were omitted in the analysis. As condensed, the main results related with the teachers are as follows:

- The teacher/school effect in Nepal is 68%; this is a very high value. This means that the students within the schools are very much alike and, simultaneously, there are wide differences between the schools. The effect size is much less in Mustang (9%) and in Kalikot (30%) where the test was administered so that the possibilities for the "social work" were minimized. (see Chapter 4.3.3.1).

- The highest achievement seems to come from the schools with the teachers aged less than 32 years. Younger age also seems to relate with more active behavior in the schools. For one reason or another, the results are significantly lower when the teachers are more than 48 years old. (Chapters 4.3.2.4 and 4.3.3.3.2)

- Higher achievement seems to be in the schools with Nepali speaking teachers. (Chapter 4.3.3.2.2 and Chapter 4.3.3.3.2)

- Mathematics teachers' confidence in their own subject matters was higher in the higher performing schools. (Chapter 4.3.3.2.4)

- In the higher achieving schools, teacher's understanding about the

curricular goals is higher. In the active schools, job contentment was also higher. Students' and parents' activity is higher in the higher achieving schools. (Chapters 4.3.3.2.3, 4.3.3.2.5, and 4.3.3.3.2)

- In the higher achieving schools, the teachers are more actively communicating with the SMCs and other teaching staff to enhance the teaching learning processes as well as being more active in evaluating their students. (Chapters 4.3.3.2.6, 4.3.3.3.1 and 4.3.3.3.2)

- In the classroom teaching activities reported by the teachers, there are no practical differences between the low- and high-performing schools. However, indirectly, in the active schools and hence in the higher performing schools, the students are lead to more constructive learning experiences (by giving them more possibilities to work with their own problems and with small groups) as well as lead to more complex thinking processes (by letting them explain the answers and decide their own procedures for solving complex problems) and more practical tasks (by giving them possibilities to relate what they have learned in their daily life). (Chapters 4.3.3.2.7 and 4.3.3.3.2)

- In the higher performing schools, the students are evaluated more frequently. The frequent and many-faceted evaluation is also more typical of active schools. In the high-performing schools, the teachers have radically more realistic perception of the students than in the low-performing schools. (Chapters 4.3.3.2.8, 4.3.3.3.1, and 4.3.3.3.2)

- In the higher-performing schools, there are less unmotivated students and their morale (assessed by the teacher) is higher-ranked. (Chapters 4.3.3.2.9)

An additional, not teacher-related, a factor is that the most effective predictor of being a low-performing school is that it origins from the Eastern Developmental region. This is discussed in detail in Section 4.1.

References for Section 4.3

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

BPEP (1998). The Effect of New Curriculum on the Achievement of Grade 5 Student.

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

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

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

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

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

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

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

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

Fullbright (2008). *A final report on national assessment of grade-V students*, Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education. Fullbright Consultancy & Cooperative Hands in restoration (CHIRAG).

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika,* **73**, 43–56.

Kansanen, P. (2003). Teacher education on Finland: current models and new developments. In M. Moon, **L. Vlãsceanu, & C. Barrows** (Eds.), *Institutional Approaches to teacher Education Within Higher Education in Europe: Current Models and New Developments*. Bucharest: Unesco-Cepes, 85–108.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics,* **29**(2), 119–127.

Metsämuuronen, J. (2013a). *Handbook of Basics of Researcher Methods in Human Sciences*. 5th Edition. Oy International Methelp Ltd. E-book. [Beta version available at www.Booky.fi]

Metsämuuronen, J. (2013b). *Changes in student achievement from grade 3 to grade 9.* Koulutuksen seurantaraportit 2013:4. Opetushallitus. Helsinki: Edita Prima. [In Finnish].

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).

Niemi, H. (2011). Educating student teachers to become high quality professionals – A Finnish case. *Center for Educational Policy Studies Journal*, **1**(1), 43–66.

Niemi, H. (2010). Teachers as high level professionals – What does it mean in teacher education? Perspectives from the Finnish teacher education. In **K. G. Karras & C. C. Wolhuter** (Eds.), *International Handbook of Teacher Education: Issues and Challenges* (Vol. l & ll, pp. 237–254). Athens Greece: Atrapos.

Niemi, H. & Jakku-Sihvonen, R. (2011). Teacher education in Finland. In **M. Valenèiè Zuljan & J. Vogrinc** (Eds.), *European Dimensions of Teacher Education: Similarities and Differences* (pp. 33–51). Slovenia: University of Ljubljana & The National School of Leadership in Education.

Niemi, H. & Jakku-Sihvonen, R. (2006). Research-based teacher education in Finland. In **R. Jakku-Sihvonen & H. Niemi** (Eds.), *Research-Based Teacher Education in Finland – Reflections by Finnish Teacher Educators* (pp. 31–51). Turku: Finnish Educational Research Association.

Niemi, E. K. & Metsämuuronen, J. (Eds.) (2010). *Learning Outcomes and Changes in Achievement in Mathematics.* Oppimistulosten arviointi 1/ 2010. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

PEDP (1998). Primary Education Development Project: An assessment. 1998. Kathmandu.

Sahlberg, P. (2011a). The Professional Educator: Lessons from Finland. *American Educator* **35**(2), 34–38.

Sahlberg, P. (2011b). Lessons from Finland: Where the Country's Education System Rose to the Top in Just a Couple Decades. *Education Digest*, **77**(3), 18–24.

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

Schleicher, A. (2011). Is the Sky the Limit to Education Improvement? *Phi Delta Kappan,* **93**(2), 58–63.

4.4 School Effect in Learning

Dr. Tulshi Thapaliya[1]

&

Dr. Jari Metsämuuronen[2]

1. Ministry of Education, Nepal

2. Helsinki University, Faculty of Behavioral Sciences, Finland

Abstract

This article deepens the knowledge of NASA 2011 by combining all the datasets into one and by explaining the differences in student achievement with school-related variables. The results are based on 1200 head teachers' background questionnaires and demographic information. The highest performing community schools are compared with the lowest performing schools. The researcher question is: what has been done differently in the highest performing schools to gain the high results?

The school effect in Nepal is 68%; this is a very high value. The student performance in the Eastern region is much lower than in other regions; the lowest performing districts come from the Eastern region, where the number of high-performing schools is radically lower than in the other regions, and the average number of schools days is also significantly lower (138 days) than the average in the other districts (152 day). A data mining tool suggests that, in Nepal, the optimal student/teacher ratio would be 31–34 students per one teacher. However, the highest achievement level is found in the schools with the student/teacher ratio of 21 or lower. The physical characteristics of the school do not explain why some schools are better than the others. In the higher performing schools, there are more computers and computer software in use than in the lowest performance schools. Those schools also have more material for disabled students and a wider repertoire of facilities for Mathematics teaching.

In the lowest performing schools, there are many more negative incidents involving students, and their seriousness is greater than in the highest performing schools. Also the working atmosphere, reported by the teachers, is significantly lower in the lowest performing schools than in the highest performing schools. In the lowest performing schools, the number of classwork- and homework evaluation sessions is lower than in the highest performing schools. The teachers in the highest performing schools participated in several kinds of professional development processes more than the teachers in the lowest performing schools. This may refer to either the better possibilities or stronger motivation of the teachers to take advantage of continuous education opportunities in the highest performing schools. The high-performing community schools are more active in improve the teaching learning processes by more intense teacher staff- and SMCs interactions as well as in co-operation with the local community compared with the low-performing schools.

Keywords: School effect, Low-performing schools, High-performing schools, School atmosphere, Teachers' continuous education

4.4.1 Introduction

Differences in the learning outcomes between the specific groups of the students are described in the literature as an "achievement gap" (originally in Reynolds *et al.*, 2002; see also Murphy, 2010). Often such comparisons are made between the groups that have some distinct advantages or disadvantages in education; the social interest or the whole purpose of such discussions is to concentrate on efforts to reduce such gaps in favor of those who are in a disadvantageous position. In other words, none of the variables like the birth place, the language of origin, the caste / ethnicity or the sex or the socioeconomic status of the family should define the future of these children (see Section 4.1 for a detailed discussion of these "equity indicators"). On the top of these equity indicators, several other variables may also explain the learning achievement. Section 2.3.1 introduced, on the basis of Metsämuuronen (2009), a sketchy model of the directions to find variables which may explain the differences in students' achievement. One possibly influential set of factors of the student performance are school-related variables. Usual concepts used in regard with school-related variables are the "School effect" and "Effectiveness of School". This branch of studies is quite wide – alone in the ERIC database (www.eric.ed.gov) one finds 7753 publications with the keywords "School effectiveness", "effectiveness of School", "School effect", or "effect of School".[43]

4.4.1.1 School effectiveness

Generally speaking, a school's effectiveness can be understood as the success of an organizational unit called school producing the desired level of output (Scheerens, 2000). Because of the complexity in measurement, school effectiveness has been accounted through a *relative* measure: comparing the performance of a school or a school system, with some curricular standards, with past performance, or with some other schools or school systems. It is rarely possible to use the *absolute* measure; it is practically not feasible to measure and compare the change in characteristics of the population who attend school with those who do not. The important question is to establish the criteria for judging the effectiveness of schools.

In many of the studies *student achievement* is considered as one of the most important criteria across the cultures (see, for example, PISA, 2001; 2003; 2005; 2007; 2010a; 2010b; Olsen, Martin & Mullis, 2008). Because student achievement is, in many cases, and specifically in the private schools, related to

---

[43] This was the situation on the 9th of July 2013 – the figures are growing constantly.

the selection processes (namely, the low-performance students are *not* selected in) a better measure for the achievement would be the *change* in the student achievement within a school (Hautamäki & Kuusela, 2005; Metsämuuronen, 2009). Hence, the more change in achievement the school would produce during a certain time frame (for the students originally at the same level), the more effective the school would be. However, it is somewhat rare to see these kinds of studies because they are difficult to organize in a feasible way.[44] The other set of relevant measures of school outcomes are the *behavioural indices* (such as student and teacher attendance and retention), *affective scales* (such as Self-Efficacy, the Locus of control; etc.) and measures of *social growth* (Teddlie, Reynolds & Sharon, 2000). These alternative measures are used by researchers just to broaden the school outcomes and not as a substitute for student achievement.

4.4.1.2 School effect

'School effect' as a concept refers to the different learning outcomes between the schools. Technically speaking, the school effect, as used in international studies, is defined as intra-class correlation related to the schools. Practically speaking, this means that when the students' achievement levels within the schools are quite similar and there are radical differences between the average achievement levels of the schools, it may indicate that the school has produced higher- or lower results. Hence, it may be indicative of more effective processes in one school compared with another school. If, on the contrary, the results within the school were deviating (that is, there are both low- and high-performing students within the same school) and the differences between the schools are small (as it is in Finland, see Schleicher, 2006, 13), the school effect will be low; i.e. the school cannot explain the differences in students' achievement levels**.**

From the methodological viewpoint, the question of school effect is somewhat difficult – in many cases it is difficult to separate the school effect from the teacher effect and in some cases the selection of the students to the school may explain the differences between the schools rather than the processes in the school. International studies – such as PISA and TIMSS studies – have shown remarkable differences between countries when it comes

---

[44] The methodological challenge comes from the fact that the change cannot be measured without (at least) two measurements from the same students. In the real-life situation related to the national-wise assessment – as NASA 2011 is – this requires that (at least part of) the same schools with the same students should be selected two times in the sample in the consecutive measurement points (say, grade 5 and grade 8). This is somewhat costly though not impossible. In Finland, Metsämuuronen (2010) reported one such comparison of the development of grade 3 students to grade 6 with 300 schools and 4000 students and another where he followed up the same students from the grade 3 to 9 (Metsämuuronen, 2013b); these are reported, however, in Finnish.

to the magnitude of the school/teacher effect. In Finland, for example, where the differences between the schools are very narrow, the school effect ranges from 9 to 12% (see Metsämuuronen, 2010; 2013b), that is, the processes in the school explains only around 10% of the student variation at grade 9 (15 years olds).[45] In Nepal, the school/teacher effect may be much higher because of the wide difference between the average achievement levels in the schools (see Sections 3.1–3.3 and 4.3). The way the teacher effect is mathematically defined restricts the discussion only to the difference between the teachers *after the children are attending school* – it does not indicate the overall effect of schooling compared with the situation where children do not attend the school at all as discussed above.

4.4.1.3 What makes schools different?

One way to see the differences between the schools is to divide the differentiating factors into Contextual factors, Inputs, and Processes. In Nepal, the *context* refers to the differences in the geographical factors of the school such as the Developmental region, the Ecological zone, and the location of school (urban and rural) or sociocultural factors such as socioeconomic status and awareness for education, the main language spoken at home of the students and teachers in the school, the main castes of the families within the school, or the type of the school (community or institutional). The *Input* factors can be divided in terms of the numbers and qualifications of teachers, school infrastructure, and financial resources. Some of the context factors (such as the family background) could be understood as input factors, too. The *Process* factors in schools can be further classified into two: first, into the *school managerial processes*, such as general school leadership, school culture, focus on instructional matters, use of instructional time, and so on, and, secondly, on the processes that directly relate to *classroom instruction* related with individual teachers. The second set of factors is often described as teacher effects rather than school effects (see Section 4.3 where the teacher effect is handled).

---

[45] The school effect in Finland varies somewhat with the school subjects, age group and language groups. For example, at the grade 6 mathematics in 2008, the average school effect was 13%, but in the Swedish-speaking minority in the rural area the effect was 34% (Metsämuuronen, 2010). This means that the Swedish-speaking pupils within a school were much more like each other compared with the Finnish-speaking pupils and there were some differences between the schools. In the urban areas the effect was 10% in the Finnish population and 11% in the Swedish population; the low- and high-performing students in the cities were more mixed within the classes. Previously, Metsämuuronen (2013b) showed that in Mathematics the school's effect in the Finnish population (majority) was round 3.5% in the cities and population density areas whereas in the rural area in the Swedish population (minority) it was 22%. The difference is notable.

Researchers continued to search for these 'process factors' that are closely associated with improved school achievement. Further, through reviews and meta reviews, there were attempts to identify and synthesize the findings regarding the critical factors for school effectiveness. For instance, Scheerens (2000) has provided a list of effectiveness-enhancing factors from earlier studies in developed and developing countries. Thapaliya (2011), reviewing Liu (2006), discusses the five such factors that were popularized by Edmonds long back in 1979. The key factors of effective schools that many of the earlier studies on effective schools converged to are: 1) strong educational leadership, 2) high expectations on pupils' achievement, 3) safe and orderly environment, 4) an emphasis on the acquisition of basic skills, and 5) frequent monitoring of student progress.

There are few (if any) pieces of research that tell us about the school effect on the learning achievement in Nepal in the same manner of the international studies. However, the Ministry of Education (MOE) in Nepal has carried out various processes to enhance the educational practices and national assessment projects to monitor the learning outcomes.[46] In order to assess the outcomes of the enhancing processes, several National Assessment projects have also been carried out (see BPEP, 1997; EDSC, 1997; BPEP, 1998; PEDP, 1998; EDSC, 1999; CERID, 1999; EDSC, 2001; 2003; CERSOD, 2001; EDSC, 2008; Fulbright, 2008). The student results of these projects are not, unfortunately, strictly comparable with each other because of the missing linking procedure between the tests (see Section 2.5). Though the reports have indicated various conditions which have some effect on achievement and have highlighted various recommendations, there is no straight study on the teacher effect on the student achievement in Nepal.

This article shows how strong the school effect is on the learning achievement in Nepal when using the same methodology as is used in the international studies. It tries to identify the effectiveness of school activities and processes by comparing the processes in the best performing schools with the lowest performing schools. As an outcome, the final aim is identifying the characteristics of an effective school in Nepal. Research questions are set as follows:

1. *How strong the effect of the school on learning achievement is in Nepal?*
2. *What kinds of activities and processes are carried out in the schools of the best performing schools compared with the lowest perform-ing schools?*
3. *What are the characteristics of effective schools*?

---

[46] Several teacher-related projects of MOE are referred to in Section 4.3. The latest project, School Sector Reform Plan (SSRP), was developed in 2009 by the MOE and it was launched for the fiscal year 2009/2010. NASA 2011 is the first assessment project of SSRP.

4.4.2 Methodological solutions

The analysis is based on the combined data comprising round 1,200 schools and 50,000 students covering the subjects of Mathematics, Nepali, and Social Studies. The sampling, achievement tests, background questionnaires, reliabilities and validities of the tests, as well as the other technical matters are discussed mainly in Section 2; they are not repeated here. However, the research design, conceptual framework, variables used in analysis, and the statistical method are discussed here in more detail.

4.4.2.1 Design

It is quite a challenging task to estimate the effect of schooling in children, especially when children with varying backgrounds attend schools with visibly distinct characteristics. In Nepal, especially in the urban areas, students are segregated by the socioeconomic status of their parents: people who can afford to, seem to choose the private schools leaving others to free or highly subsidized community schools. The general performance pattern, which is also supported by this study, is that the private schools do better in terms of measured student achievement; this can be explained by the selection of the students as discussed above.

A basic reality of the mean achievement level of the schools in Nepal is very interesting from the comparison viewpoint. There are two distinctive populations of schools in each sample of different subjects – schools with the lower- and higher socioeconomic status (SES). Another distinction comes from the mean achievement in the school: lower-performing schools and higher-performing schools (see, for example, Figures 3.1.5 and 3.1.6 in Section 3.1.3). The situation does not change when taking into account all the 1200 schools into the analysis (Figure 4.4.1a): most of the institutional schools are performing very well but the community schools vary from the very high-performing schools to the very low-performing schools. When simplifying the situation and by focusing on the highest and lowest performing schools, the phenomenon looks as in Figure 4.4.1b.
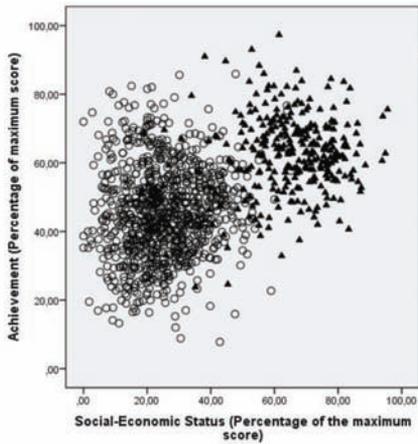
Figure 4.4.1a Connection of SES and the mean achievement in the school
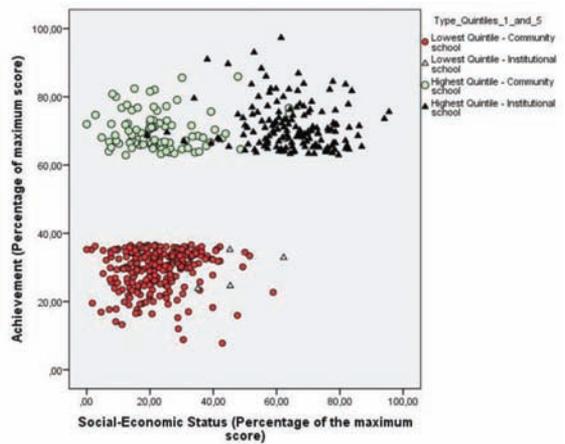
Figure 4.4.1b Connection of SES and school achievement – the lowest- and highest quintiles

In NASA 2011 dataset, there are very few private schools where the learning outcomes are very low except for a couple of interesting schools; mainly, the institutional schools are in the highest performing quintile. The community schools vary from low- to high-performing schools. From the methodological point, it is possible to form two sets of comparisons as visualized in Figure 4.4.2.
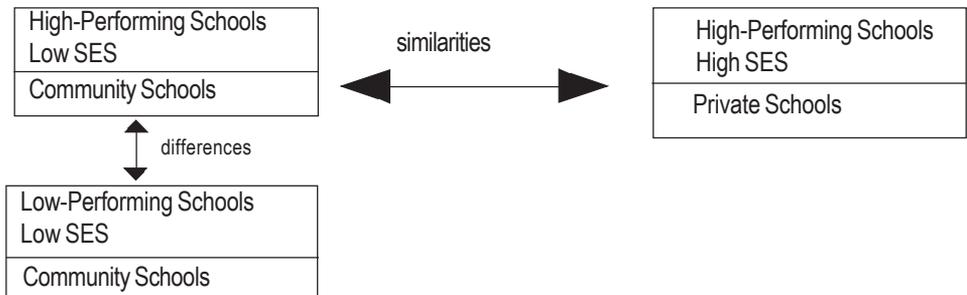
Figure 4.4.2 Research design

With the exception of some rare cases, the economic and educational background of the families (SES) is somewhat the same in all the community schools. Nevertheless, the learning outcomes are radically different when it comes to the lowest and highest quintile community schools (in what follows, Q1 and Q5

323

respectively). Then an adequate question is what happens differently in the community schools of the highest quintile compared with the community schools in the lowest quintile? Another comparison comes from the fact that results in the highest performing schools with the high educational and economic status (private schools) do not differ from those without the economic and educational support (high performing community schools). An adequate question is, what are the common elements shared by the high performing schools? The latter question is not, however, handled here because of space. It seems evident, however, that the community schools in the highest quintile are effective schools: with much lower input they are able to produce the same results as in the private schools with much more input. This article tries to find the answer to the questions above by addressing the schools' processes and characteristics; Section 4.3 focuses on the teacher-related factors and Section 4.5 on the student factors.

For a more effective statistical analysis the division into quintiles was made again within the community schools; originally out of 250 schools in the highest performing group there were only 102 community schools (the rest were private schools) but in the lowest performing group there were 244 community schools.[47] After the new division, there were the 187 lowest and 187 highest performing community schools to be compared. Some of the schools (6 + 6) did not provide us with the head teacher or teacher information. However, these schools gave the student information and demographical information. The number of schools (n = 374) is sufficient for the most statistical analysis. The variables found by focusing on the extreme schools only are used, at the final phase, in explaining the differences within all the schools.
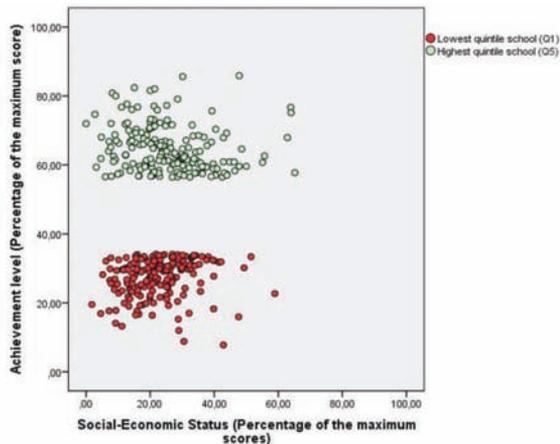
Figure 4.4.3 Distribution of the community schools in comparison (n = 383)

---

[47] The situation is somewhat more complicated for the comparison. Namely, from several schools, the students did the test and answered the background questionnaire. More, the students might have been quite good in Mathematics but not in Nepali – or vise versa. In 36 cases, there was no teacher- or head teacher information available. These schools were mainly omitted from the analysis. If the school participated the assessment in 2–3 subjects, the head teachers' information is replicated while the student means are different.

4.4.2.2 Conceptual framework of the school factors in NASA 2011

In NASA 2011, there are a number of questions for the Head Teacher (HT) related to the school that can be used to analyse the effectiveness of school on student achievement. Figure 4.4.3 shows the conceptual framework of the HT background questionnaire. Some of the factors are also added from the Teachers' and Students' questionnaire to detect the differences between the low- and high-performing schools.
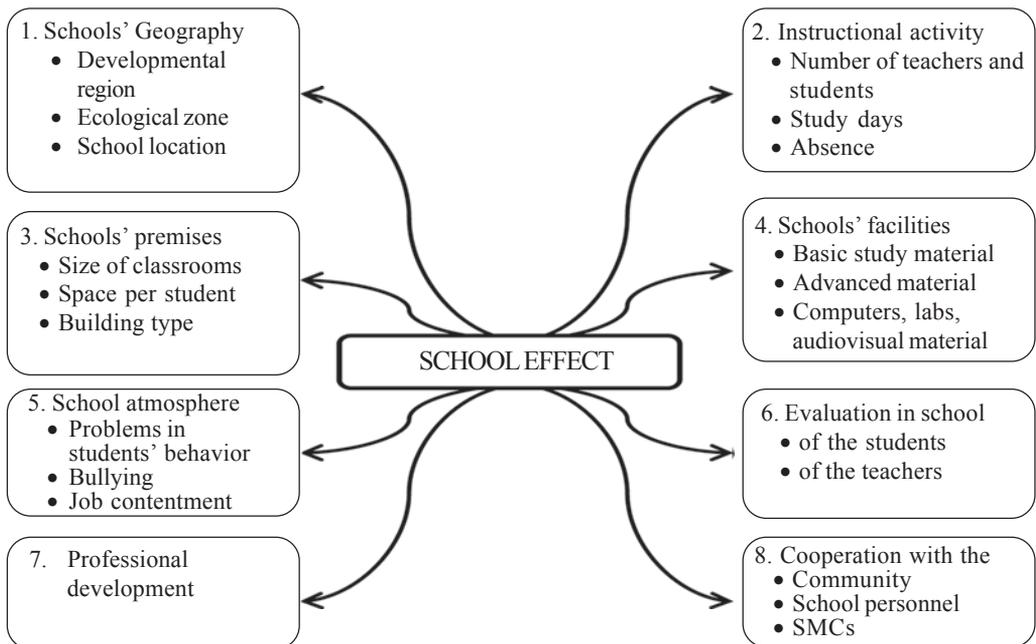
Figure 4.4.4 Conceptual framework of the school factors

The framework is not discussed here, it can be criticized and debated, however, a civilized guess is that all the factors in Figure 4.4.4 may be valuable when explaining the differences in students' learning outcomes in the different schools. Nevertheless, it is good to note that, for all the students in the same class, the teacher is behaving in exactly the same manner and the school remains the same but some students just do not learn whereas some others do. In many cases it might be too much to say that these characteristics or processes in the school are the real reason for the learning. However, even a small shift in students' achievement levels may be valuable and a secondary factor may be a valuable asset in the hands of a wise head teacher.

4.4.2.3 Variables used in analysis

In what follows, the school effect is estimated in terms of variation of student achievement throughout the sample. Student achievement is estimated as the latent ability (θ) by using IRT modeling (see Section 2.5). In the first phase, all the different versions within each subject were equated – that is, their difficulty levels were adjusted to make them comparable – after which all three student datasets were merged. It proved that the Mathematics test was more demanding in the population than the other tests.[48] Hence, without any further transformations, the lowest schools would automatically be those schools with the Mathematics test. That is why the schools' scores were further transformed by using so called 'mean equating', which means that the means of the different subject populations were shifted so that all the distributions had the same mean.[49] Technically, the mean of the means (46.91) was taken as the baseline for the transformation. All the school means were shifted depending on the subject tested in the school; Mathematics schools were shifted up and Nepali and Social Studies were shifted down. In the graphs and statistical outputs one may see a variable label 'P_Eqd_Total_mean_SHIFTED to the mean of 46.9 for all subjects'**.**

The equated and shifted total score in this full data is used for estimating the school/teacher effect with Multi-level modeling (Goldstein, 1986) or Hierarchical linear modelling (Bryk & Raudenbush, 1987). The large student dataset of approximately 50,000 students was further aggregated to the school level with a total of 1200 units. Hence, the students' mean in each school forms the estimated "schools proficiency level". It is worth noting that the individual school mean in the aggregated data no longer carry the information whether the results came from a small school or a large school; information on the school size was stored as an independent variable**.**

Another variable derived from the previous one is the indicator of the quintiles (Q1 and Q5). Q1 schools were given the code 0 and Q5 schools the code 1. Hence the variable can be used two ways. First, it is a nominal variable which groups the schools into two categories. This way the variable can be used as a fixed factor in ANOVA modeling; as such one asks whether there is a significant difference between the groups with regard to some other variables, such as the

---

[48]   In the original data of all students, the mean in Mathematics was 42.8, in Nepali 48.6, and in Social Studies 49.3.

[49]   In the 'mean equating' one assumes that the distributions are equal but that their locations are different. Now it is known that the distribution of mathematics subject differed from the other subjects (see Section 3.1.3). It was taken as a minor thing compared with the radically lower mean in the population. Other option would have been so called 'linear equating' which is used when there is a difference between the means as well as between the variance.

attitudes of the teachers. Second, the same variable can be interpreted as an interval variable; so that the mean of the variable tells the proportion of better achieving schools in relation to some grouping variable, such as the language background of the teachers. This way the variable can be used in DTA modeling (see the next Section); in this way one can ask whether there is a significant difference between the proportions of the highest performing schools in different classes based on the grouping variable.

4.4.2.4 Statistical methods

Four main methodological tools are used in the analysis: Multilevel modeling (MLM), Decision tree analysis (DTA), Analysis of Variance (ANOVA) and both Logistic- and Traditional Linear regression analysis (LRA and RA). MLM is used only in estimating the teacher effect as the proportion of two variance components: teacher-related variance and student-related variance. MLM is run by using the original Theta value and the full student data (around 50,000 students) – the Theta values are comparable over the subjects; zero refers to the average students in each subject. ANOVA, LRA, and RA are basic methods referred in many textbooks (see, for example, Metsämuuronen 2013a) and hence they are not described here.

DTA is one of the data mining tools in SPSS. It is used when desiring to segment the data into statistically the most significantly deviating groups. Figure 4.3.4 shows an example of a DTA output.

In the example, on the left hand side the Q1 and Q5 schools have been analyzed; the proportion of the highest performing schools (**Q1_5_NEW2_COMMUNITY**) has been used as the *dependent* variable and the average number of study days (pq6a) as an *independent* one. The first box, the "mother node", tells that in the data the proportion of the Q5 schools is 49% (0.491), standard deviation $s = 0.501$, and the number of schools where the answer was given is $n = 344$. In the statistical process, the variable pq6a has been divided in countless ways to find the grouping which produces statistically the most significant difference between the groups in regard to the variable "proportion of the highest performing schools". Output, that is, two "child nodes", tells that in regard to maximizing the number of the highest performing schools, the average days for studying can be divided into two groups by using the following cut-offs: 182 days or lower (<= 182.0) and 183 days or more (>182.0). The number of highest performing schools is lower (39.7%) if there are 182 (or less) study days and higher (58.8%) when there are more than 182 study days. This division of average study days to exact these two groups produces the statistically highest test statistics (here F = 13.05 with the degrees of freedom of $df_1 = 1$ and $df_2 = 342$) which produces the lowest p-value (here as an adjusted one: Adj. P-value = 0.003).

327

Proportion of Q5 schools

Q1_5_NEW2_COMMUNITY_
Based_on_more_schools USE
THIS

```
           Node 0
  Mean          0,491
  Std. Dev.     0,501
  n             344
  %             100,0
  Predicted     0,491
                        [-]
```

pq6a study day
Adj. P-value=0,003, F=13,048,
df1=1, df2=342

```
        <= 182,0              > 182,0
```

```
          Node 1                       Node 2
  Mean          0,397         Mean          0,588
  Std. Dev.     0,491         Std. Dev.     0,494
  n             174           n             170
  %             50,6          %             49,4
  Predicted     0,397         Predicted     0,588
```

Number of study days

pq6a study day

```
           Node 0
  Mean          183,721
  Std. Dev.     23,051
  n             344
  %             100,0
  Predicted     183,721
                         [-]
```

Q1_5_NEW2_COMMUNITY_
Based_on_more_schools USE
THIS
Adj. P-value=0,002, F=9,815,
df1=1, df2=342

```
        <= 0,000              > 0,000
```

```
          Node 1                       Node 2
  Mean          179,943       Mean          187,633
  Std. Dev.     23,977        Std. Dev.     21,428
  n             175           n             169
  %             50,9          %             49,1
  Predicted     179,943       Predicted     187,633
```

Figure 4.4.5 Two examples of DTA output

In the example on the right-hand side, the question is set in a different way with the same variables as in the example on the left-hand side. Here, it is asked how the division into Q1- and Q5 schools explains the number of average study days. DTA shows that there is a statistically significant difference between the schools: in the lowest level schools, the average days that students are present is 180 days and in the highest level schools it is 188 days. The difference is statistically significant ($F_{(1, 342)} = 9.81$, $p < 0.002$). In what follows, the latter question is analyzed by using Analysis of Variance (ANOVA), or t-test, rather than DTA.

DTA does not explain the reasons for the segmentation; it just gives the cold fact. In what follows, DTA is used with the CHAID algorithm (Kass, 1980). When the graphs are wide, as they are easily with several variables, only selected parts of the graph are shown, or the results are explained verbally.

4.4.3 Results

4.4.3.1 School effect

The school effect is estimated by using the whole student dataset of 50,000 students and by knowing from which school the students came from. From all the school, there came only one teacher paper and in most cases there actually was only one teacher of Mathematics, Nepali or Social Studies. In larger schools, this is not true and hence, the estimation is somewhat crude when assessing the teacher effect (see Section 4.3). The figures tell the school effect more accurately than the teachers' effect though they can be very close to each other. The original estimate for the latent ability (Theta) is used as a basis of the calculation.[50] MLM with Restricted Maximum Likelihood Estimation (RMLE) gives the following results:

Estimates of CovarianceParameters[a]		Estimate	Std. Error
Residual		,386516	,002509
Intercept [subject = New_sch_code]	Variance	,831243	,034640

a. Dependent Variable: Theta_TOTAL.

In the output, the variance related to school is $\tau_{00} = 0.83$ and the residual variance related to students is $\sigma^2 = 0.39$. On the basis of these figures, the value of the intra-class correlation (Rho) is

$$\rho = 0.831243/(0.831243 + 0.386516) = 0.68$$

In other words, the variable SCHOOL explains 68% of the student variation. This is remarkably high value – much higher than in Finland, for example. The result means that the students within the individual schools are very much alike; value ρ = 1 would mean that they were identical within the schools. When knowing that there is a wide difference between the schools and that the schools/teachers effect is as high as 68%, the result reflects a huge inequity between the students and their opportunities to receive equal educational possibilities. Practically speaking, all students in low-performing schools get low results and in the high performing schools all get good results. The reality is quite opposite in Finland, for example,

---

[50] If using the equated score as a basis for the calculation/estimation, the results would have been 59%. However, the equated score, compared with the original Theta, is cruder; in the equating process, several different values of Theta were transformed to be as equal score. Hence Theta gives a more accurate picture of the students' performance than the equated score. The values of Theta are difficult to report as they are standardized values, though. Hence, the equated scores (and not Theta) are used otherwise in the report.

where in each class there are very good students as well as very poor students and hence, in all schools the results are practically equal. In Table 4.4.1, the estimated intra-class correlations in the different parts of the country are collected.

Table 4.4.1 Intra-class correlation in Developmental regions and Ecological zones

Developmental region	Ecological zone	Intra-class correlation
Eastern	Mountain	0.49
	Hill	0.69
	Terrai	0.69
Central	Mountain	0.60
	Hill	0.61
	Terrai	0.78
Western	Mountain	0.091
	Hill	0.61
	Terrai	0.66
Mid-Western	Mountain	0.311
	Hill	0.66
	Terrai	0.76
Far-Western	Mountain	0.71
	Hill	0.74
	Terrai	0.64
Valley	Valley	0.60

Three notes of the table may be worth highlighting. First, in Mustang (Western Mountain area), the effect size seems to be at the same level as in Finland – less than 10%. It can be interpreted in a way that regardless of which school the children are in, the results would be equal though the students' achievement may vary within the school. Second, remarkably lower effect than the average can be found also in Kalikot (30%) and to some extent in Sankhuwasabha (49%). All these districts come from the Mountain zones. The reason for the low school effect in Mustang and Kalikot may be that the schools are usually small and hence, the students may be more or less independent actors in the school. Another option for explaining the result is that the test was administered in a slightly different way in those two districts; in Mustang and Kalikot, all three subjects were tested in the same class at the same time. This means that the effect of "social work" in the testing situation was minimized. Third, the most unequal areas seem to be Central Terrai (78%), Mid-Western Terrai (76%), Far-Western Hill (74%), and Far-Western Mountain (71%) areas. This may refer either to a strict selection of the students in the schools on the basis of their achievement level or to a high level of "social work".

4.4.3.2 Schools characteristics and processes related to high- and low performance

In this section, several sets of variables are analyzed to find out what kind of difference can be found between the highest and lowest performing community schools in the highest and in the lowest quintile. The variables are segmented on the basis of Figure 4.3.2 into eight groups: 1) Schools' geographical characteristics, 2) Instructional activities, 3) Schools' premises, 4) Schools' facilities, 5) Schools' atmosphere, 6) Evaluation and monitoring in the school, 7) Professional development, and 8) Cooperation with the community and school personnel.

4.4.3.2.1 Geographical differences between the high- and low-performing schools

From the statistical point of view, the Q1 and Q5 schools are randomly distributed to rural- and urban schools; there is no statistically significant difference between rural- and urban schools when it comes to the number of Q1 and Q5 schools (Table 4.4.2a). [51] From the equity viewpoint, this is a good signal. There are differences between the Ecological zones though[52]– practically because the Valley differs remarkably from the others. However, the differences are much milder than between the Developmental regions (see Table 4.4.2b).[53].

Table 4.4.2a Ecological zone and School location of the Q1 and Q5 schools

	Ecological zone				School location	
	Mountain	Hill	Valley	Terrai	Rural	Urban
1 Lowest quintile (Q1)	40.4	53.1	15.4	57.5	50.2	48.2
5 Highest quintile (Q5)	59.6	46.9	84.6	42.5	49.8	51.8
Binomial probability (p=0.5)	0.028	0.242	<0.001	0.067	0.460	0.382

Table 4.4.2b Developmental region of the Q1- and Q5 schools

	Developmental region					
	Eastern	Central	Western	Mid-Western	Far-Western	Valley
1 Lowest quintile (Q1)	84.50	38.50	34.10	63.90	47.20	15.40
5 Highest quintile (Q5)	15.50	61.50	65.90	36.10	52.80	84.60
Total	100	100	100	100	100	100
Binomial probability (p=0.5)	< 0.001	0.010	0.001	0.003	0.309	< 0.001

---

[51] $\chi^2(1) = 0.21$, exact $p = 1.00$
[52] $\chi^2(3) = 18.58$, $p < 0.001$
[53] $\chi^2(5) = 71.59$, $p < 0.001$

There are more Q1 schools in the Eastern- and Mid-Western region than what is expected in a random case[54] and more Q5 schools than expected in a random case in the Valley, Western-, and Central regions.[55] Except in the Valley, where the differences are wide, only in the Mountain region can one find a statistically significant difference between the Q1 and Q5 schools; there are more high-performing schools in the Mountain region than is expected in a random case ($p = 0.028$). The Eastern region and the Valley seem to represent opposite realities; the ratio of Q1- and Q5 schools is 85:15 in Eastern region and 15:85 in Valley

The Eastern region seems to be interesting from the student achievement viewpoint; the student performance is much lower than in other regions, most lowest performing districts come from the Eastern region, and the number of highest performing schools is radically lower than in the other regions. The characteristics of the Eastern region are handled in Section 4.1 when addressing the diversity of the country. However, when it comes to the working conditions of the teachers, it may be worth noting that the average number of schools days in the Eastern region is significantly ($p = 0.001$) lower (138 days) than the average in the other districts (152 day) (compare with Figure 4.4.5).

4.4.3.2.2 Instructional activity in the high- and low-performing schools

A set of questions covering the number of the students, teachers, study days, and teachers with high qualifications were set to HTs. On the basis of the responses, there is no difference between the Q1- and Q5 schools when it comes to the qualifications of the teachers, the total number of teachers, and the number of male teachers in the schools (Table 4.4.3).

Table 4.4.3 Instructional activity of the Q1 and Q5 schools

Variables[1]	Lowest quintile (Q1)	Highest quintile (Q5)	t	df	Sig. (2-tailed)	Cohen's d (Effect size)
pq5A_per_PQ5B Students per teacher ratio	31.8	26.8	3.282	338	0.001[2]	0.36
pq6a Number of study days	179.9	187.6	-3.133	342	0.002	-0.34
pq5a Number of the students total	474.6	385.4	3.074	339	0.002	0.33
pq5D_F_per_PQ5D_M Ratio of female teacher and male teachers	0.29	0.42	-2.640	334	0.009	-0.29
pq5d_F Number of Fem. teachers at Grade 8	1.2	1.7	-2.634	356	0.009	-0.28
pq6c Today class 8 present	52.6	42.5	2.229	316	0.027	0.25
pq5b Total number of Teachers	15.6	14.9	0.908	354	0.364	0.10
pq5d_M Number of Male teachers at Grade 8	5.3	5.2	0.849	356	0.396	0.09
pq5c Teachers with qualification above BA	4.0	3.7	0.883	343	0.378	0.10

1) The variables are ordered on the basis of p-value and Cohen's d
2) Significant differences are highlighted

---

[54] Binomial probabilities are $p < 0.001$ and $p = 0.001$ respectively
[55] Binomial probabilities are $p < 0.001$, $p = 0.001$, and $p = 0.010$ respectively

On the basis of the head teachers' reports, the students/teacher ratio is statistically lower in the highest performing community schools (27 students per a teacher) than in the lowest performing school (32). The difference is statistically significant ($p = 0.001$) though the difference is moderate ($d = 0.36$). The number of study days in the highest performing schools is, on average, eight days more than in the lowest quintile schools. The difference is notable though not necessarily remarkable (Cohen's $d = 0.34$ showing a moderate effect size). DTA suggests that when there are 182 school days or less, the probability to find a high-performing school is statistically less (40% of higher-performing schools) than when the number of days is more than 182 days (59% at the higher-performing schools) (see Figure 4.4.5 in Chapter 4.4.2.4). Also, the Q5 schools are smaller (average 385 students) than the Q1 schools (475) (Fig. 4.4.6) and there are more female teachers (on average 2) compared with Q1 schools (on average only 1). The latter is also seen when comparing the ratio of female- and male teachers: on the basis of the reports of the HTs, the female/male teacher ration shows on average of 1.5 times more female teachers in the highest performing schools (ratio is 0.42 female teachers per one male teacher) compared with the lowest performing schools (ratio is 0.29).
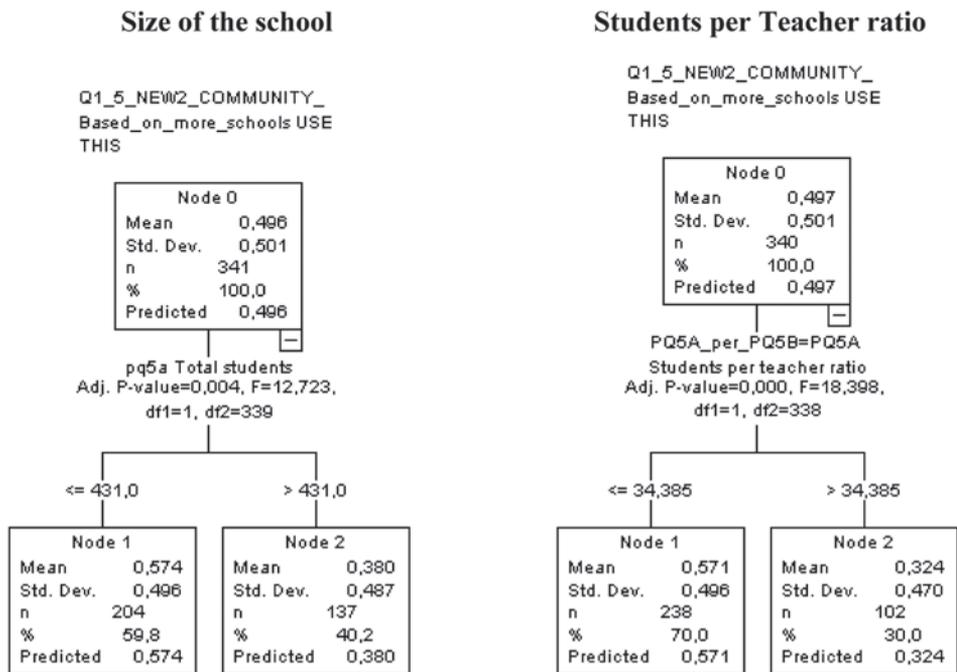
Figure 4.4.6 DTA of "optimal" size of school and the student/teacher ratio

The data gives a clue to an "optimal" size of schools and students per teacher ratio. Though the differences between the proportions of the highest and lowest performing community schools are not very wide, DTA suggests that when the school size is round 430 students or less, the number of the highest achieving schools is statistically speaking the most high (Figure 4.4.6).[56] In the same manner, DTA suggests that, in Nepal, the optimal student/teacher ratio would be 34 students or less per one teacher. In the whole data, including the institutional schools, the cut-offs would be 22 students or less (mean score is 55.3%) compared with the class size of 23–31 students which produces the average results (49.6%) nothing to say the classsize of more than 31students in a class (43.7%).

4.4.3.2.3 School's premises in the high- and low-performing schools

Generally speaking, the physical premises of the schools– Ownership of the premises, Size of the classrooms, or Sufficient desk space – does not explain the differences between the highest and lowest performing schools (Table 4.4.4).There is, however, a small difference between the schools when it comes to the building type: in the Q1 schools, there were more "Kacchi" type of schools than in the Q5 schools ($p = 0.055$). The difference is small ($d = 0.21$).

Table 4.4.4School premises in the Q1- and Q5 schools

Variables[1]	Lowest quintile (Q1)	Highest quintile (Q5)	t	df	Sig. (2-tailed)	Cohen's d
pq7 Own Building	1.0	1.0	-1.935	262	0.054[2]	-0.24
pq10_aL Measurements of a classroom A Length (metres)	7.5	6.9	2.041	335	0.042	0.22
pq8 BuildyingType	1.3	1.2	1.921	347	0.055	0.21
pq9 Sufficient desk, bench	1.2	1.1	1.656	343	ns.	0.18
pq10_aB Measurements of a classroom A Breadth (metres)	5.0	4.8	1.302	335	ns.	0.14
PQ10_ALB_class size AL*AB	39.0	36.1	0.788	335	ns.	0.09

1) The variables are ordered on the basis of p-value and Cohen's d
2) Significant differences are highlighted

The data gives a signal that the physical characteristics of the school do not explain why some schools are better than the others.

---

[1] When explaining the school's average result – not the belonging to the Q1 or Q5 schools – the cut-off would be 386 students (see Chapter 4.3.3.3.2).

4.4.3.2.4 Schools' facilities in the high- and low-performing schools

Two sets of questions concerning the schools' facilities to produce quality teaching and processes were addressed to the head teachers. One set of these questions included availability of study materials (books and teacher's guidelines) and curricula. None of these differed statistically in a significant way between the lowest and highest performing schools; in both groups, 44–50% of the schools got access to all of the material in the subjects in focus.

Another set of questions focused on the observed shortages in relevant facilities to administer quality teaching. These variables are collected into Table 4.4.5. There are no statistically significant differences between the Q1 and Q5 schools when it comes to the number of teachers, school buildings and play grounds, place for teaching, or educational materials; on average, the head teachers evaluate that the situation is quite good (the mean varies 2.7–3.2). Somewhat poorer, though equal between Q1 and Q5 schools, is the situation with the audiovisual material for Social Studies- and Nepali teaching, calculators in the Mathematics studies, budgeting for the basic tools such as pencils and copies, as well as with the electricity system, coldness and hotness in school (the mean ranges 1.8 – 2.3).

Table 4.4.5 Difference in the availability in schools' facilities between Q1 and Q5 schools

Availability of…[1]	Lowest quintile (Q1)	Highest quintile (Q5)	t	df	Sig. (2-tailed)[3]	Cohen's d
pq13h computer software for Mathematics teaching	1.2[2]	1.4	-3.317	294	0.001	-0.39
pq13g computer for Mathematics teaching	1.2	1.5	-2.864	326	0.004	-0.32
pq13n computer for Social Studies	1.3	1.5	-2.690	334	0.007	-0.29
pq13f special material for disabled students	1.2	1.4	-2.340	294	0.020	-0.28
pq13p staff for the computer maintenance and technical assistance	1.2	1.4	-2.372	331	0.018	-0.26
pq13k audiovisual material for Mathematics teaching	1.5	1.7	-2.244	351	0.025	-0.24
pq13j lab materials for Mathematics teaching	1.8	2.0	-2.130	352	0.034	-0.23
pq13l audiovisual material for Nepali teaching	1.6	1.7	-1.746	347	ns.	-0.19
pq13e place for teaching (e.g. classroom)	3.2	3.2	0.936	353	ns.	0.10
pq13b budget for purchasing (e.g. copies, pencils)	2.3	2.2	1.563	346	ns.	0.17
pq13i calculator for Mathematics teaching	2.2	2.3	-1.268	356	ns.	-0.14
pq13d hot, cold, system of electricity	1.9	2.0	-0.870	353	ns.	-0.09
pq13c school building and play ground	2.9	3.0	-0.064	352	ns.	-0.01
pq13m audiovisual material for Social Studies teaching	1.8	1.8	-0.958	345	ns.	-0.10
pq13o Teacher	2.7	2.8	-0.498	349	ns.	-0.05
pq13a educational material (e.g. text books)	3.1	3.1	-0.421	356	ns.	-0.05

1) Variables are profiles on the basis of p-value and Cohen's d.
2) For all variables the scale in anchored to 1= not at all and 4 = enough. Hence, the higher the value the better situation is.
3) Statistically significant differences are highlighted.

The widest difference between the lowest and highest performing schools comes with the computers and related matters. In the highest performing schools – though having severe shortages in those too – there is significantly more computer software for Mathematics teaching ($p = 0.001$, $d = 0.39$), more computers for Mathematics teaching ($p = 0.004$), Social Studies ($p = 0.007$), as well as more staff for the computer maintenance and technical assistance ($p = 0.018$). For the latter three variables, the effect sizes are moderate or low (Cohen's d ranges 0.23–0.32). Significant differences are also seen in audiovisual material for Mathematics teaching ($p = 0.0013$), special materials for disabled students ($p = 0.025$), and in laboratory materials for Mathematics teaching ($p = 0.034$). Note that in all cases, there are more shortages in the lowest performance schools. Though the judgment is based on the subjective feeling of the head teachers, the result is credible because the head teachers would know whether the school has or does not have such facilities as computers; in most cases, the values in the variables are very close to the "not at all" situation.

The dataset gives a strong signal that, in the higher performing schools, there are more computers and computer software in use than in the lowest performance schools. The highest performing schools also have more staff for maintenance and technical assistance, material for disabled students and a wider repertoire of facilities for Mathematics teaching.

4.4.3.2.5 Schools' atmosphere in the high- and low-performing schools

The school atmosphere is entered from two angles: from the students' perspective and from the teachers' perspective. The students' self-reported view is handled in Section 4.5 – here the head teachers' view is covered mainly. Teachers' job contentment is handled in Section 4.3 – here some selected parts are condensed.

Table 4.4.6 condenses the school atmosphere related to the students' behavior reported by the head teacher. Altogether 13 questions were given to the head teacher covering the topics from school leaving to breaking the dress code and from the destructive behavior of a single student to bullying. Two things were asked: how frequently these kinds of actions happen against the school, teacher, students or oneself and how serious the problem is. In Table 4.4.6 these two dimensions are multiplied to produce an index of frequent and serious incidents: the lowest value 1 means that there is no problem and (obviously) it is not a serious problem and the highest value 12 means that this happen every day and so it creates a serious problem in the school.

Table 4.4.6 Difference in the study atmosphere between Q1 and Q5 schools

Variables[1] Problems in…	Lowest quintile (Q1)	Highest quintile (Q5)	t	df	(2-tailed) Sig.	Cohen's d
pq19AXB_a being late in school	5.4[2]	4.0	4.189	317	<0.001[3]	0.48
pq19AXB_c school leaving (number of periods, hour)	4.92	3.5	4.005	302	<0.001	0.46
pq19AXB_f cheating	2.9	2.2	3.793	333	<0.001	0.42
pq19AXB_d breaking dress code	3.9	2.9	3.479	337	0.001	0.38
pq19AXB_h destroying	2.8	2.2	2.902	341	0.004	0.32
pq19AXB_b absence (without reason)	5.7	4.6	2.852	323	0.005	0.32
pq19AXB_j bullying	2.3	1.8	2.64	311	0.009	0.30
pq19AXB_e side talking	3.4	3.0	1.542	342	ns.	0.17
pq19AXB_g vandalism	2.3	2.1	0.961	337	ns.	0.11
pq19AXB_k hurting other students	2.2	2.1	0.923	339	ns.	0.10
pq19AXB_l using vulgar words for teacher	1.9	1.8	1.055	339	ns.	0.12
pq19AXB_m physically hurting teacher and student	1.7	1.5	0.851	340	ns.	0.09
pq19AXB_i stealing	2.1	2.0	0.693	337	ns.	0.08

1) Variables are profiled on the basis of p-value and Cohen's d; the most significant variables are at the top of the list. Original scales were as follows: Frequency: 1 = not at all – 4 = every day; Seriousness: 1 = not serious, 3 = very serious. The variable is composed by multiplying the two dimensions of the original variables.
2) Scale: 1= does not happen – 12 = every day and very serious
3) Statistically significant differences are highlighted.

There are remarkable differences between the highest and lowest performing schools when it comes to the destructive behavior of the students in the schools reported by the head teachers. In the lowest performing schools, the number of school leaving (that is absence) is remarkably higher ($p < 0.001$, $d = 0.46$), the greater frequency of the students being late in school ($p < 0.001$, $d = 0.48$), higher absenteeism rate without a reason ($p = 0.005$), they break the dress code more frequently ($p = 0.001$), they cheat more ($p < 0.001$) and they destroy property more ($p = 0.004$) than the students in the highest performing schools. In is also notable that in the higher performing schools, there is less bullying than in the lower performing schools ($p = 0.009$). The effect sizes are moderate, which indicates that the differences are quite remarkable.

When condensing the variables in Table 4.4.6 into three sums, behavior against the *other students* (items pq19A_e to items pq19A_k), behavior against the *teacher* (items pq19A_l to items pq19A_m), and behavior against *oneself* (items pq19A_a to items pq19A_d), the differences are very wide when it comes to the destructive behavior against oneself (being late and absent, leaving the school

and breaking the dress code).[57] In the lowest performing schools, there are many more negative incidents and their seriousness is greater than in the highest performing schools; the difference is obviously statistically significant[58] and the effect size is moderate or somewhat high ($d = 0.49$) indicating a remarkable difference (Figure 4.4.7). There is also a significant difference in the behavior against the other students[59] though the difference is not very wide ($d = 0.29$). There is no difference in the behavior against the teacher; the incidences are very rare and nor are they very serious cases.
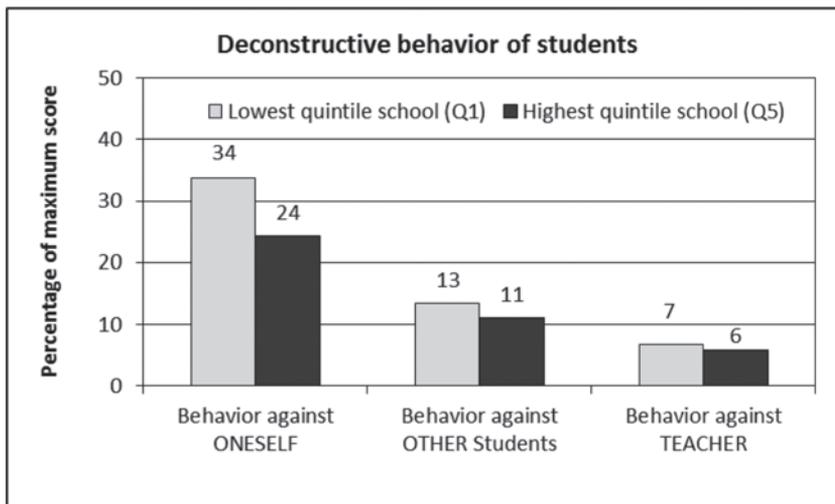
Figure 4.4.7 Differences in students deconstructive behavior between Q1 and Q5 schools

Some of the teacher-related factors in Section 4.3, T20a (Teachers' job satisfaction), T20b (Teachers' understanding of school's curricular goals), T20c (Teachers' degree in success in implementing the school's curriculum), and T20d (Teachers' expectations of student achievement), may also reflect the schools' atmosphere as a place to work and study. The differences between the Q1- and Q5 schools can be seen in Table 4.4.7.

---

[57] It may be arguable why this dimension is called "behavior against oneself". The other two dimensions clearly indicate the behavior is against other students and teachers. For the symmetrical reasons, the first set was named with the same logic. In any case, being absent or late harms just the student him- or herself – not anyone else.

[58] $t(332) = 4.451$, $p < 0.001$

[59] $t(340) = 2.738$, $p = 0.007$

Table 4.4.8 Differences in teacher-related factors between Q1- and Q5 schools

Variables	Lowest quintile (Q1)	Highest quintile (Q5)	t	df	(2-tailed) Sig.	Cohen's d
T20a Teachers' job satisfaction	2.54[1]	2.31	2.387	359	0.017[2]	0.25
T20b Teachers' understanding of school's curricular goals	2.57	2.36	2.628	357	0.009	0.28
T20c Teachers' degree in success in implementing the school's curriculum	2.57	2.35	2.523	355	0.012	0.27
T20d Teachers' expectations of student achievement	2.19	2.05	1.621	358	ns.	0.17

1) Scale: 1= very high to 5 = very low. Hence ,the higher the value the worse.
2) Statistically significant differences are highlighted.

Except in teachers' expectations of student achievement, there are significant differences between the highest and lowest performing schools when it comes to the working and studying atmosphere. In the highest performing schools, teachers' job contentment is higher ($p = 0.032$), they understand the school's curricular goals better ($p = 0.042$), and their degree of success in implementing the school's curriculum is higher than in the lowest performing schools ($p = 0.052$). Though the differences are not wide (d ranges $0.21 - 0.23$) they are systematic: it seems that the working atmosphere in the lowest performing schools is lower than in the highest performing schools.

Data gives a strong signal that in the lowest performing schools, there are many more negative incidents involving students, and their seriousness is greater than in the highest performing schools. This may result in the working atmosphere, as reported by the teachers, being significantly lower in the lowest performing schools than in the highest performing schools.

4.4.3.2.6. Evaluation and monitoring in the high- and low-performing schools

Inferring from the teacher's questionnaires (see detailed in Section 4.4.3.2.8), the teachers in the lowest performing schools use significantly less classroom evaluation ($p = 0.001$) and homework evaluation ($p = 0.002$) per week than in the highest performing schools (Table 4.4.9). It is also notable that the teachers in the lowest performing schools left many more missing values in the variables; this may indicate that the teacher actually does not use these methods.

Table 4.4.9 Difference in the amount of evaluation between the Q1- and Q5 schools

How often you use the following:	Lowest quintile (Q1)	Highest quintile (Q5)	t	df	(2-tailed) Sig.[2]	Cohen's d
T13_1B classwork evaluation (per week)	2.95	3.66	-3.221	339	0.001	-0.34
T13_2B homework evaluation (per week)	3.56	4.21	-3.097	342	0.002	-0.34
T13_3B discussion (per week)[1]	2.54	2.84	-1.357	318	0.176	-0.15
T13_4B project work (per month)[1]	1.64	1.66	-0.094	225	0.925	-0.01
T13_5B observation of behavior (per month)[1]	3.83	3.74	0.128	300	0.898	0.01

1) There are many missing values in the variable, especially in the lowest performing schools. If all the missing values were replaced by 0 (indicating that no active evaluation of the kind was observed), the result would remarkably favour the highest performing schools; the situation would look even more deviating between the schools.
2) Statistically significant differences are highlighted.

The data gives a signal that in the lowest performing schools the number of classwork and homework evaluation sessions is lower than in the highest performing schools.

4.4.3.2.7 Teachers' professional development in the high- and low-performing schools

Inferred from the teachers' background questionnaire (see detailed in Section 4.3.3.2.7), the teachers in the highest performing schools participated in several kinds of professional development processes much more than the teachers in the lowest performing schools (Figure 4.4.8); double amount of teachers in Q5 schools (33%) attended two or more out of six given possibilities for professional development compared with the Q1 teachers (16%). Further, three times more teachers in Q5 schools (21%) attended three or more types of education courses compared with the Q1 teachers (7%). Differences are statistically significant ($p = 0.007$ and $p = 0.010$ respectively).The result seems to refer to the better possibilities or stronger motivation of the teachers to attend continuous education in Q5 schools.
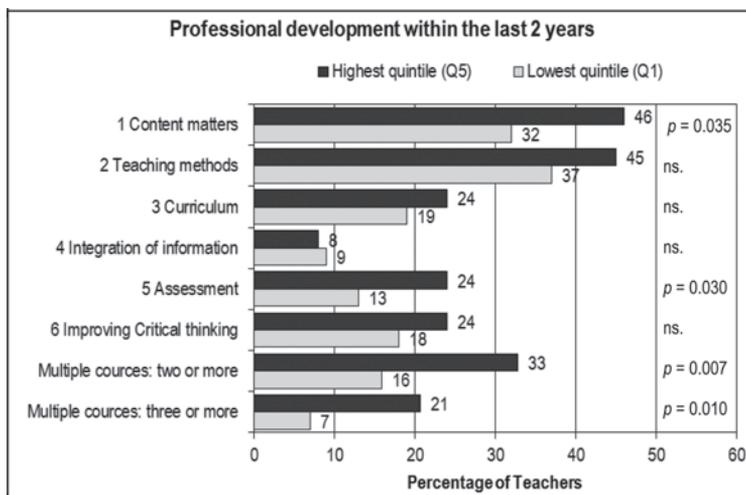
Figure 4.4.8 Differences in the professional development between Q1 and Q5 schools

4.4.3.2.8 Cooperation with the community, school personnel, and SMC

Combining the questions from the head teachers'- and teachers' questionnaires it seems evident that the highest performing schools are statistically more active in the community ($p < 0.001$), with the teaching staff ($p < 0.001$) as well as with School Management Committees ($p = 0.006$) (Table 4.4.10). The differences are radical when it comes to those teachers who reported more than 7 pedagogical interactions per year with the teaching staff; almost double the amount of those teachers who came from the highest performing schools (55%) when compared with the lowest performing schools (31%). The same happens with the SMC; from those teachers who reported that they have not met the school management committees, twice the number came from the Q1 schools (22%) than from the Q5 schools (11%). Also, when it comes to the interaction with the community, twice the amount of head teachers in the highest performing schools reported that the number of interactions is good or very good (28%) compared with the head teachers in the lowest performing schools (14%). This indicates that in the high-performing schools both head teachers, teachers and SMCs are more active in developing the school than in the low-performing schools.

Table 4.4.10 Difference in the interaction in the Q1- and Q5 schools

Variables[1]	Lowest quintile (Q1)	Highest quintile (Q5)	t	df	(2-tailed) Sig.	Cohen's d
T16_dummy Interaction with Teacher staff for improving teaching-learning; how many times during the year (=<7 = 0, >7 =1) (Percentage)	0.31	0.55	-4.723	360	<0.001	-0.51
pq21 co-operation of community in school (1 = minimal – 5 = very good)	2.69	3.04	-4.024	351	<0.001	-0.44
T16 Interaction with Teacher staff (TS) of improving teaching-learning (how many times during the year)	6.04	7.59	-3.508	339	0.001	-0.37
pq22 co-operation between school personnel (1 = minimal – 5 = very good)	3.85	4.06	-2.802	352	0.005	-0.31
T15 Interaction with the School Management Committee (SMC) of improving teaching-learning (1 = Yes, 2 = No)	1.22	1.11	2.742	338	0.006	0.30

1) The variables are ordered on the basis of p-value and Cohen's d

The data gives a strong signal that the high-performing community schools are more active in teacher staff interactions and SMCs to improve the teaching learning processes as well as in co-operation with the local community compared with the low-performing schools.

4.4.3.3 Summary of the teachers characteristics –modeling the phenomenon

Above, several individual school-related factors have been detected which *individually* explain the difference between the lowest and highest performing community schools. These factors are collected in Table 4.4.11. Some of them may be strongly related with each other and hence they may not add value in explaining why some community schools are performing much better than others. It is also worth remembering that the school-related factors are not the only relevant variables in explaining the phenomenon; related diversity factors are handled in Section 4.1, teacher factors are handled in Section 4.3 and Student factors in Section 4.5.

4.4.3.3.1 Modeling the highest and lowest quintile community schools with the school-related variables

Logistic regression analysis was used to analyse which of the factors are independent predictors for belonging to either the lowest or highest performing schools. When taking several variables in the analysis into account at the same time, all the cases with even one missing value are omitted during the analysis by the software. Hence, such variables with a remarkable number of missing values were not included in the analysis. In Table 4.4.11, only one variable was detected with a remarkably low number of cases; this variable was not taken into the analysis.

Table 4.4.11 School-related factors differing significantly between the Q1- and Q5 schools

Developmental region Dummy Eastern = 1/ other = 0
Developmental region Dummy Valley = 1/ other = 0
Ecological zone Dummy Mountain = 1/ other = 0
pq5a Number of the students total
PQ5A_per_PQ5B Students per teacher ratio
pq5d_F Number of Female teachers at Grade 8
PQ5D_F_per_PQ5D_M ratio of female teacher and male teachers
pq6a number of study days
pq8 Type of school building
pq13f special material for disabled students
pq13g computer for Mathematics teaching
pq13h computer software for Mathematics teaching
pq13j lab materials for Mathematics teaching
pq13k audiovisual material for Mathematics teaching
pq13n computer for Social Studies
pq13p Staff for the computer maintenance and technical assistance
pq19AxB_a being late in school
pq19AxB_b absence (without reason)
pq19AxB_c school leaving (number of periods, hour)
pq19AxB_d breaking dress code
pq19AxB_f cheating
pq19AxB_h destroying
pq19AxB_j bulling
pq21 co-operation of community in school(1 = minimal – 5 = very good)
pq22 co-operation between school personnel (1 = minimal – 5 = very good)
T13_1B classwork evaluation (per week)
T13_2B homework evaluation (per week)
T15 Interaction with the School Management Committee (SMC) of improving teaching-learning (1 = Yes, 2 = No)
T16_dummy Interaction with Teacher staff of improving teaching-learning; how many times per year (0-7 = 0, >7 =1)
T16 Interaction with Teacher staff (TS) of improving teaching-learning(how many times during the year)
1) T18a Dummy Professional development last 2 years: combined (1-6 = 0, >6 = 1)

1) the variable is not in the model because of the high number of missing values

343

Table 4.4.12 shows the statistically best model (Conditional-selection) after some transformations. It is notable that actually only a couple of variables really have their own (main) effect on the whole model. These factors are discussed below. During the analysis, the statistically significant predictors were first detected. After this, DTA was used individually with each variable and the best cut-offs of the variables were detected. These cut-offs were used to dichotomize the predictors; the original variable was transformed as a dummy variable to simplify the analysis and the interpretation.

Table 4.4.12 Statistically the best model of Logistic regression analysis explaining belonging to the Q1- and Q5 schools

Variables[1]	B	S.E.	Wald	df	Sig.	Exp(B)[3]
Developmental region dummy Eastern = 1/ other = 0[2]	-2.277	0.467	23.737	1	<0.001	0.103
Developmental region dummy Valley = 1/ other = 0	1.833	0.823	4.957	1	0.026	6.25
PQ21_dummy co-operation with community (1 = minimal – 5 = very good) (4-5 = 1, 1-3 = 0)	1.304	0.412	10.047	1	0.002	3.686
PQ5A_dummy Total number of Students (431 or less = 1, more than 431 = 0)	0.851	0.305	7.776	1	0.005	2.343
T13_1B_dummy How often classwork evaluation (per week) (0-2 = 0, 3 or more = 1)	0.849	0.306	7.681	1	0.006	2.338
TQ16_dummy Interaction with Teacher staff of improving teaching-learning; how many times during the year (0-7 = 0, 8 or more =1)	0.794	0.3	7.02	1	0.008	2.213
PQ19AxB_A_dummy Frequency and seriousness of being late in school(range 1 – 15) (1 – 3 = 1, 4 or more =0)	0.785	0.387	4.121	1	0.042	2.192
PQ19AxB_H_dummy Frequency and seriousness of destroying the schools' premises (range 1 – 15) (1 – 3 = 1, 4 or more =0)	0.772	0.311	6.18	1	0.013	2.165
PQ6A_dummy Number of Study days (182 or less = 0, 183 or more = 1)	0.023	0.008	9.435	1	0.002	1.024
Constant	-4.984	1.5	11.034	1	0.001	0.007

1) Method = Forward: Conditional
2) The cut-offs for the dummy variables are made on the basis of suggestions by DTA.
3) The variables are ordered on the basis of the risk value; the variable indicating the highest "risk" of being within the highest performance schools is at the top.

All the variables in the combined model (Table 4.4.12) are statistically significant predictors for a school to belong to either Q1- or Q5 schools. The last column on the table (Exp(B)) tells the "risk" of being in the group of the highest performing schools when the value of the dummy variable is higher, that is 1. Out of eight variables, the most effective predictors are the location of the school: It is very "risky" for being within the lowest performance schools when the school comes from the Eastern Developmental region (9.7 times risk = 1/0.103) and, parallel, the "risk" is very high to be within the highest performing school if the school was located in Valley (6.2 times risk) in comparison with the other districts.

The remaining variables are dichotomized so that the higher value favors the high-performing schools. Hence, when the head teacher evaluates that the *co-operation with the community* is at a good or very good level, the "risk" of being one of the highest performing school is 3.7 times higher than belonging to the group of the lowest–performing schools. Slightly lower is the risk when there are *fewer than 431 students in the school* (2.3 times risk). More than double is the "risk" when there are at least 3 *classroom evaluations* a week (2.3 times risk), there are more than 7 times per year *interaction between the teaching staff* for improving the teaching learning processes (2.2 times risk), when *being late from the school is not frequent* or a serious problem (2.2 times "risk"), and when the *destruction of a property is not frequent* or a serious problem (2.2 times risk). Altogether, that is, when taking the variables into account at the same time, they explain 45% of the phenomenon (Nagelkerke $R^2 = 0.450$). This means, in practice, that 74% of the schools can be classified correctly by using the model although it is somewhat better detecting the lowest performing schools (76% correct).

4.4.3.3.2 Modeling the overall achievement on the community schools with school-related variables

The methodological decision made (to concentrate on profiling only the most extreme schools), lead to a situation that actually most of the schools were not in the analysis; usually this is reasonable because nothing can explain the difference between similar schools with the average student achievement. The decision and analysis done above, however, was carried out just to *clarify* the phenomenon; it is expected that the predicting variables found in the analysis above can explain, to some extent, the general differences of the average achievement level in the schools. The traditional linear regression analysis with Stepwise regression was used to model the phenomenon; the equated (and shifted) mean score of the school is explained by the same variables as above (see Table 4.4.11). After the variables were found, they were dichotomized in order to simplify the interpretation by using DTA. Table 4.3.13 shows the results.

Table 4.4.13 Statistically the best model of Linear regression analysis explaining the average of student achievement in the community schools

Model[1]	Unstdzed Coefficients		Stdzed Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	41.33	1.034		39.952	<0.001
DevelopmentalRegiondummyEastern=1	-7.58	1.144	-0.227	-6.625	<0.001
DevelopmentalRegiondummyValley = 1	8.94	2.011	0.151	4.445	<0.001
T16_dummy Interaction with Teacher staff of improving teaching-learning; how many times during the year (7 or less = 0, 8 or more = 1)	3.257	0.931	0.12	3.499	<0.001
PQ5A_dummy Total number of Students (386 or less = 1, more than 386 = 0)	3.131	0.912	0.116	3.434	0.001
pq6a_dummy Number of study days (<182 = 0, 183 or more = 1)	2.267	0.918	0.084	2.47	0.014
PQ19AxB_H_dummy Frequency x Seriousness of destroying school premises (1-3 = 1, >3 = 0)	2.266	0.934	0.083	2.426	0.015
Dependent Variable: P_Eqd_Total_mean_SHIFTED to the mean of 46.91 for all subjects $R = 0.38$ $R^2 = 0.14$ $R^2_{adj} = 0.13$					

1) (Method = Stepwise)

The model in Table 4.4.13 can be interpreted as follows: The average mean of the community schools is 41.3% of the maximum score (the value of the coefficient B of 'Constant').[60] The coefficient B tells how much the average score would change when one unit change is seen in the predictor variable; all the predictor variables are statistically significant (at the same time). Now, all the predictors are dummy variables and hence the coefficient strictly indicates how much the school would have gained by having the higher category of the dummy variable, that is, when having the value of 1. For example, if the school came from the Eastern region, the score was, on average, 7.6 percent points *lower* than in the case that the school came from any other Developmental region. Similarly, if the school came from the Valley, it would have advanced 8.9 percent points. The differences between the Eastern district and the other districts are further analysed in Section 4.1 with the diversity analysis.

---

[60] It deviates from 46.9 because of missing values in the variables – not all schools are in the analysis.

The school cannot do much for its location; a school from the Eastern region cannot make a decision of being a school from the Valley. However, the other seven variables are more or less of a kind that could be used as a basis for developing the school. When using the variables in Table 4.4.13 as a checklist, one gets an idea of how much more a school would gain when the higher category of the dummy variable was observed. **If the school met all the circumstances indicated by the higher category**, that is, it had *more than 8 times a year interaction with the teaching staff* to improve the teaching learning processes, *less than 386 students* in the school, *more than 183 study days* per year, and none or only a *few serious incidents of destroying school's premises*, **the school would have gained 11 percent points more than those schools observing the opposite**. As is the nature of the regression modeling, the model cannot explain all the variance in the schools' population – but 13% of the variation can be explained. Hence, there are certainly some other strengths of the school than in the list in Table 4.4.13 which may compensate for the disadvantages – or there may be some weaknesses which may lower the level even though all the items on the checklist are met positively.

It is worth noting that the result does *not* mean that if the school just decided to raise the number of study days from 182 to 183 and the number of interactions between teaching staff from 7 per year to 8 per year, the results would change radically. Behind these variables, there lies a set of other variables and practices which are condensed with the variables found with the model above. The complexity of the phenomenon is illustrated in Section 4.3 in Table 4.3.12 where the teacher-related variables showing the differences between the "low-activity schools" and "high-activity schools" are seen in a rough way. On the basis on that table, the situation looks logical: Compared with the low-activity schools (which explains the lower student achievement), in the high-activity schools the teachers are *younger*, they *evaluate their students more frequently*, they are *more active with the SMCs*, they are *more content with their work* (which may be explained by the good results, perhaps), the other *teachers seem to understand the educational goals* better and they *expect students to succeed*, the *parents are more actively* involved in the school, and the *students are working constructively to reach their goals*, the students are lead to *more constructive learning experiences* (by giving them more possibilities to work with their own problems and with small groups) as well as lead them to *more complex thinking processes* (by letting them explain the answers and decide their own procedures for solving complex problems) and *more practical solutions* (by giving them possibilities to relate what they have learned in their daily life).

4.4.4 Conclusions

Several school-related variables were detected which explain, at least to some extent, the differences between the low and high performing community schools. Some of the variables showed a high percentage of missing values; these were omitted in the analysis. In summary, the main results related to the school-related factors explaining the student achievement are as follows:

- The school effect in Nepal is 68%; this is a very high value. This means that the results within the schools are very much alike and there are wide differences between the schools. The effect size is much less in Mustang (9%) and in Kalikot (30%) where the test was administered so that the possibilities for the "social work" were minimized. (Section 4.4.3.1).

- The student performance in the Eastern region is much lower than in other regions; most of the lowest performing districts come from the Eastern region, where the number of high-performing schools is radically lower than in the other regions. The average number of schools days in the Eastern region is also significantly lower (138 days) than the average in the other districts (152 day). (Section 4.4.3.2.1)

- DTA suggests that, in Nepal, the optimal student/teacher ratio would be less than 34 students per one teacher. However, the highest achievement level is found in the schools with the student/teacher ratio of 34 or lower. (Sections 4.4.3.2.2)

- The physical characteristics of the school do not explain why some schools are better than the others. (Section 4.4.3.2.3)

- In the higher performing schools, there are more computers and computer software in use than in the lowest performance schools. Those schools also have more material for disabled students and a wider range of facilities for Mathematics teaching. (Sections 4.4.3.2.4 and 4.4.3.3.2)

- In the lowest performing schools, there are many more negative incidents involving students, and their seriousness is greater than in the highest performing schools. Also the working atmosphere, as reported by the teachers, is significantly lower in the lowest performing schools than in the highest performing schools. (Section 4.4.3.2.5)

- In the lowest performing schools, the number of classwork and homework evaluation sessions is lower than in the highest performing schools. (Sections 4.4.3.2.6)

- The teachers in the highest performing schools participated in several kinds of professional development processes more than the teachers in the lowest performing schools. This may refer to the better possibilities or stronger motivation of the teachers to attend continuous education opportunities in the highest performing schools.(Sections 4.4.3.2)

- The high-performing community schools are more active in improving the teaching learning processes by more intense teacher staff- and SMCs interactions as well as in co-operation with the local community, compared with the low-performing schools. (Sections 4.3.3.2.8)

An additional, not school-related a factor, is that the most effective predictor of being a low-performing school is that it originates from the Eastern Developmental region. This is discussed in detail in Section 4.1.

It is good to keep in mind that school factors are just one part of a wide variety of factors explaining the student achievement. Another set of relevant factors related to family factors and the socioeconomic status (the SES factor), have been widely discussed in the literature (starting mainly from Coleman *et al*. 1966 on; see Jencks *et al*., 1972; Reynolds *et al*., 2000). Whatever may be the effect of these familial factors in learning achievement, the extent of influence of social policies is very much limited. It is because social policies try to address the gaps that are malleable and it is easier to deal with such factors that are directly associated with the schools. Hence, the 'school factor' is one of the most important factors for drawing the attention of the educationists, policy makers and social scientists alike. The extent of the gap that is attributed to school factors varies across the studies. However, as reviewed by Scheerens, (2000) the gap is higher in the developing countries. Also, the input differentials in these schools explain a little of these gaps in the developed countries, whereas the effect of the input differentials is significant in the developing countries. Moreover, the interplay between the inputs and the processes may have a greater influence in the achievement of the students.

References for Section 4.4

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

BPEP (1998). The Effect of New Curriculum on the Achievement of Grade 5 Student.

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

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

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

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., Weinfeld, F. D., & York R. L. (1966). *Equality of Educational Opportunity* . Washington, DC: US Department of Health, Education & Welfare. Office of Education. U.S. Government Printing Office.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership,* **37**(10), 15–24.

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

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

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

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

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

Fullbright (2008). *A final report on national assessment of grade-V students*, Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education. Fullbright Consultancy & Cooperative Hands in restoration (CHIRAG).

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika,* **73**, 43–56.

Hautamäki, J. &Kuusela, J. (2005). Measuring the children, infering the school. In **H.K. Lyytinen & A. Räisänen** (eds), *Kehittämissuuntaa arvioinnista.* [*Directions to develop in Evaluation*] Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä: Jyväskylänyliopistopaino. 229–242. [In Finnish]

Jencks, C., Smith, M., Ackland, H., Bane, M.J., Cohen, D., Ginter, H., Heyns, B. & Michelson, S. (1972). *Inequity: A reassessment of the effect of the family and schooling in America.* New York: Basic Books

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics,* **29**(2), 119–127.

Liu S (2006). *School effectiveness research in China.* An Unpublished Doctoral dissertation at Louisiana State University. Retrieved from http://etd.lsu.edu/docs/available/etd-10222006-172424/unrestricted/Liu_dis.pdf.

Metsämuuronen, J. (2009). *Methods assisting the assessment.* Oppimistulosten arviointi 1/2009. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

Metsämuuronen, J. (2010). Changes in Mathematical achievement and attitudes during the first years in compulsory education. In **E. K. Niemi & J. Metsämuuronen** (Eds.), Learning Outcomes and Changes in Achievement in Mathematics. Koulutuksen seurantaraportit 2010:2. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

Metsämuuronen, J. (2013a). *Handbook of Basics of Researcher Methods in Human Sciences*. 5th Edition. Oy International Methelp Ltd. E-book. [Beta version available at www.Booky.fi]

Metsämuuronen, J. (2013b). *Changes in student achievement from grade 3 to grade 9.* Koulutuksen seurantaraportit 2013:4. Opetushallitus. Helsinki: Edita Prima. [In Finnish].

Murphy, J. (2010). *The educator's handbook for understanding and closing achievement gaps.* . Thousand Oaks, CA: Corwin Press. Retrieved from http://knowledge.sagepub.com/view/the-educators-handbook-for-understanding-and-closing-achievement-gaps/SAGE.xml

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.) **(2008).** *TIMSS 2007 Technical Report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://pirls.bc.edu/timss2007/mathreport.html.

PEDP (1998). Primary Education Development Project: An assessment. 1998. Kathmandu.

PISA (2001). *Knowledge and Skills for Life.* First results from PISA 2000. Paris: OECD.

PISA (2003). *Education at a Glance.* OECD Indicators 2003. Paris: OECD.

PISA (2005)._Equity in Education._Thematic review.Finland, Country note. Retrieved from http://www.oecd.org/dataoecd/49/40/36376641.pdf

PISA (2007).PISA 2006 results. Retrieved from http://www.pisa.oecd.org/document/ 2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html #ES

PISA (2010a).PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Vol. I). Paris: OECD.

PISA (2010b) Finland: Slow and Steady Reform for Consistently High Results. Retrieved from www.oecd.org/dataoecd/34/44/46581035.pdf

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., and Townsend, T. (2000). An introduction to school effectiveness research. In **C. Teddlie & D. Reynolds** (eds.), _The international handbook of school effectiveness research_ (pp. 3−25). London: Falmer.

Scheerens, J. (2000). _Improving school effectiveness_. In International Institute for Educational Planning, _Fundamentals of Educational Planning, 68_. Paris: IIEP.

Teddlie, C., Reynolds, D. & Sharon, P. (2000). Current topics and approaches in School Effectiveness Research: The contemporary field. In **C. Teddlie & D. Reynolds** (eds.), _The International Handbook of School Effectiveness Research_ (pp. 26−51). London: Falmer.

Thapaliya, T. P. (2011). _Leadership Process and School effectiveness_. An unpublished Ph. D. Dissertation submitted to the School of Education, Kathmandu University, Nepal.

4.5 Student Effect in Learning

Dr. Jari Metsämuuronen[1]

1. Helsinki University, Faculty of Behavioral Sciences

Abstract

This article deepens the knowledge of NASA 2011 by combining all the datasets into one and by explaining the differences in student achievement with student-related variables. The results are based on 50,000 students' background questionnaires. The highest performing community schools are compared with the lowest performing schools. The researcher question is: what do the students in the highest performing school do differently to gain the high results?

Because the school effect in Nepal is 68% the remaining 32% is assigned to the students. The student effect is much higher in Mustang (91%) and in Kalikot (70%) where the test was administered so that the possibilities for the "social work" were minimized. It is more probable that the highest performing students will be aged 14 or lower rather than 15 or higher. There is no difference in the achievement level between boys and girls in the lowest performing students. However, the boys seem to be better than girls within the highest performing students.

The higher-performing students are more content with subject, teachers, and school and they sense more utility in the subject than the lowest performing students though the differences between the groups are not very wide; both groups feel, for example, the subject to be equally useful for them. A reasonable amount of time is spent on homework given by the teacher, being present in school, and time *not* spent in paid work raises the achievement level remarkably.

There is a remarkable inequity between the language groups when it comes to the highest and lowest performing student groups: Rai, Limbu and Magar students are over-represented in the lowest performing student group and Urdu, Tamang, Newar, and Sherpa students in the highest performing student group. There is also a difference between the castes: Dalits, Alpasankhyaks, and Madhesis are over-represented in the lowest performing student group and Brahman and Cheetri students are over-represented in the highest performing group.

The socioeconomic status (SES) plays a strong role in the educational processes in Nepal – not only in the private schools but also in the community schools. Within the community schools, the difference in achievement between the students from the lowest and highest SES groups is remarkable (22–25 percent points). Especially challenging is the situation in the families where both parents are illiterate or they both work in the agricultural field. When the children have only some home possessions (like a table to work on, a dictionary and so on), the achievement level is statistically lower than if there are more than four out of 12 indicators met. It is the same with the home accessories (mobile phones, bathrooms and so on): with no accessory indicator

met, the results are very low and when there are three or more met out of five, the results are remarkably higher.

Keywords: Student effect, Inequity, Students' own activity, SES

4.5.1 Introduction

When knowing, on the basis of Sections 4.3 (Teacher effect in achievement) and 4.4 (School effect in achievement), that, in Nepal, the teachers' and schools' processes, actions, and characteristics explain 68% of the student performance, the rest can be allocated to the students and their families and some other related factors such as the peer group. On the basis of the rough conceptual map of achievement and related factors in Section 2.7 (Metsämuuronen, 2009), the student effect can be divided into three components: 1) strictly individual characteristics of the student including such elements as sex, age, working habits, time spent on studies, the attitude toward the subject and school, motivation, academic ability, or intellectual capacity, 2) factors related to family background such as the socioeconomic status (SES), language used at home, caste, parents' profession, education and literacy levels, given help at home for homework, or home possessions for assisting the studies (such as a table or a peaceful place for studying, a computer, literature or dictionaries), and 3) factors related to their peer group, such as interpersonal relationship or bullying (e.g. Monks, Robinson & Worlidge, 2012). More factors can be found for each component – especially the theories of personal psychology may add such concepts as coping[61], commitment[62], self-efficacy[63], Self-Efficacy[64], and learning strategies[65]. Because of the wide range of different possible viewpoints, a radical selection of the factors were done when preparing the student background questionnaire (see Section 4.5.2.2)**.**

This article shows how strong the student effect is on the learning achievement in Nepal when using the same methodology as is used in international studies. It tries to identify the effective student activities and processes by comparing the processes in the highest performing students with the lowest performing students. As an outcome, the final aim is to identify the characteristics of an effective student in Nepal. Research questions are set as follows*:*

---

[61] e.g. Metsämuuronen, 1995; Holen *et al*., 2012; MacCan *et al*., 2012; Yeung, Lau & Nie, 2011; Shin & Ryan, 2012.
[62] e.g. Shamir 1986; Metsämuuronen, 1995.
[63] e.g., Kiran & Sungur, 2012; Gafoor & Ashraf, 2012; Phan, 2012.
[64] e.g. Roebers et al., 2012; Park, 2011; Phan, 2012.
[65] e.g. What Works, 2012a; 2012b; Tavakolizadeh & Oavam, 2011; Postholm, 2010; Vandevelde, Van Keer & De Wever, 2011; Lennon, 2012.

(i)   *How strong is the effect of the student in the learning achievement in Nepal?*

(ii) *What kinds of characteristics, activities and processes can be found within the highest performing students compared with the lowest performing students?*

(iii) *What are the characteristics of the effective student?*

4.5.2 Methodological solutions

The analysis is based on the combined data comprising round 50,000 students covering the subjects of Mathematics, Nepali, and Social Studies. The sampling, achievement tests, background questionnaires, reliabilities and validities of the tests, as well as the other technical matters are discussed mainly in Section 2; they are not repeated here. However, the research design, conceptual framework, variables used in analysis, and the statistical method are discussed here in more detail.

4.5.2.1 Design

It is quite a challenging task to separate the student effect from the school effect or family effect, especially when children with varying backgrounds attend schools with visibly distinct characteristics. As seen in the previous Sections, there are a multitude of factors explaining the achievement on different subjects (Sections 3.1 to 3.3) as well as differences between the teachers and schools (Sections 4.3. and 4.4). In Nepal, especially in the urban areas, students are segregated by the socioeconomic status of their parents: the parents who can afford to, seem to choose the private schools leaving others to free or highly subsidized community schools (see Sections 4.3.2.1 and 4.4.2.1). It appears that private schools do better in terms of measured student achievement; this can be explained by the selection of the students as discussed in the Sections mentioned.

A basic reality of the mean achievement level of the schools in Nepal is very interesting from the comparison viewpoint: the schools can be divided strictly into the lower- and higher socioeconomic status (SES) schools and lower-performing schools and higher-performing schools (See Sections 4.3.2.1 and 4.4.2.1). The average achievement of the students in the lowest performing schools was below 15% while the students in some of the highest performing schools gained, on average, over 90%. When knowing that the average score in a school is below 15%, it means that, in those schools, many students fall below that boundary. The differences are waste and the result means that the low- and high-performing students are concentrated in certain schools. From Sections 4.3 and 4.4 it is known

that most of the institutional schools are performing very well but the community schools vary from the very high-performing schools to the very low-performing schools.

When it comes to the student data, the picture is not that clear. It is evident that there are low- and high-performing students. Some students in the survey (n = 10) did not give a correct answer to a single test items (0% of the maximum marks correct) while the best students (n = 46) gained more than 95%. It seems also evident that the students in the institutional schools perform better than the students in the community schools (Figure 4.5.1) and most probably the difference is due to the selection process rather than efficiency in these schools: the poorest students are not applying/selected into the private schools.
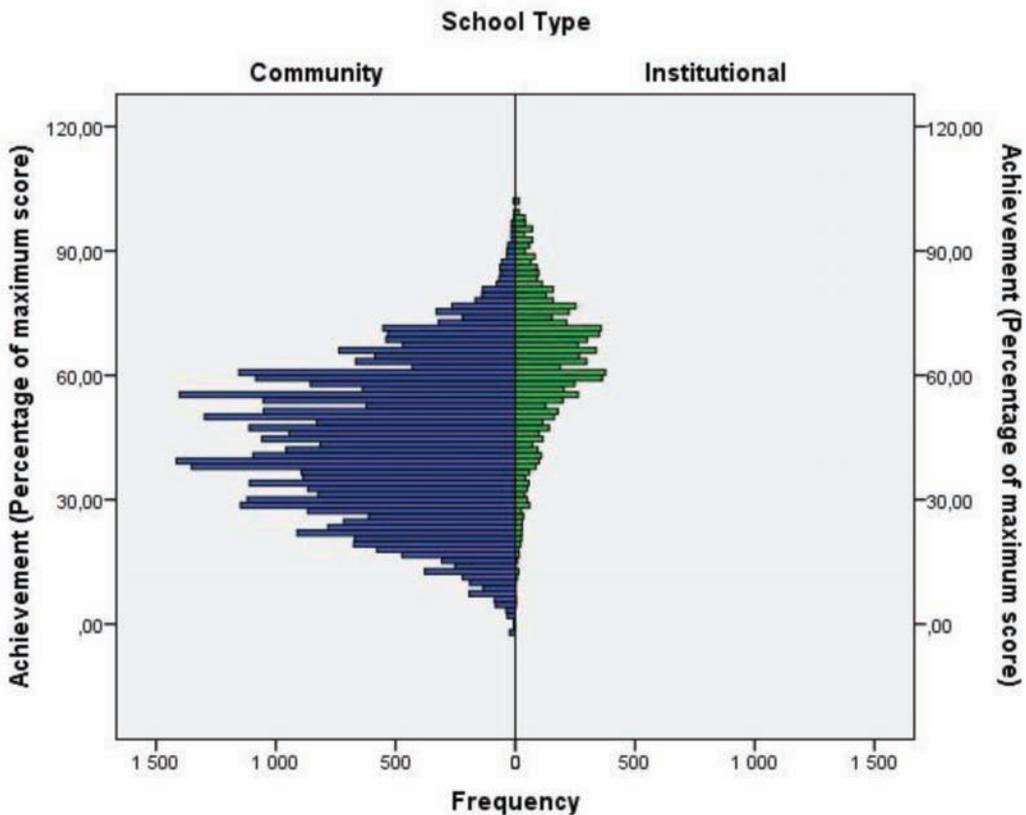
Figure 4.5.1 Comparison of student achievement in the Community- and Institutional schools

356

Most students, 83%, come from the community schools; community schools therefore are responsible for the main educational thrust in Nepal. Methodologically thinking, because of the vast difference between the populations, there is no point in comparing the students from the community schools with the students from the institutional schools even though there are even more highly-performing students in the community schools than in the institutional schools. A more interesting comparison can be made when comparing the high and low-performing students in the community schools. Hence, in what follows, only the community schools are handled. To clarify the design for the analysis, only the highest and lowest deciles of the community schools' students are used in comparison (n = 4097 and n = 3977 respectively). After dividing the students in the community schools into ten equal-sized groups on the basis of their achievement, the lowest decile comprises students with 21% of the maximum score or less and the highest decile comprises students with 67% or more of the maximum score (Figure 4.4.2).
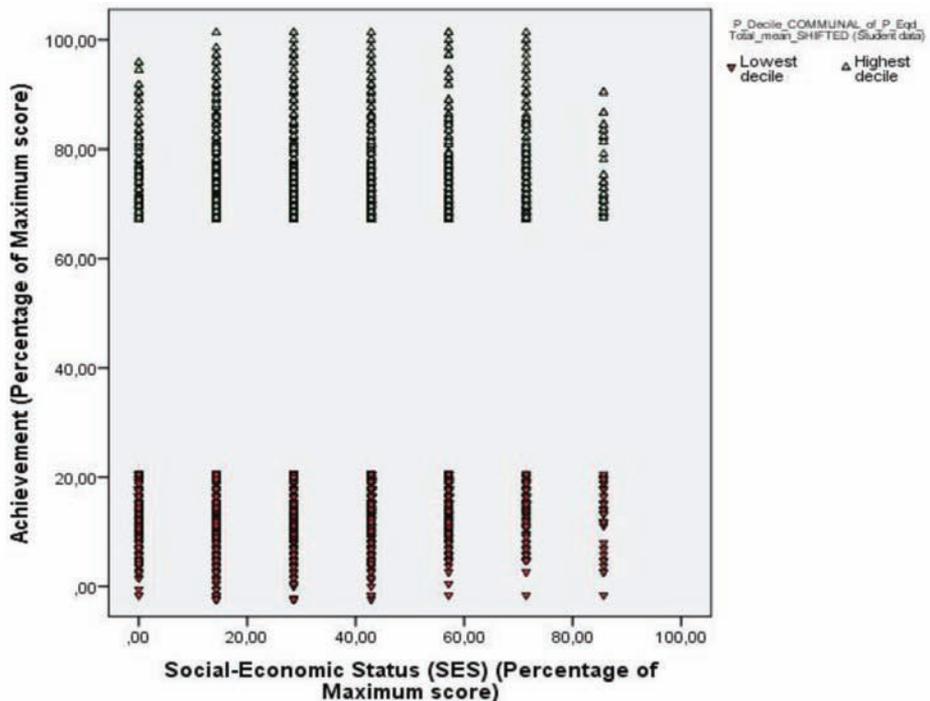
Figure 4.5.2 Connection of SES and student Achievement – the lowest and highest deciles

In Figure 4.5.2 one may notice that contrary to the school-wise data where the mean SES of schools differed radically between the community and institutional schools, the SES of individual students deviates quite widely within the community schools. Except for the highest level of SES (100%), all the SES categories are found in both the highest and lowest performing student group. Technically then,

357

the SES will be taken as a covariate in the statistical models (see Section 4.5.2.4). From the methodological point of view, the design is as visualized in Figure 4.4.2 (compare with the designs in Sections 4.3.2.1 and 4.4.2.1).
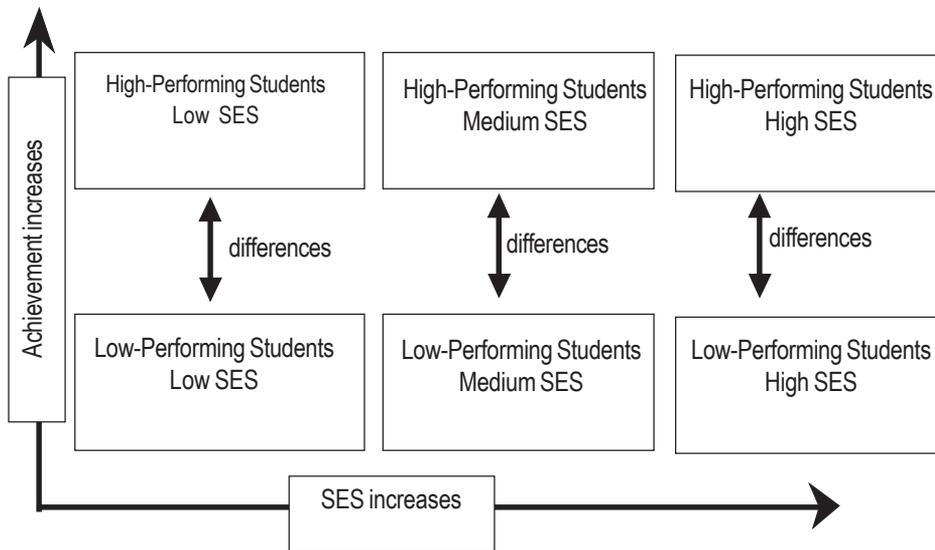
Figure 4.5.3 Design for analyzing the differences in student achievement

Then an adequate question is what kinds of differences are present in the community schools of the highest quintile compared with the community schools in the lowest quintile when taking into account the differences in the SES of the family? It seems evident that the students in the highest quintile are effective ones: with the same (parents' economic and educational) input they are able to produce the same results as the students in the private schools who have much more input. This article tries to find the answer to the question above by concentrating on students' characteristics; whereas Section 4.3 focused on the teacher-related factors and Section 4.4 on school-related factors. The variables found by focusing only on the extreme students are used, at the final phase, in explaining the differences within all the students**.**

4.5.2.2 Conceptual framework of the student factors in NASA 2011

In NASA 2011, there are a number of questions for the students that can be used to analyse the effect of individual characteristics of the students and family background in achievement. Figure 4.5.4 shows the conceptual framework of the Student background questionnaire.
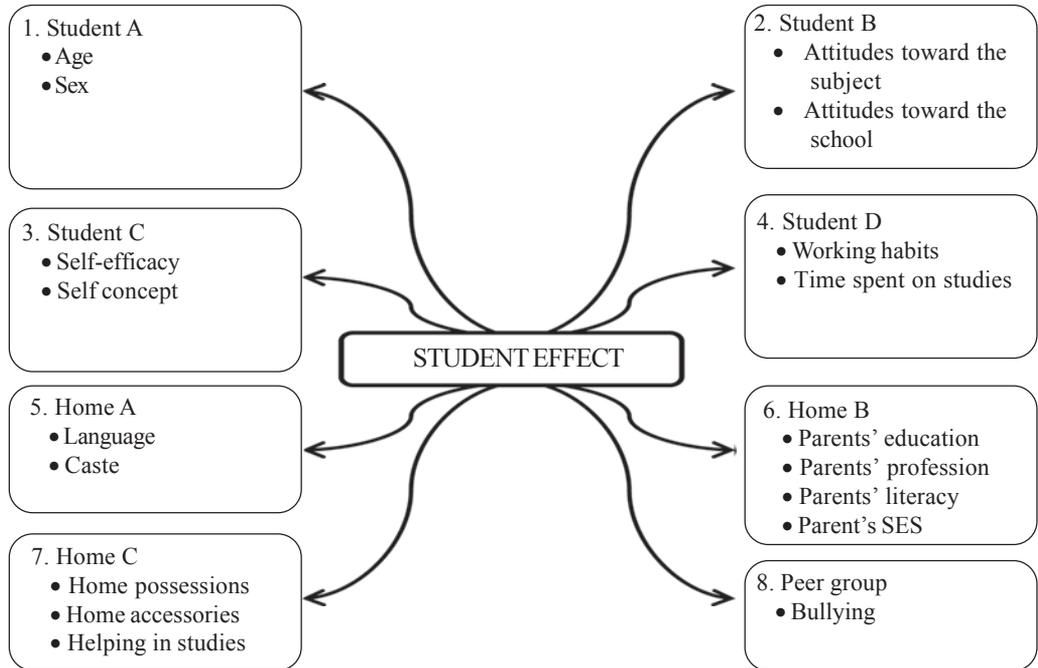
Figure 4.5.4 Conceptual framework of the student factors

The framework is not discussed here. Nevertheless, it can be criticized and debated. However, a civilized guess is that all the factors in Figure 4.5.4 may be valuable when explaining the differences in students' learning outcomes.

4.4.2.3 Variables used in analysis

Mainly the individual variables are explained within the text. However, some main variables are addressed here. In what follows, the student effect is estimated in terms of variation of the students' achievement in the whole sample. Students' achievement is estimated as the latent ability (θ) by using IRT modeling (see Section 2.4.1). In the first phase, all three different versions within each subject were equated – that is, their difficulty levels were adjusted to make them comparable – after which all three student datasets (Mathematics, Nepali, and Social Studies) were merged. However, as the Mathematics test was more demanding than the other tests,[66] hence, without any further transformations, the lowest students would automatically be those with the lowest achievement in the

---

[66] The mean in Mathematics was 42.8, in Nepali 48.6, and in Social Studies 49.3.

Mathematics test. That is why the equated scores were further transformed by using so called 'mean equating' which means that the means of the different subject populations were shifted so that all the distributions had the same mean.[67] Technically speaking, the mean of the means (46.91) was taken as the baseline for the transformation. All the student scores were shifted depending on the subject tested in the school; Mathematics schools were shifted up and Nepali and Social Studies were shifted down. In the graphs and statistical outputs one may see a variable label 'P_Eqd_Total_mean_SHIFTED' which refers to this variable. After the shift, a very few rare cases of the student appear to have a lower value than 0 or higher than 100. These are out-of-range values but they were not altered.

The equated and shifted total score in this whole data is used when estimating the student effect with Multi-level modeling (Goldstein, 1986) or Hierarchical linear modeling (Bryk & Raudenbush, 1987).

Another variable derived from the previous one is the indicator of the deciles (Q1 and Q10). Q1 students were given the code 0 and Q10 schools the code 1; this division was made for the students in community schools only. This variable can be used in two ways. First, it is a nominal variable which groups the students into two categories. This way the variable can be used as a fixed factor in ANCOVA modeling; as such, one asks whether there is a significant difference between the Q1 and Q10 with regard to some other variables, such as the attitudes towards the school. Second, the same variable can be interpreted as an interval variable; so that the mean of the variable tells the proportion of better achieving students in relation to some grouping variable, such as the language background of the student. This way the variable can be used in DTA modeling (see the next Section); in this way one can ask whether there is a significant difference between the proportions of highest performing schools in different classes based on the grouping variable.

4.4.2.4 Statistical methods

Four main methodological tools are used in the analysis: Multilevel modeling (MLM), Analysis of Variance (ANOVA), both Logistic- and Traditional Linear regression analysis (LRA and RA), and Decision tree analysis (DTA). MLM is used only in estimating the school effect as the proportion of two variance components: school-

---

[67] In the 'mean equating' one assumes that the distributions are equal but their locations are different. Now it is known that the distribution of mathematics subject differed from the other subjects (see Section 3.1.3). It was taken as a minor thing compared with the radically lower mean in the population. Other option would have been so called 'linear equating' which is used when there is a difference between the means as well as between the variance.

related variance and student-related variance. MLM is run by using the original Theta value and the full student data (around 50,000 students) – the Theta values are comparable over the subjects; zero refers to the average students in each subject. ANOVA, LRA, and RA are basic methods referred in many textbooks (see, for example, Metsämuuronen 2013) and hence they are not described here. DTA is one of the data mining tools in SPSS software. It is used when desiring to segment the data into statistically the most significantly deviating groups. Figure 4.5.5 shows two examples of a DTA output.

Proportion of Q10 schools

P_Decile_only_Q1_and_Q10_
COMMUNAL_of_P_Eqd_Total_
mean_SHIFTED

Node 0	
Mean	0,544
Std. Dev.	0,498
n	6864
%	100,0
Predicted	0,544

q9e Before or after school: I work
at a paid job
Adj. P-value=0,000, F=664,298,
df1=1, df2=6862

<= 0,0   > 0,0

Node 1	
Mean	0,635
Std. Dev.	0,481
n	4961
%	72,3
Predicted	0,635

Node 2	
Mean	0,305
Std. Dev.	0,460
n	1903
%	27,7
Predicted	0,305

Working after the school

q9e Before or after school: I work
at a paid job

Node 0	
Mean	0,604
Std. Dev.	1,122
n	6864
%	100,0
Predicted	0,604

P_Decile_only_Q1_and_Q10_
COMMUNAL_of_P_Eqd_Total_
mean_SHIFTED
Adj. P-value=0,000, F=457,418,
df1=1, df2=6862

<= 0,0   > 0,0

Node 1	
Mean	0,910
Std. Dev.	1,266
n	3132
%	45,6
Predicted	0,910

Node 2	
Mean	0,347
Std. Dev.	0,910
n	3732
%	54,4
Predicted	0,347

Figure 4.5.5 Two examples of DTA output

In the example on the left-hand side, the Q1- and Q10 students are analyzed; the proportion of the highest performing students (P_Decile_only_Q1_ and_Q10_COMMUNITY_of_P_Eqd_ Total_mean_SHIFTED) has been used as the depende*nt* variable and the amount of paid work as a side of school (q9e) as the *independent* one. The first box, the 'mother node', indicates that, in the data, the proportion of the Q10 students is 54% (0.544), standard deviation $s = 0.498$, and the number of the students who gave the answer, is $n = 6864$. In the statistical process, the variable "I work at a paid job" has been divided in as many ways as is possible to find the grouping which produces statistically the most significant difference between the groups in regard to the variable "proportion of the highest

performing students". Output, that is, two 'child nodes', indicates that in regard to maximizing the number of the highest performing students, the amount of work in a paid capacity can be divided into two groups by using the following cut-offs: not at all ($<= 0$) and other than nothing (> 0). The number of the highest performing students is significantly lower (30.5%) in the group of the students who worked in a paid capacity and higher (63.5%) if not worked at all. This division of average study days to exact these two groups produces the statistically highest test statistics (here $F = 664.30$ with the degrees of freedom of $df_1 = 1$ and $df_2 = 6862$) which produces the lowest p-value (here as an adjusted one: Adj. P-value < 0.001).

In the example on the right-hand side, the question is set in a different way with the same variables as in the example on the left-hand side. Here, it is asked how the division into Q1- and Q10 students explains the amount of worked at a paid job. DTA shows that there is a statistically significant difference between the students: in the group of the lowest level students ($<= 0$), the average amount of working is 0.91 (that is, almost 1 at the scale of maximum 4) and in the highest level schools it is 0.35 (that is, much nearer the value 0 indicating no work at all). The difference is statistically significant ($F_{(1, 6862)} = 457.42$, $p < 0.001$). In what follows, the latter question is analyzed by using Analysis of Covariance (ANCOVA) or t-test rather than DTA.

DTA does not know the reasons for the segmentation; it just gives the cold fact. In what follows, DTA is used with the CHAID algorithm (Kass, 1980). When the graphs are wide – as they are easily with several variables, only selected parts of the graph are shown, or the results are explained verbally.

4.5.3 Results

4.5.3.1 General note of the student effect and deviance within students

4.5.3.1.1 Student effect

In Sections 4.3.3.1 and 4.4.3.1, it is seen that teacher/school effect is, on average, 0.68.[68] The figures derived from MLM with Restricted Maximum Likelihood Estimation (RMLE) gives the following results:

---

[68] If using the equated score as a basis for the calculation/estimation, the results would have been 59%. However, the equated score, compared with the original Theta, is cruder; in the equating process, several different values of Theta were transformed to be as equal score. Hence, Theta gives a more accurate picture of the students' performance than the equated score. The values of Theta are difficult to report as they are standardized values, though. Hence, the equated scores (and not Theta) are used otherwise in the report.

Estimates of Covariance Parameters[a]

	Parameter		Estimate	Std. Error
	Residual		.386516	.002509
Intercept [subject =	New_sch_code]	Variance	.831243	.034640

a. Dependent Variable: Theta_TOTAL.

In the output, the variance related to school is $\tau_{00} = 0.83$ and the residual variance related to students is $\sigma^2 = 0.39$. On the basis of these figures, the value of the intra-class correlation (Rho) is

$\rho = 0.831243/(0.831243 + 0.386516) = 0.68$

In other words, the variable SCHOOL explains 68% of the student variation and hence, something else explains the remaining 32%. The share of the school is remarkably high, much higher than in Finland, for example, where the value has been round 10%. The result means that the students within the individual schools are very much alike; value $\rho = 1$ would mean that they were identical within the schools. This extreme high value of the school effect means, practically speaking, that all students in the low-performing schools get low results and in the high performing schools all get good results.

The student effect is much higher in Mustang (91%) and Kalikot (69%) – see Sections 4.3.3.1 and 4.4.3.1. It can be interpreted in a way that regardless of which school the children go to, the results would be equal though the students' achievement may vary within the school; in Mustang and Kalikot the schools are usually small and hence, the students may be more or less independent actors in the school. Another option for explaining the result is that the test was administered in slightly different way in those two districts; in Mustang and Kalikot, all three subjects were tested in the same class at the same time. This means that the effect of "social work" in the testing situation was minimized.

4.5.3.1.2. Wide deviance in learning outcomes within the students

There are two kinds of measures for achievement: absolute and relative. Within the national testing, it is very difficult to say much of the absolute achievement, exept in languages. In Nepali, the Common European Framework of Reference for Languages (CEFR) was employed to identify the *absolute* achievement of students in two major language skills: Reading and Writing. It was found that, although there are many good readers and writers in the sample, an average 8th grader reader of Nepali cannot read and understand newspapers independently, for example (see Section 4.6). The risk of being a weak reader is higher in language groups of Tamangs, Tharus, and Limbus and especially amongst the Madhesi

population. An average 8th grader writer of Nepali cannot make, for example, lecture notes or brief summaries independently. 8% of the students are very weak in writing – the risk is very high in the language groups of Tharus, Limbus, and Gurungs and especially in Madhesi societies.

In Mathematics, the population is not normally distributed as it is Nepali and Social Studies (see Section 3.1). The distribution shows that there are three distinctive student populations: low-, mediocre-, and high performing students. The low-performing students achieve, on average, 20–25% of the maximum score, the medium-performing students 40–50%, and the high-performing students as high as 70–80%. The majority of the students in Mathematics lie in the low-performing group.

The study confirms the well-known fact of a great difference between the achievement levels of the students and schools. Some students (n = 10) in the survey did not give a correct answer to a single test items (0% of the maximum marks correct) while the best students (n = 46) gained more than 95%. The average achievement of the students in the lowest performing schools (n = 5) was below 15% while the students in some of the highest performing schools (n = 3) gained, on average, over 90%. When knowing that the average score in a school is below 15%, it means that, in those schools, many students fall below that boundary. The differences are waste and the result means that the low- and high-performing students are concentrated in certain schools. From the equal opportunities viewpoint, this is not a good sign.

The remaining Chapters try to find out why many students are at the very low level and the other at the very high level.

4.5.3.2 Student's, family's and peer group's characteristics and processes related to high- and low-performing students

In this section, several sets of variables are analyzed to find out what kind of difference can be found between the highest and lowest performing community schools in the highest and lowest quintiles. The variables are segmented on the basis of Figure 4.5.2 into eight groups: 1) Students' age and sex, 2) Students' attitudes toward the subject and toward the school, 3) Students self-efficacy and Self concept, 4) Students' working habits and time spent on studies, 5) Home background in language and caste, 6) Home background in socioeconomic status, 7) Home possessions, accessories, and given help in studies, and 8) Bullying.

4.5.3.2.1 Students' age and sex within the high- and low-performing students

In the Nepalese context, the age of the students attending to grade 8 studies varies remarkably. Some students have responded (or the external coders have inputted) their age as below 10 and some above 20. All the ages of the students below 13 were encoded as '13 or less', and all students above 19 as '19 or more'. The descriptive statistics of the mean in each year are given in Table 4.5.1 and visualized in Figure 4.5.6.

Table 4.5.1 Descriptive statistics of the students' achievement in different age groups

Age	N		Mean		Std. Deviation	
	Lowest decile (Q1)	Highest decile (Q10)	Lowest decile (Q1)	Highest decile (Q10)	Lowest decile (Q1)	Highest decile (Q10)
13 or less	516	1036	13.6	74.5	5.33	6.63
14	915	1494	15.0	74.7	4.71	6.77
15	1165	911	14.6	73.9	4.71	6.24
16	738	441	15.1	73.1	4.42	5.78
17	370	132	14.6	73.4	4.36	5.92
18	159	48	15.1	73.7	4.58	6.60
19 or more	93	25	14.6	72.3	4.23	6.38
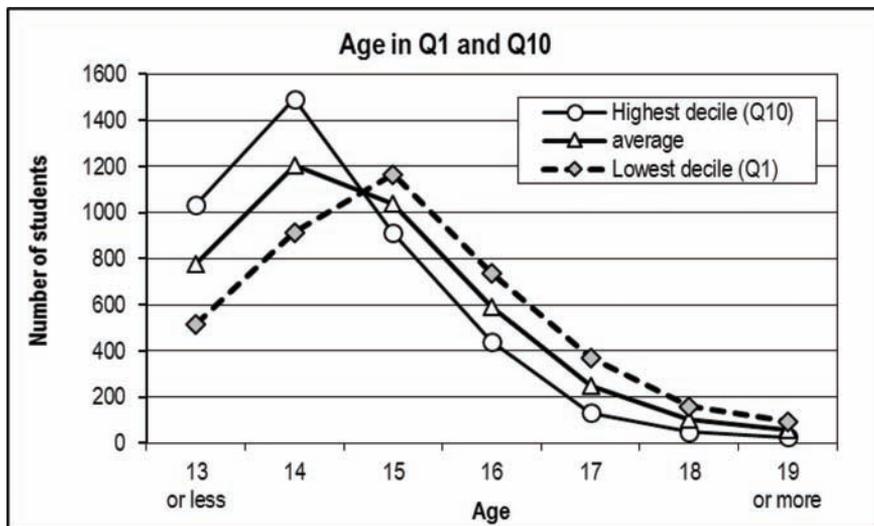
Figure 4.5.6 Ages of the lowest and highest performing students

365

All the age groups are observed in both the lowest and highest achieving groups. However, it is worth emphasizing that the "over-aged" students (15 years or higher) are overpopulating the lowest performing group.[69] Practically speaking, there are two times more students with the age *14 or lower in the highest* decile compared with the students of 15 years or more. In parallel, there is more than 3 times the risk when the student is *17 or older*, *that they will be in the lowest* decile group. It seems evident that the best achievers are those students who are at the proper age for grade 8 studies (13 to 14 years old). The higher the age is – meaning that the students have either started much later than they should have, or they have repeated classes – the less probable it is to find them within the highest performing group. Another side of the matter is that it is good that these "over-aged" students *are* in school to learn; they should have been identified at a much earlier age for extra support with their studies.

After taking into account the socioeconomic background, there is no difference in the achievement level between the boys (14.5) and girls (14.8) when it comes to the lowest level students (ANCOVA $p = 0.080$) (Table 4.5.2). Within the highest level students, the boys seem to be performing better (74.5) than girls (73.9) though the difference is not wide ($p = 0.001$, $\eta^2 = 0.003$). It is also noteworthy that, statistically thinking, there are the same number of girls in the lowest and highest decile ($p = 0.217$) but there are more boys in the highest (52%) than the lowest decile (48%, Binomial probability $p = 0.001$).

---

[69] Statistically thinking, there should be the same amount of students from each age group in both the lowest- and highest achieving groups ("average" in Figure 4.5.6). Now the differences between the number of students in each age group are so wide that the Binomial probability, which tests whether the difference is within the boundaries of random error, give the signal that the difference is absolutely statistically significant. For example, in the age group 13 or lower, the probability of observing only 516 cases out of 1552 (= 516+1036) cases when expecting around 776 (= [516+1036]/2) cases is 1.1×10^{-40}, which is ultimately a low probability; behind the figures there is a phenomenon that the younger students are performing better than the older students.

Table 4.5.2 ANCOVA table of SES (Covariant) and Sex (Fixed factor) explaining the achievement

Tests of Between-Subjects Effects

Dependent Variable: P_Eqd_Total_shifted (Mean-equated subjects)

P_Decile	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
	Corrected Model	68.556[a]	2	34.278	1.543	.214	.001
	Intercept	268357.024	1	268357.024	12083.134	.000	.753
	P_SES	.246	1	.246	.011	.916	.000
Lowest	STQ4_Gender	68.243	1	68.243	3.073	.080	.001
decile	Error	87881.901	3957	22.209			
	Total	939510.785	3960				
	Corrected Total	87950.457	3959				
	Corrected Model	959.633[b]	2	479.816	11.415	.000	.006
	Intercept	6140735.982	1	6140735.982	146091.564	.000	.973
	P_SES	532.178	1	532.178	12.661	.000	.003
Highest	STQ4_Gender	486.922	1	486.922	11.584	.001	.003
decile	Error	171958.943	4091	42.033			
	Total	22721308.584	4094				
	Corrected Total	172918.575	4093				

a. R Squared = .001 (Adjusted R Squared = ,000)

b. R Squared = .006 (Adjusted R Squared = ,005)

The dataset gives a strong signal that it is significantly less probable to find a student aged 15 or older within the highest performing student group of 8th graders. The highest performance is found with those students studying within their normal age group, that is, at the age of 13 and 14 years. There is no difference in the achievement level between boys and girls in the lowest performing students. However, the boys seem to be better than girls within the highest performing students.

4.5.3.2.2 Students' attitudes toward the subject, teacher, and school within the high- and low-performing students

Several sets of questions covering the attitudes toward the subject (STQ17a to STQ18e), the teacher (STQ27 and STQ28), and school (STQ28a and STQ28b) were set at the students. The attitudes toward the subject were assessed by

using the same test which is in use in PISA and TIMSS studies: the shortened Fennema-Sherman test. In the European- and Anglo-Saxon countries the intended structure of the test includes three dimensions: Attitudes toward the subject, Self-Efficacy in relation with the subject, and utility of the subject. Here the attitudes and utility are reported; the self-efficacy is handled in the next Section 4.5.3.2.3. The variables used in the analysis are collected in Table 4.5.3.

Table 4.5.3 Variables used in assessing students' attitudes toward the subject, teacher, and school

Attitude toward the subject (Fennema-Sherman test)
q17d FS: I enjoy learning [Subject]
q17b FS: I would like to take more [Subject] in school
q17a FS: I usually do well in [Subject]
Utility in the Subject (Fennema-Sherman test)
q18a FS: I think learning [Subject] will help me in my daily life
q18d FS: I would like a job that involved using [Subject]
q18c FS: I need to do well in [Subject] to get into the <university> of my choice
q18e FS: I need to do well in [Subject] to get the job I need
q18b FS: I need [Subject] to learn other school subjects
Attitude toward the teacher
q27d If I need extra help, I will receive it from my teacher
q27b Most teachers are interested in student's well-being
q28d Teacher wants the students to do their best
q28c Teacher in the school care about the students
q27e Most of my teachers treat me fairly
q27c Most of the teachers really listen to what I have to say
q27a Students get along well with most teachers
Attitude toward the School
q28a School: I like come and stay in school
q28b School: Students in my school try to do their best

Summed variables were formed on the basis of the sets of variables after Factor analysis.[70] On the basis of the responses, the higher-performing students are significantly ($p < 0.001$) more content with subject, teachers, and school and they

---

[70] It may be worth mentioning that the phenomenon pointed by Metsämuuronen (2012a; 2012b) is found also in Nepal. Namely, in a number of Asian countries, the expected factor structure cannot be found. In many other countries, also in Nepal, the negative formulated items correlate with each other more than they should.

sense more utility in the subject (Table 4.5.4) though the differences between the groups is not very wide; both groups feel the subject somewhat equally useful for them (effect size is $d = 0.18$), for example.

Table 4.5.4 Students' attitudes toward the subject, teacher, and school within the Q1- and Q10 students

Variables	Lowest decile (Q1)	Highest decile (Q10)	t	df	Sig. (2-tailed)	Cohen's d
Attitude toward Subject	1.4[1]	1.2	13.274	7056	<0.001	0.31
Utility of the Subject	1.5	1.4	7.984	7444	<0.001	0.18
Attitude toward teacher	1.5	1.3	19.369	6804	<0.001	0.46
Attitude toward school	1.3	1.2	11.651	6680	<0.001	0.28

1) All the scales were anchored to 1 = fully agree and 4 = fully disagree. Hence, the lower the value is, the more positive the answer.

It seems that the differences are the most notable in the area of attitude towards the teacher ($d = 0.46$); the effect size is moderate. The set of variable may, however, reflect the teachers' activity more than the attitude of the students; it is known that in the highest performing schools, the teachers act more actively than in the lowest performing schools (see Section 4.4.3.3.1).

4.5.3.2.3 Students self-efficacy and self-confidence within the high- and low-performing students

The self-efficacy or self-confidence of the students in managing the subject matter were assessed with a set of four variables embedded into the Fennema-Sherman test. It is an interesting fact that the students' self-efficacy, measured by using the sum of the Fennema-Sherman test does not differ between the lowest and highest performing students. A technical reason for this is that in some items the lowest performing students felt themselves more confident than the highest performing students (Table 4.5.5).

Table 4.5.5 Students' self-efficacy within the Q1- and Q10 students

Variables[1]	Lowest decile (Q1)	Highest decile (Q10)	t	df	(2-tailed) Sig.	Cohen's d
q17a FS: I usually do well in [Subject]	1.4	1.3	7.556	7130	<0.001	0.17
q17c FS*[2]: [Subject] is more difficult for me than many of my classmates	2.6	2.6	-0.576	7211	0.565	-0.01
q17e FS*[2]: Sometimes, when I do not initially understand a new topic in [Subject], I know that I will never really understand it	2.5	2.5	1.778	7588	0.075	0.04
q17f FS*[2]: [Subject] is not one of my strengths	2.5	2.4	5.055	7483	<0.001	0.11
q17g FS: I learn things quickly in [Subject]	1.7	1.8	-5.097	7717	<0.001	-0.11

1) The variables did not load with the same factor. Traditionally, there is at most two negative items per dimension – in Nepal, all three negative items loaded to the same variable.
2) The variables with an asterisk (*) were originally negative; they were reversed before analysis. Hence they should be read in a positive way.

It is a mystery why the very lowest performing students think that they perform very well and are learning quickly; either their self-efficacy is radically distorted or they did not understand the questions. It is evident, that the students in Nepal use a different logic in answering the attitude questions than what was expected. One example of this comes from the variables which are mutually excluding themselves (q17a and q17f). Logically, when one feels strongly that (s)he "usually does well in the subject" (q17a) (s)he should strongly oppose the idea that "the subject is *not* one of my strengths" (q17f). However, there is practically no correlation between these variables ($r = 0.04$). This raises the question whether the students (at least in the lowest achievement group) understood the negative items at all. If not, the assessment of the self-efficacy is not necessarily meaningful (see discussion on the matter in Metsämuuronen, 2012a; 2012b).

The data gives a signal that there is no difference in self-efficacy between the lowest and highest performing students; the analysis of self-efficacy should be interpreted with caution because the lowest performing students seemed not to understand the questions.

4.5.3.2.4 Students' working habits and time spent on studies within the high- and low-performing students

A set of questions concerning the time spent before or after the school is used to assess the students' working habits and time spent on studies. Additionally, a question was asked as to how many days the student were absent from school. The results are condensed in Table 4.5.6. and illustrated in Figure 4.5.7.

370

Table 4.5.6 Differences in the time consumption within the Q1- and Q10 students

Variables[1]	Lowest decile (Q1)	Highest decile (Q10)	t	df	(2-tailed) Sig.	Cohen's d
q9g Before or after school: I do homework	1.8[2]	2.4	-22.53	6656	<0.001	0.52
q9e Before or after school: I work at a paid job	0.92	0.3	20.80	5559	<0.001	0.51
q13A How many days you were absent from the school	4.7	2.3	19.15	5820	<0.001	0.49
q9c Before or after school: I play sports	0.82	0.6	11.13	6544	<0.001	0.26
q9b Before or after school: I play or talk with friends	0.72	0.5	9.37	6753	<0.001	0.22
q9f Before or after school: I read a book for enjoyment	1.12	0.9	7.83	6731	<0.001	0.18
q9a Before or after school: I watch television and videos	0.62	0.5	5.32	6863	<0.001	0.12
q9d Before or after school: I do jobs at home	1.52	1.4	4.59	6873	<0.001	0.10

1) Variables are profiled on the basis of the value in Cohen's d.
2) The scale is 0= not at all, 1 = less than 1 hour, 2 = 1 – 2 hours, 3 = 2 – 4 hours, and 4 = more than 4 hours.
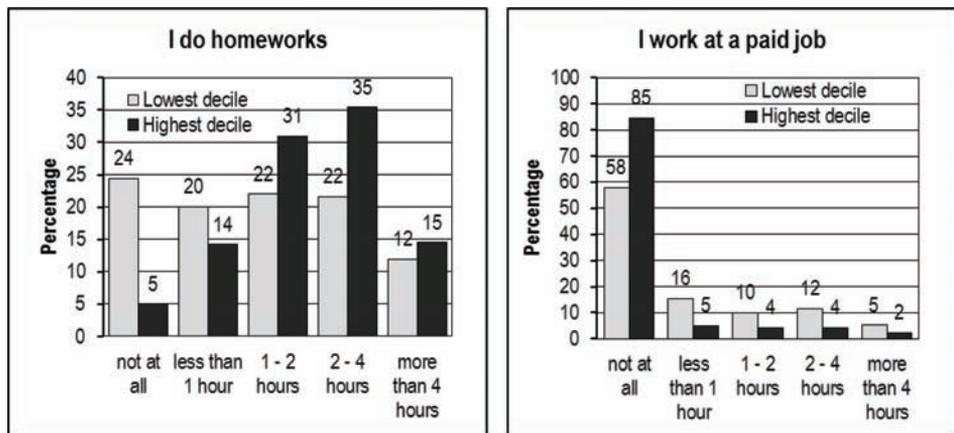
Figure 4.5.7   Time spent with the homework and paid work within the lowest and highest performing students

Though all the variables in Table 4.5.6 show statistically significant differences between the lowest and highest performing students ($p < 0.001$), actually there are three variables which stand above the others: The highest performing students spend *remarkably more time on homework* (1–4 hours/week), they *work less in a paid capacity* (practically speaking not at all), and they are *less absent from school* (average 2.3 days per year) compared with the lowest performing students; the effect sizes are moderate ($d > 0.49$).

The data gives a strong signal that a reasonable amount of time spent on homework given by the teacher, presence in school, and time not spent in paid work raises the achievement level remarkably.

4.5.3.2.5 Language background and caste within the high- and low-performing students

On the basis of the language background of the family, the students from the Rai and Limbu background are remarkably less represented in the highest performing group and highly represented in the lowest level groups. It is actually more than a remarkable inequity that there are 50 times more Rai students and 9 times more Limbu students in the lowest performing group than in the highest performing group (Table 4.5.7. and Figure 4.5.8). The effect sizes are huge (Cohen's *h* equal 2.6 and 1.8 respectively). On the other hand, there are remarkably more Urdus, Tamangs, and Newars in the highest performing group than is expected by a change (Effect sizes equal 1.70, 0.77 and 0.71 respectively). On the basis of the data, it seems evident that the students in the different language background do not have equal opportunities to advance towards higher education. However, it is good to note that the students with Rai and Limbu background were found practically only in the Eastern region. Their achievement may be higher in some other areas which were not selected in the sample. However, the data gives a strong signal that at least in the groups of the Eastern region's Rai and Limbu students more actions should be addressed to find out how large this inequity is.

Table 4.5.7 Differences in language background within the Q1- and Q10 students (number of cases)

Language group[1]	Lowest decile (Q1)	Highest decile (Q10)	Binomial p[2]	Cohen's h[3]
Rai (n = 102)	100	2	< 0.001	2.57
Limbu (n = 29)	26	3	< 0.001	1.83
Magar (n = 196)	117	79	0.004	0.39
Tharu (n = 252)	140	112	0.044	0.22
Urdu (n = 8)	1	7	0.035	-1.70
Tamang (n = 279)	87	192	< 0.001	-0.77
Newar (n = 31)	10	21	0.035	-0.72
Sherpa (n = 8)	3	5	0.363	-0.51
Gurung (n = 71)	31	40	0.171	-0.25
Nepali (n = 5087)	2275	2812	< 0.001	-0.21

1)   The language groups are profiled on the basis of the effect size (Cohen's h)
2)   Test for equality of numerous in both groups
3)   Cohen's *h* is an effect size for proportions. Interpretation is the same as with Cohen's *d*: lower than 0.20 is taken as "low" and more than 0.80 as "high" effect size.
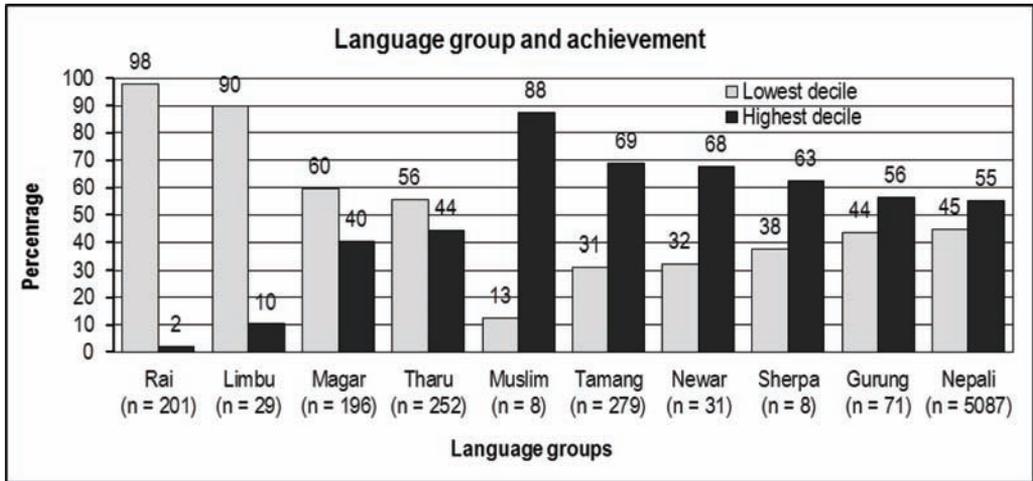
Figure 4.5.8   Differences in language background within the Q1- and Q10 students (%)

When it comes to the castes, the differences are not at all that large as they are with the language groups (Table 4.5.8. and Figure 4.5.9). However, there are remarkably more Dalits, Alpasankhyaks, and Madhesis in the lowest performing group than what is expected by change (Cohen's h > 0.32). The students from the Brahman or Cheetri background seem to be statistically significantly overpopulated in the highest performing group. Actually then, only the Janjati students are evenly distributed in both the highest and lowest performing groups.

Table 4.5.8 Differences in caste background within the Q1- and Q10 students

Caste[1]	Lowest decile	Highest decile	Binomial p[2]	Cohen's h (effect size)[3]
Dalit (n = 850)	510	340	< 0.001	0.40
Alpasankhyak (n = 125)	74	51	0.024	0.37
Madhesi (n = 1290)	747	543	< 0.001	0.32
Brahman/ Cheetri (n = 2763)	1112	1651	< 0.001	-0.39
Janjati (n = 3008)	1501	1507	0.464	0.00

1)   The language groups are profiled on the basis of the effect size (Cohen's h)
2)   Test for equality of numerous in both groups
3)   Cohen's h is an effect size for proportions. Interpretation is the same as with Cohen's d: lower than 0.20 is taken as low and more than 0.80 as high effect size.
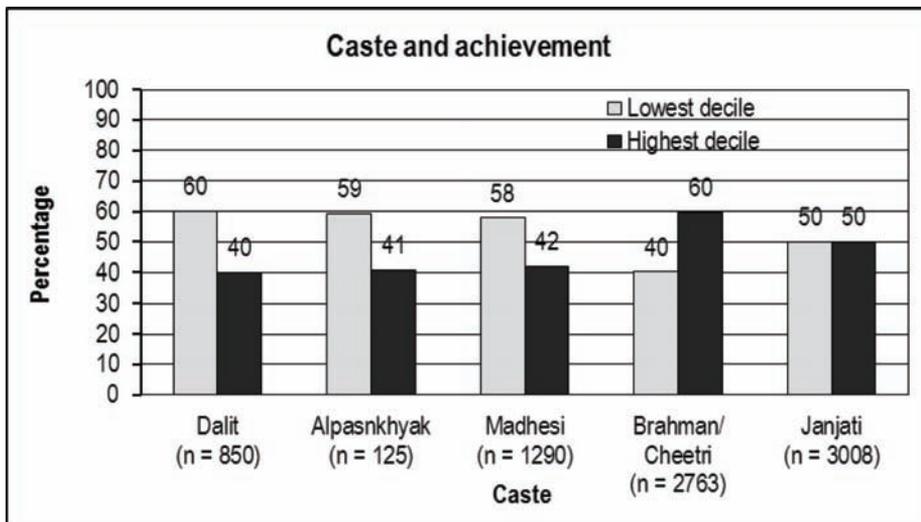
Figure 4.5.9 Differences in caste background within the Q1- and Q10 students (%)

Data gives a strong signal that there is a huge inequity between the language groups when it comes to the highest and lowest performing student groups: Rai, Limbu and Magar students are over-represented in the ultimately lowest performing student group and Urdu, Tamang, Newar, and Sherpa students in the ultimately highest performing student group. Dalits, Alpasankhyaks, and Madhesis are over-represented in the lowest performing student group and Brahman or Cheetri students are overpopulated in the highest performing group.

4.5.3.2.6 Home background and socioeconomic status within the high- and low-performing students

Several variables in the student background questionnaire highlight their socioeconomic status (SES). These were categorized into parents' education, parents' occupation, home possessions (whether or not the student has their own space to do homework, or a dictionary, for example), home accessories (how many mobile phones, cars or bathrooms there is in the student's home), and whether the student attends 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 detailed in Section 2.7.2). Of the variables related to SES, parents' education is handled in this Section and home possessions in the next one.

The socioeconomic status was formed on the basis of seven indicators which were all first dichotomized (see Section 2.7.2). The variables (mother's education, father's education, mother's occupation, father's occupation, home

possessions, home accessories, and type of school where the students studying) were summed (as SES) and changed into a percentage of the maximum score (P_SES). The P_SES represents the percentage of SES the student possesses; 100 means that the student has met all the seven indicators of SES and 0 means that none are met. Even though the analysis of the P_SES shows a strong relation between SES and achievement ($p < 0.001$, $\eta^2 = 0.089$, effect size $f = 0.31$ indicating remarkable differences between the lowest and highest group, Table 4.5.9), it does not explain the differences within the lowest and highest performing students (Figure 4.5.10).[71]

Table 4.5.9 ANOVA table of SES explaining the differences in achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
	Tests of Between-Subjects Effects					
	Dependent Variable: P_Eqd_Total_shifted USE THIS (Mean-equated subjects)					
Corrected Model	1523394,928[1]	7	217627,847	676,179	,000	,089
Intercept	64998777,821	1	64998777,821	201953,934	,000	,806
P_SES	1523394,928	7	217627,847	676,179	,000	,089
Error	15665703,818	48674	321,850			
Total	124305667,979	48682				
Corrected Total	17189098,747	48681				

1) R Squared = ,089 (Adjusted R Squared = ,088)
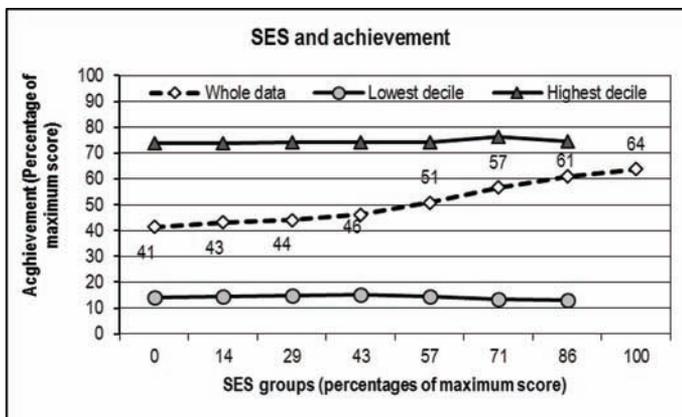
Figure 4.5.10 Connection of SES and achievement in the whole data

---

[71] By using SES as a covariate when explaining the differences between the Q1- and Q10 students, it appears to be a statistically significant covariant ($p = 0.003$) but the explanatory power in explaining the differences within the student groups is marginal $(\eta^2 = 0.001)$; the effect size $f = 0.03$ shows that the differences in achievement are very small *within* the lowest- and highest SES group. This can be seen clearly in Figure 4.5.10.

A closer view to the SES variables is given by focusing on parents' education and occupation in relation to the achievement level. Parents' education is divided into nine categories: 1) illiterate, 2) literate, 3) grades 1–5 pass, 4) grades 6–10 pass, 5) SLC pass, 6) IA pass, 7) BA pass, 8) MA pass, and 9) Above MA pass.[72] In order to simplify the situation, the parents' educational level is divided into "illiterate" and "other than illiterate". Table 4.5.10 condenses and Figure 4.5.11 illustrates the differences between the groups. Though there are statistically significantly more illiterate mothers and fathers within the lowest performing students ($p < 0.001$), the differences in numbers are not necessarily very wide when it comes to illiterate mothers (Cohen's $h = 0.18$). The difference is much wider when it comes to the illiterate fathers ($h = 0.43$) though there are much more illiterate mothers than fathers.

Table 4.5.10 Differences in parents' educational background within the Q1- and Q10 students

Variable	Lowest decile (Q1)	Highest decile (Q10)	Binomial p[1]	Cohen's h[2]
Mother Iliterate	2025	1688	< 0.001	0,18
Mother other than illiterate	1887	2397	< 0.001	0,24
Father Illiterate	989	641	< 0.001	0,43
Father other than illiterate	2900	3441	< 0.001	0,17

1) Test for equality of numerous in both groups
2) Cohen's h is an effect size for proportions. Interpretation is the same as with Cohen's d: lower than 0.20 is taken as "low" and more than 0.80 as "high" effect size.
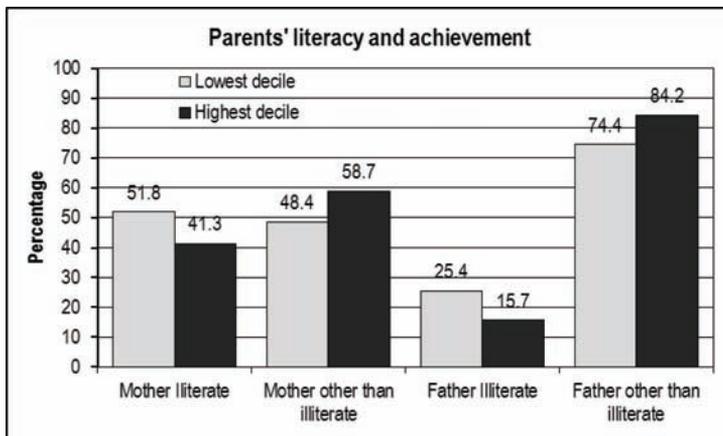
Figure 4.5.11 Differences in parents' educational background within the Q1- and Q10 students (%)

[72] The questions of parents' education and occupation were asked from the students and hence there may be some impurities embedded in the data. It is credible that, within the lowest-performing students, there are significantly more illiterate mothers and fathers than within the highest-performing students. However, it is difficult to explain why, within the *lowest*-performing students, there are almost four times more mothers (144 to 39) and two times more fathers (196 to 91) *above* MA pass level compared with the highest-performing students. According to the students, some of their illiterate mothers and fathers worked as teachers. Though the latter two things may be possible, it still raises the question whether the lowest-performing students have understood the question in the same way as the highest-performing students.

From the number of the students, one can infer a positive fact: the students with illiterate parents are not doomed to be the lowest performing students. Quite the opposite: many students from illiterate families can reach a high achievement level. However, when both parents are illiterate, or the father is illiterate, the probability of being within the highest performing students is quite low (39%) compared with the situation when both the parents are literate and they are not working in agriculture (64%) (Figure 4.5.12).
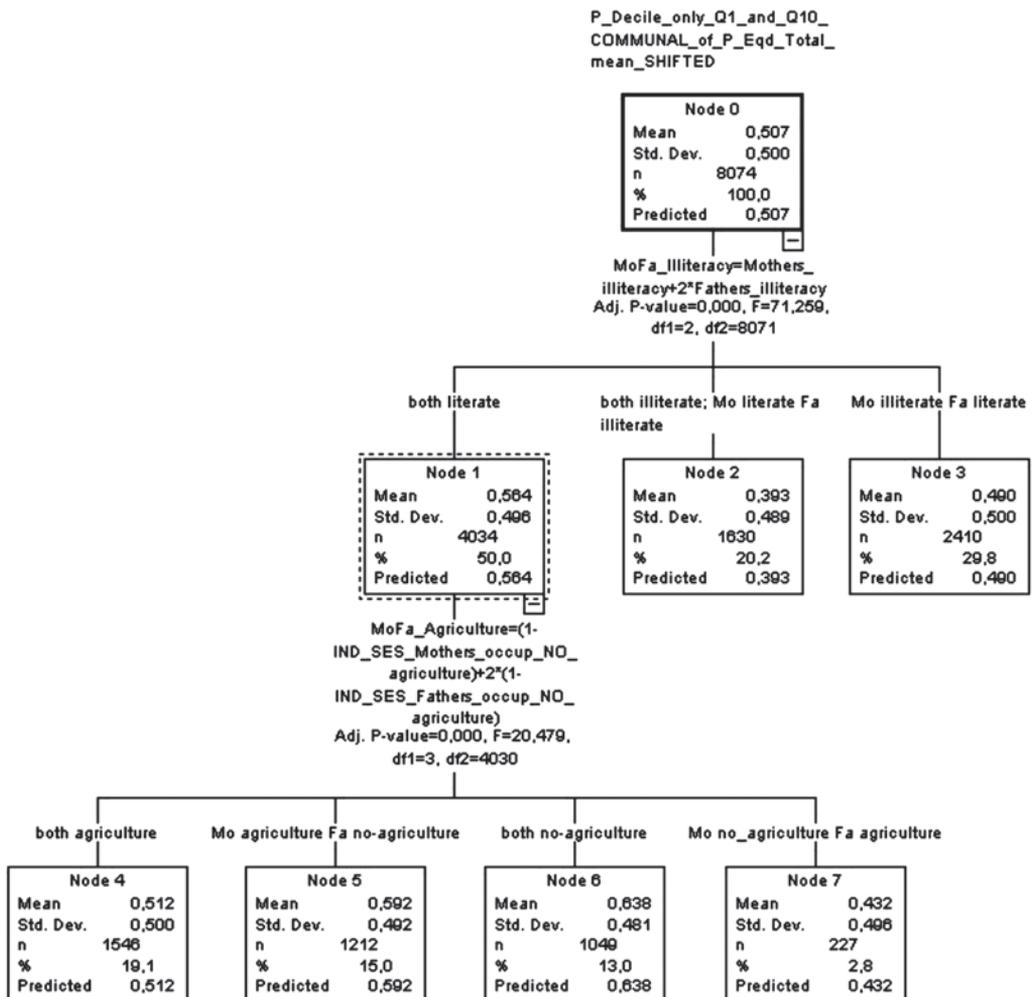
Figure 4.5.12 DTA of parents' education and occupation within the Q1- and Q10 students

The dataset gives a strong indication that the socioeconomic status plays a strong role in the educational processes in Nepal. The difference between the lowest and highest SES groups is remarkable (22–25 percent points). This means that if the social economic standard of the lowest performing students was raised to a decent level, that is, in practice, that the problems in four out of seven indicators were solved, the results in these groups would raise remarkably. Especially challenging is the situation in the families where both parents are illiterate or they both work in agriculture.

4.5.3.2.7 *Home possessions, accessories, and given help in studies within the high- and low-performing student*

The facilities and resources available at home may have some effect on achievement levels. There were two kinds of home possessions defined in the background information questionnaire for the students. One is related to the facilities that help them to study at home: whether they have a table for their studies, 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, classical literature, poetry books, artistic things like pictures, books that help them in their studies, 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 highlight their difference from home possessions) such as the number of mobile phones, televisions, computers, cars, and bathrooms.

There are 12 questions in the student background questionnaire related to home possessions and 5 related to home accessories. Each possession was scored 1 if the student had access to this possession (e.g. having a separate room or a table to study). Adding these items up, the maximum score was 12 indicating that the student expressed to have access to *all* of the possessions and the lower the score the less possession they have access to at home. Table 4.5.11 and Figure 4.5.13 show the connection between home possessions and achievement levels: In the whole data the achievement level of the students' raises logically, though moderately, the more there is access to the home possessions. However, the home possessions do not explain the division into Q1 and Q10 groups remarkably though it is a statistically significant predictor ($p < 0.001$); the effect size estimated on the basis of the Eta squared ($\eta^2 = 0.010$) is $f = 0.10$ which is very low a value. The same pattern – the more possession, the better results – can be seen with home accessories. Though the trend is clear in the whole data, the number of home accessories does not explain the differences within the lowest and highest performance students ($f = 0.12$).

Table 4.5.11 ANOVA table of home possessions and accessories within the Q1- and Q10 students

| | Tests of Between-Subjects Effects | | | | | |
| | Dependent Variable: P_Eqd_Total_shifted USE THIS (Mean-equated subjects) | | | | | |
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7045748.375[1]	18	391430.465	12302.930	.000	.965
Intercept	1975398.573	1	1975398.573	62088.141	.000	.887
P_Decile	6665508.889	1	6665508.889	209501.545	.000	.963
home_possessions	2519.501	12	209.958	6.599	.000	.010
home_accessories	3561.899	5	712.380	22.391	.000	.014
Error	252523.885	7937	31.816			
Total	23564476.038	7956				
Corrected Total	7298272.259	7955				

1) R squared = .965 (Adjusted R Squared = .965)
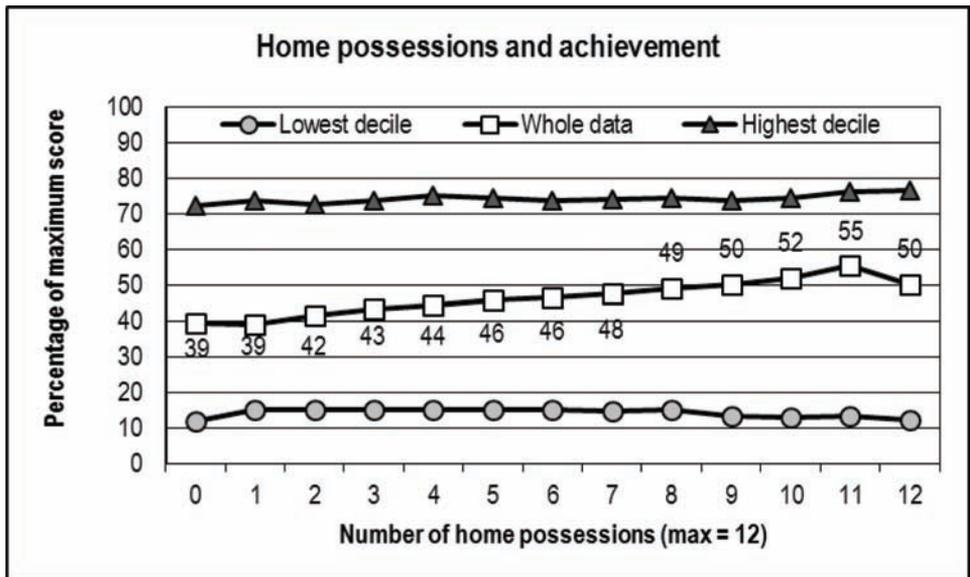
Figure 4.5.13 Connection of home possessions and achievement within the Q1- and Q10 students

Additional to the economic input (seen in the possessions and accessories above), the parents can also contribute their intellectual capacity to their children's

379

education. One of these kinds of questions was set on the students: "Who helps you when you do not understand what you have read?" The main helper in both the lowest and highest performing group is the *teacher* though more of the highest performing students have taken advantage of their teacher's time (61%) than of the lowest performing students (53%). The difference is notable though not necessarily remarkable (Effect size $h = 0.17$). Fathers have helped more of the lowest level students (8%) than of the highest level students (5%); this difference is remarkable ($h = 0.44$). Otherwise there are no vast differences between the percentages – both groups have been equally helped by mother or tuition master or no one.
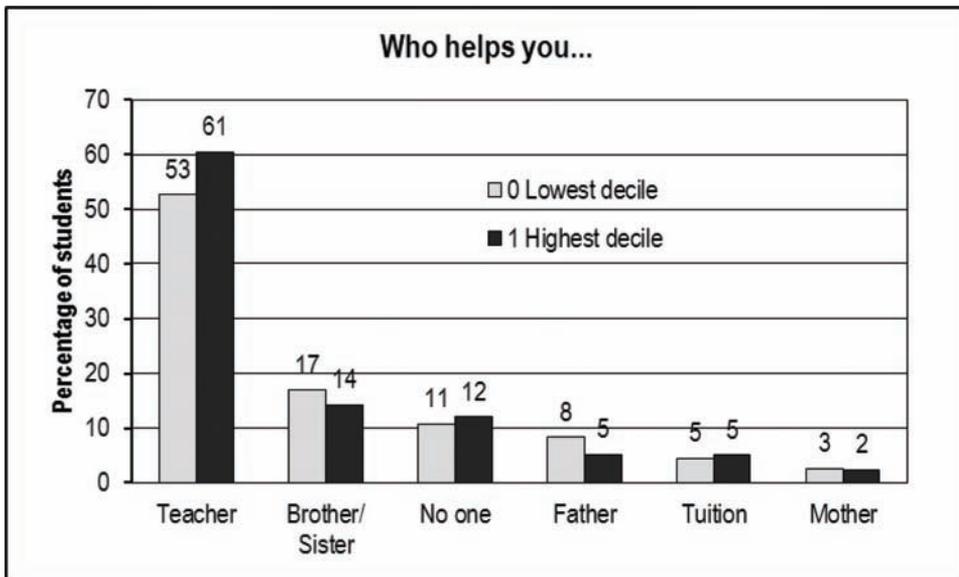
Figure 4.5.14   Who helps the student and its relation to achievement within the Q1- and Q10 students?

The data gives a signal that when the children only have very few home possessions helping them in their studies the achievement level is statistically lower (39–42%) than if there are more than four indicators met (> 46%). With ten to twelve possessions, the average score is very high (> 52%) compared with the national average. The same is shown with home accessories: with no accessory indicator met, the results are very poor (43%) and when there are three or more met, the results are remarkably higher (> 52%).With four or five indicators met the results are the best (56–58%). The highest performing students use their teacher more frequently when they experience difficulties in their studies.

4.5.3.2.8 Bullying within the high- and low-performing students

One set of questions was used to measure the negative effect of the peer group in studies and learning; the physical or social bullying was asked by five questions (Table 4.5.12, Figure 4.5.15). All the indicators show that bullying occurs statistically significantly more often within the lowest performing students (17 – 23% of the students) compared with the highest performing students (6 – 16% of the students) ($p < 0.001$). The difference is remarkable with two indicators of social bullying: the lowest performing students were made to do three times more things they didn't want to do ($d = 0.42$) and their fellow students kept them outside without involving them in activities ($d = 0.42$). The reason for this kind of social isolation may be the low achievement and related actions (see, for example, the less effective working habits in Section 4.5.3.2.4), language group, or caste (see Section 4.5.3.2.5). After summing up all the components of bullying, there is two times more bullying in the lowest performing group (23% of the students) than in the highest performing students (11% of the students). Regardless of the reason for the bullying, it is hardly possible to think that the negative behavior from the peer group would help the lowest level students to make better results in school. In a civilized society, efforts are made to root out this kind of behavior from schools.

Table 4.5.12 Bullying within the Q1- and Q10 students

Variables[1]	Lowest decile (Q1)	Highest decile (Q10)	t	df	Sig. (2-tailed)	Cohen's d
q29c I was made to do things I didn't want to do by other students	0.20[2]	0.07	17.158	5908	< 0.001	0.42
q29e fellow students kept outside without involving me in activities	0.18	0.06	16.828	5807	< 0.001	0.42
q29a Something of mine was stolen	0.28	0.16	12.161	7266	< 0.001	0.27
q29d I was made fun of or called names	0.18	0.12	7.207	6978	< 0.001	0.16
q29b I was hit or hurt by other student(s)	0.17	0.12	7.024	7109	< 0.001	0.16
SUM_Bullying=mean (STQ29_A to STQ29_E)	0.23	0.11	20.316	6654	< 0.001	0.48

1) The variables are profiled on the basis of Cohen's d; the most remarkable differences are at the top of the list.
2) The scale was 0 = no and 1 = yes. Hence the lower the value is, the less bullying.

381
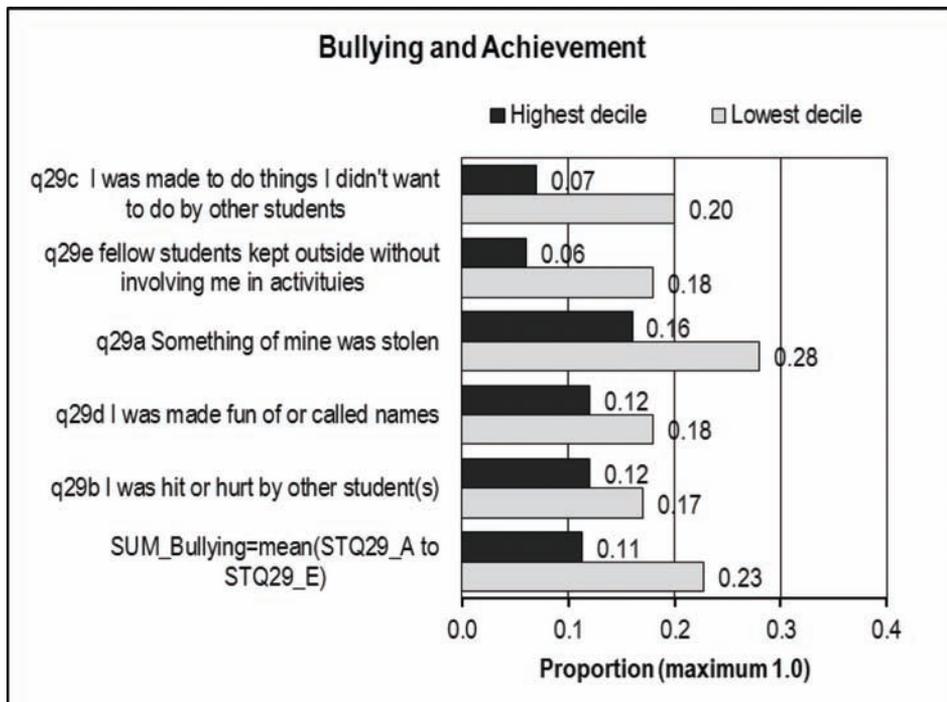
Figure 4.5.15 Bullying and its relation to the achievement within the Q1- and Q10 students

The data gives a strong signal that there are more occurrences of bullying within the lowest performing students; too often they are excluded from actions and they are made to do things that they did not want to.

4.5.3.3 Summary of the students characteristics – Modeling the phenomenon

Above, several individual student-related factors have been detected which *individually* explain the difference between the lowest and highest performing students in the community schools. These factors are collected in Table 4.5.13. Some of them may be strongly related to each other and hence they may not add value in explaining why some community schools are performing much better than others. It is also worth remembering that the school-related factors are not the only relevant variables in explaining the phenomenon; related diversity factors are handled in Section 4.1, teacher factors are handled in Section 4.3 and school factors in Section 4.4.

Table 4.5.13 Student-related factors differing statistically significantly between the Q1- and Q10 students

Personal characters
Q4 Gender
Q5 Age
Q6 Language at Home
q8 Cast

Student's working habits
q9a Before or after school: I watch television and videos
q9b Before or after school: I play or talk with friends
q9c Before or after school: I play sports
q9d Before or after school: I do jobs at home
q9e Before or after school: I work at a paid job
q9f Before or after school: I read a book for enjoyment
q9g Before or after school: I do homework

Family background and SES
q10 who helps you when you do not understand what you have read
Mother's illiteracy = 1, other 0
Father's illiteracy = 1, other 0
home_possessions=sum(NSTQ25_A to NSTQ25_L)
home_accessories=sum(C_NSTQ26_A to C_NSTQ26_E)
P_SES=sum(IND_SES_Private_School to IND_SES_Fathers_education)/7*100

Attitudes toward school and the subject
q13A how many Days you were Absent from the school
q17a FS: I usually do well in [Subject]
q17g FS: I learn things quickly in [Subject]
q17f FS*2: [Subject] is not one of my strengths
Self_Efficacy_in_Subject=mean(STQ17_C_INV,STQ17_E_INV,STQ17_F_INV,STQ17_G)
Attitude_toward_Subject=mean(STQ17_a,STQ17_b,STQ17_d)
Utility_of_the_Subject=mean(STQ18_A to STQ18_E)
Attitude_toward_teacher=mean(STQ27_A to STQ27_E,STQ28_C,STQ28_D)
Attitude_toward_school=mean(STQ28_A,STQ28_B)

Peer group factors and bullying
q29a Something of mine was stolen
q29b I was hit or hurt by other student(s)
q29c I was made to do things I didn't want to do by other students
q29d I was made fun of or called names
q29e fellow students kept outside without involving me in activities
SUM_Bullying=mean(STQ29_A to STQ29_E)

4.5.3.3.1 Modelling the highest and lowest quintile community schools with the school-related variables

Logistic regression analysis was used to analyse which of the factors are independent predictors for belonging to either the lowest or highest performing schools. Table 4.5.14 shows the statistically best model (Conditional-selection) after some transformations. Compared with the models with the teacher-related- (Section 4.3.3.3.1) and school-related variables (Section 4.4.3.3.1) it is notable that quite a number of the student-related variables in Table 4.5.14 actually have their own (main) effect on the whole model. During the analysis, the statistically significant predictors were first detected by using the original variables. After this, DTA was used individually with each variable and the best cut-offs of the variables were detected to dichotomize the predictors; the original variable was transformed as a dummy variable to simplify the analysis and the interpretation. [73]

All the variables in the combined model in Table 4.5.14 are statistically significant predictors for a student to belong to either the lowest or highest performing groups of the students. The whole model explains 39% of the grouping (Nagelkerke $R^2 = 0.386$). This means, in practice, that 74% of the students could be classified correctly though the model is much better detecting the higher-performing students (81% correct).

The last column on the table, Exp(B), tells the "risk" of being in the group of the highest performing students when the value of the dummy variable is higher. Out of the many variables, five are worth noticing; all of them show more than two times the "risk" of belonging to the highest performing student group: Compared with the lowest performing students, the highest performing students spend *more time on homework* and they are *less absent from school,* more of them *do not need to work while studying*, they are *younger*, and their *attitude toward the subject is positive*. All variables make sense: high-performing students in the community schools are working hard and they have enough economic support to concentrate on the studies.

---

[73]   In many cases, though, DTA suggests three or more categories for the most significant grouping. For example, when analyzing the question "who helps you when you do not understand what you have read?", DTA suggests three groups: the group where the mother or father helped the student, the group helped by the brother or sister, and the group helped by the teacher, tuition master, or no one. The last group predicted the best the belonging to the highest-performing group.

Table 4.5.14 Statistically the best model of Logistic regression analysis explaining belonging to the groups of Q1- and Q10 students (Method = Forward: Conditional)

Variables in the Equation [1,2]	B	S.E.	Wald	df	Sig.	Exp(B)
q9g_dummy Before or after school: I do homework less than one hour per week = 0, other 1	1.088	0.065	278.575	1	< 0.001	2.969
STQ13A_dummy Absence <5 days = 1, = > 5 days = 0	0.879	0.066	178.972	1	< 0.001	2.409
q9e_dummy Before or after school: I work at a paid job not at all = 1, other 0	0.87	0.061	200.568	1	< 0.001	2.388
STQ5_Age_dummy 14 years or less = 1, = >15 = 0	0.842	0.058	209.751	1	< 0.001	2.321
Attitude_toward_subject_dummy <2 = 1, =>2 = 0	0.735	0.089	68.696	1	< 0.001	2.085
q29e_dummy I was hit or hurt by other student(s) no = 1, yes = 0	0.636	0.107	35.166	1	< 0.001	1.888
Attitude_toward_teacher_dummy <1.5 = 1, = >1.5 = 0	0.623	0.066	89.589	1	< 0.001	1.865
q29c_dummy I was hit or hurt by other student(s) no = 1, yes = 0	0.623	0.099	39.378	1	< 0.001	1.864
q9c_dummy Before or after school: I play sports 0,1 = 1, other 0	0.56	0.077	53.503	1	< 0.001	1.751
q9f_dummy Before or after school: I read a book for enjoyment 0, 1, 2 = 1, other 0	0.53	0.079	44.716	1	< 0.001	1.698
STQ6_language_dummy Magar, Tharu, Rai, Limbu,else= 0, other = 1	0.454	0.067	46.295	1	< 0.001	1.575
STQ10_dummy who helps you? mother, father, brother/ sister = 0, teacher, tuition, no one =1	0.446	0.068	43.078	1	< 0.001	1.562
q9d_dummy Before or after school: I do jobs at home 0,1, 2 = 1, other 0	0.382	0.069	30.512	1	< 0.001	1.465
q29a_dummy Something of mine was stolen no = 1, yes = 0	0.363	0.075	23.5	1	< 0.001	1.437
home_accessories_dummy 0 indicators (out of 5) met = 0, >0 = 1	0.358	0.062	33.647	1	< 0.001	1.430
STQ4_Gender	-0.258	0.058	19.73	1	< 0.001	0.773
STQ17_G_dummy I learn things quickly in Nepali language 1 (very confident) = 0, other = 1	0.241	0.059	16.693	1	< 0.001	1.273
home_possessions_dummy <3 indicators (out of 12) met = 0, =>3 = 1	0.23	0.104	4.931	1	0.026	1.259
q29d_dummy I was made fun of or called names no = 1, yes = 0	-0.196	0.09	4.744	1	0.029	0.822
STQ8_Caste_dummy Brahman/Cheetri, Janjati = 1, other = 0	0.194	0.07	7.784	1	0.005	1.214
Fathers_illiteracy_dummy literate = 1, illiterate = 0	0.18	0.072	6.167	1	0.013	1.197
Constant	-5.749	.238	585.784	1	.000	.003

1) The cut-offs for the dummy variables are made on the basis of suggestions of DTA.
2) The variables are ordered on the basis of the risk value; the variable indicating the highest risk is at the top.

4.5.3.3.2 Modeling the overall achievement on the community schools with student-related variables

The methodological decision made (to concentrate in profiling only the most extreme students), lead to a situation that actually most of the students were not in the analysis; usually this is reasonable because nothing can explain the difference between students with average achievement. The decision and analysis done above, however, was carried out just to *clarify* the phenomenon; it is expected that the predicting variables found in the analysis above can also explain, to some extent, the general differences of the average achievement level in the schools. The traditional linear regression analysis with Stepwise regression was used to model the phenomenon; the equated (and shifted) total score of the students is explained by the same variables as above (see Table 4.5.13). After the predicting variables were found, they were dichotomized in order to simplify the interpretation by using DTA. Table 4.3.15 shows the results.

The model in Table 4.5.15 can be interpreted as follows: The starting value for the students in the community schools is 15.5% of the maximum score (the value of the coefficient B of 'Constant'). The coefficient B tells how much the average score would change when one unit change is seen in the predictor variable; all the predictor variables are statistically significant (at the same time). Now all the predictors are dummy variables and hence the coefficient strictly tells how much the student would have gained by having the higher category of the dummy variable. For example, if the student spent one hour or more per week on homework, the result was 5 percent points higher than if spending less than 1 hour. Similarly, if the student was absent less than 5 days per year, (s)he would have advanced 4.2 percent points.

Some of the variables in the list are areas where the student cannot do much to alter the situation: The student cannot decide to be Newari instead of being Rai or Limbu; or a boy instead of a girl; or being 13 instead of 17; (s)he cannot make his/her father literate; or add accessories to his/her home. Here comes the important role of society; society is responsible to reduce the differences between the languages, castes, gender groups or between the agricultural- or other occupational of children's parents.

However, some of the factors the student can affect. Table 4.5.16 condenses these variables.

386

Table 4.5.15 Statistically the best model of Linear Regression analysis explaining the average of student achievement in the community schools

Model[1]	Unstdzed Coeff.		Stdzed Coeff.		
	B	Std. Error	Beta	t	sig.
(Constant)	15.46	0.64		24.09	<0.001
q9g_dummy Before or after school: I do homework, less than one hour per week = 0, other = 1	4.96	0.19	0.13	25.44	<0.001
STQ13A_dummy Absence 4 days or less = 1, => 5 = 0	4.21	0.20	0.10	20.82	<0.001
STQ5_Age_dummy 14 years or less = 1, =>15 = 0	4.07	0.18	0.12	22.86	<0.001
q9e_dummy Before or after school: I work at a paid job not at all = 1, other 0	3.91	0.19	0.11	20.10	<0.001
q29e_dummy I was hit or hurt by other student(s) no = 1, yes =0	3.29	0.33	0.05	10.01	<0.001
Attitude toward teacher dummy <1.5 (very positive or positive) = 1, >=1.5 = 0	2.77	0.20	0.07	13.83	<0.001
Attitude_toward_Subject_dummy <2 (extreme positive) = 1, =>2 = 0	2.60	0.28	0.05	9.33	<0.001
q9f_dummy Before or after school: I read a book for enjoyment 0,1, 2 hours = 1, more than 2 = 0	2.35	0.25	0.05	9.45	<0.001
q29c_dummy I was hit or hurt by other student(s) no = 1, yes =0	2.33	0.31	0.04	7.49	<0.001
q17f_INV_dummy [Subject] is not one of my strengths <4 = 0, 4 (very confident) =1	2.25	0.20	0.06	11.04	<0.001
STQ6_language_dummy Magar, Tharu, Rai, Limbu, else= 0, other = 1	1.87	0.20	0.05	9.28	<0.001
STQ10_dummy who helps you? mother, father, brother/ sister = 0, teacher, tuition, no one =1	1.53	0.20	0.04	7.48	<0.001
q9d_dummy Before or after school: I do jobs at home 2 hours or less per week = 1, other 0	1.48	0.21	0.04	6.99	<0.001
q29a_dummy Something of mine was stolen no = 1, yes =0	1.40	0.23	0.03	6.20	<0.001
Q4 Gender	-1.06	0.18	-0.03	-6.06	<0.001
Fathers_illiteracy_dummy Fathers education literate = 1, illiterate 0	1.26	0.24	0.03	5.35	<0.001
q9c_dummy Before or after school: I play sports, less than one hour per week = 1, other 0	1.12	0.24	0.02	4.60	<0.001
q9a_dummy Before or after school: I watch television and videos, not at all = 0, >0 =1	0.72	0.19	0.02	3.80	<0.001
home_accessories_dummy, 0 indicators (out of 5) met = 0, >0 = 1	0.61	0.20	0.02	3.04	0.002
SES_dummy, <3 indicators met (out of 7) = 0, other = 1	0.58	0.20	0.02	2.97	0.003

1) Method = Stepwise, Dependent Variable: P_Eqd_Total_shifted USE THIS (Mean-equated subjects)

Table 4.5.16 Checklist of the statistically the most effective factors explaining the average achievement level of the students

Variable[1]	gain when the lower value	gain when the higher value
I do homework: one hour or more per week = 1, less than one hour = 0	0	4.96
Absence: 4 days or less = 1, => 5 = 0	0	4.21
I work at a paid job: not at all = 1, other 0	0	3.91
I read a book for enjoyment: 0,1, 2 hours = 1, more than 2 = 0	0	2.35
I do jobs at home: 2 hours or less per week = 1, other 0	0	1.48
I play sports: less than one hour per week = 1, other 0	0	1.12
Attitude toward teacher: <1.5 (very positive or positive) = 1, =>1.5 = 0	0	2.77
Attitude toward Subject: <2 (extreme positive) = 1, =>2 = 0	0	2.60
I was hit or hurt by other student(s): no = 1, yes =0	0	3.29
I was hit or hurt by other student(s): no = 1, yes =0	0	2.33
Something of mine was stolen: no = 1, yes =0	0	1.40
SUM of the gains	0	30.4
Total score (expected)	15.5	43.9

1)   variables are grouped content-wise

On the basis of the data of around 50,000 students, two simple things may radically boost the learning outcomes: if the students were spending time on their *homework at least one hour per week* and they are *not absent from school*, they would score 10 percent point higher. Additionally, if the economic situation allowed the student *not to work in a paid capacity*, (s)he would *read books* (but less than 2 hours per week), did *home chores at home* (2 hours or less per week), and spent not too much, or no time at all *for sports*, (s)he would have scored another 9 percent points more. If the students were *positive toward the teacher and subject*, they would have scored an additional 5 percent point. Hence, if slightly moderating the study habits and attitudes the students would easily gain 24 percent points more. Another thing is that if the school could have made sure that there is *no bullying* within the school, it may be possible to raise the standard 7 percent points more. It is good to keep in mind that the estimations above are just mathematical averages. It is worth noting that the result does *not* mean that if the student just decides to reduce the number of absent days from 5 to 4 and the time spent on homework from 45 minutes to 60 minutes per week, the results would change radically. It is also good to remember that there are teacher- and school-related factors which may raise the achievement level of the students in the community schools (see Sections 4.3.3.2.2 and 4.4.3.3.2.2).

4.5.4 Conclusions

Several student-related variables were detected which explain, at least to some extent, the differences between the low and high performing students in the community schools. In summary, the main results related with the teachers are as follows:

- The school effect in Nepal is 68% and hence, something else (usually the rest is given to the students) explains 32% of the variance in the student data; the share for the students is not a very high value. The student effect is much higher in Mustang (91%) and in Kalikot (70%) where the test was administered so that the possibilities for the "social work" were minimized. (Section 4.5.3.1).

- It is significantly more probable to find a student aged 14 or lower within the highest performing students than of age 15 or higher. The highest performance is found with those groups of students studying within their normal age group, that is, at the age of 13 and 14 years. There is no difference in the achievement level between boys and girls in the lowest performing students. However, the boys seem to be better than girls within the group of highest performing students. (Section 4.5.3.2.1)

- The higher-performing students are more content with subject, teachers, and school and they sense more utility in the subject than the lowest performing students though the difference between the groups is not very wide; both groups feel, for example, the subject is equally useful for them. (Section 4.5.3.2.2)

- There is no difference in self-efficacy between the lowest and highest performing students; the analysis of self-efficacy should be interpreted with caution because the lowest performing students seemed not to understand the questions. (Section 4.5.3.2.3)

- A reasonable amount of time spent on homework given by the teacher, presence in school, and time *not* spent on paid work raises the achievement level remarkably. (Sections 4.5.3.2.4 and 4.5.3.3.2)

- There is a huge inequity between the language groups when it comes to the highest and lowest performing student groups: Rai, Limbu and Magar students are over-populated in the lowest performing student group and Urdu, Tamang, Newar, and Sherpa students in the highest performing student group. There is also a difference between the castes: Dalits, Alpasankhyaks, and Madhesis are over-populated in the lowest performing student group and Brahman and Cheetri students are overpopulated in the highest performing group. (Section 4.5.3.2.5)

389

- The socioeconomic status (SES) plays a strong role in the educational processes in Nepal, not only in the private schools but also in the community schools. Within the community schools, the difference in achievement between the students from the lowest and highest SES groups is remarkable (22–25 percent points). This means that if the social economic standard of the lowest performing students be raised to a decent level, that is, in practice, that the problems in four out of seven indicators would be solved, the results in these groups would rise remarkably. Especially challenging is the situation in the families where both parents are illiterate or they both work in agriculture. (Sections 4.5.3.2.6)

- When the children have very few home possessions (like a table to work, a dictionary and so on) helping them in the studies, the achievement level is statistically lower (39–42%) than if there are more than four out of 12 indicators met (> 46%). With ten to twelve possessions, the average score is very high (> 52%) compared with the national average. The same is true of home accessories (mobile phones, bathrooms and so on): With no accessory indicator met, the results are very low (43%) and when there are three or more met out of five, the results are remarkably higher (> 52%). With four or five indicators met, the results are the very high (56–58%). (Sections 4.5.3.2.7)

- There are more occurrences of bullying within the lowest performing students; too often they are excluded from the actions and they are made to do things that they did not want to. (Sections 4.5.3.2.8)

References for Section 4.5

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

Gafoor, K. A. & Ashraf, P. M. (2012). Contextual Influences on Sources of Academic Self-Efficacy: A Validation with Secondary School Students of Kerala. *Asia Pacific Education Review*, **13**(4), 607–616.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika,* **73**, 43–56.

Holen, S., Lervag, A., Waaktaar, T. & Ystgaard, M. (2012), Exploring the Associations between Coping Patterns for Everyday Stressors and Mental Health in Young Schoolchildren. *Journal of School Psychology*, **50**(2), 167–193.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics,* **29**(2), 119–127.

Kiran, D. & Sungur, S. (2012). Middle School Students' Science Self-Efficacy and Its Sources: Examination of Gender Difference. *Journal of Science Education and Technology*, **21**(5), 619–630.

Lennon, P. (Ed.) (2012). Learner Autonomy in the English Classroom: Empirical Studies and Ideas for Teachers. Frankfurt: Peter Lang.

MacCann, C., Lipnevich, A. A., Burrus, J. & Roberts, R. D. (2012). The Best Years of Our Lives? Coping with Stress Predicts School Grades, Life Satisfaction, and Feelings about High School. *Learning and Individual Differences*, **22**(2), 235–241.

Metsämuuronen, J. 1995. *Interests and independent learning. Commitment, motivation, and coping*. Doctoral Thesis. Helsingin yliopiston opettajankoulutuslaitos. Tutkimuksia 146. Tummavuoren Kirjapaino Oy, Vantaa. [In Finnish].

Metsämuuronen J 2009. *Methods assisting the assessment*. Oppimistulosten arviointi 1/2009. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

Metsämuuronen, J. (2012a). Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies,* **4**(3) September.

Metsämuuronen, J. (2012b). Comparison of Mental Structures of Eighth-Graders in Different Countries on the basis of Fennema-Sherman test**.** *International Journal of Psychological Studies,* **4**(4) December, 1–17.

Monks, C. P., Robinson, S. & Worlidge, P. (2012). The Emergence of Cyberbullying: A Survey of Primary School Pupils' Perceptions and Experiences. *School Psychology International*, **33**(5), 477–491.

Park, Y. (2011). How Motivational Constructs Interact to Predict Elementary Students' Reading Performance: Examples from Attitudes and Self-Efficacy in Reading. *Learning and Individual Differences*, **21**(4), 347–358.

Phan, H. P. (2012). The Development of English and Mathematics Self-Efficacy: A Latent Growth Curve Analysis. *Journal of Educational Research*, **105**(3), 196–209.

Postholm, M. B. (2010). Self-Regulated Pupils in Teaching: Teachers' Experiences. *Teachers and Teaching: Theory and Practice*, **16**(4), 491–505.

Roebers, C. M., Cimeli, P., Rothlisberger, M. & Neuenschwander, R. (2012). Executive Functioning, Metacognition, and Self-Perceived Competence in Elementary School Children: An Explorative Study on Their Interrelations and Their Role for School Achievement. *Metacognition and Learning*, **7**(3), 151–173.

Shamir B (1988). Commitment and leisure. *Sociological perspectives*, **31**(2), 238–258.

Shin, H. & Ryan, A. M. (2012). How Do Young Adolescents Cope with Social Problems? An Examination of Social Goals, Coping with Friends, and Social Adjustment. *Journal of Early Adolescence*, **32**(6), 851–875.

Tavakolizadeh, J. & Qavam, S. E. (2011). Effect of Teaching of Self-Regulated Learning Strategies on Attribution Styles in Students. *Electronic Journal of Research in Educational Psychology*, **9**(3), 1087–1102.

Yeung, A. S., Lau, S. & Nie, Y. (2011). Primary and Secondary Students' Motivation in Learning English: Grade and Gender Differences. *Contemporary Educational Psychology*, **36**(3), 246–256.

Vandevelde, S., Van Keer, H. & De Wever, B. (2011). Exploring the Impact of Student Tutoring on At-Risk Fifth and Sixth Graders' Self-Regulated Learning. *Learning and Individual Differences*, **21**(4), 419–425.

What Works (2012a). *Peer-Assisted Learning Strategies*. What Works Clearinghouse Intervention Report January 2012. What Works Clearinghouse (ED), Institute of Education Sciences. Available at http://www.eric.ed.gov/PDFS/ED531596.pdf.

What Works (2012b). *Peer-Assisted Learning/Literacy Strategies*. What Works Clearinghouse Intervention Report May 2012. What Works Clearinghouse (ED), Institute of Education Sciences. Available at http://www.eric.ed.gov/PDFS/ED531596.pdf.

4.6 Setting Standard in Nepali – CEFR Levels

Dr. Jari Metsämuuronen[1],
Mr. Shyam Prasad Acharya[2] &
Mr. Hari Arial[2]

1. Helsinki University, Faculty of Behavioral Sciences
2. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal

Abstract

This article introduces two methodological solutions to find out how good readers and writers the 8th grader students really are. The first solution is one of the widely used systems for setting standards, the Common European Framework of Reference for Languages (CEFR). It was selected for the basis of the standard setting in Nepal because the procedures and standards are well-described in the literature and the levels are transformable to other standards. Within the different systemics, there are several methods for the standard setting. A new method used in Finland, the three-phased Theory-based and Test-centered method for the Wide range of proficiency levels (3TTW, Metsämuuronen, 2013), is introduced and its application in Nepal is described.

An average 8th grader reader of Nepali is at the CEFR level B1.1 which means that (s)he "*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. [The student c]an follow the main points, key words and important details of a few pages of text dealing with a familiar topic*." The level is not very high. The level means that an average 8th grader reader cannot read and understand newspapers independently, for example. The risk of being a weak reader is high in language groups of Tamangs, Tharus, Newars, and Limbus and especially amongst the Madhesi population.

An average 8th grader writer of Nepali is at the CEFR level A2.2 which means that (s)he "*can manage in routine everyday situations in writing. [The student c]an 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).*" The level is somewhat lower than that of reading even though the writing skills are highly focused in the Nepali curriculum. The level means that an average 8th grader writer cannot make, for example,

lecture notes or brief summaries independently. 8% of the students are very weak in writing – the risk is very high in the language groups of Tharus, Limbus, and Gurungs and especially in Madhesi societies.

Keywords: Language Proficiency, Reading skills, Writing skills, CEFR-levels

4.6.1. Introduction

On the basis of National Assessment of Student Achievement (NASA) in Nepali 2011, 30.4% of the 8th graders speak something other than Nepali as their first language. These "other" languages are quite fragmented; the largest groups in the student dataset are Magars (3.2%), Tamangs (3.1%), and Tharus (2.2%). After dividing the languages into ten groups excluding Nepali, there were still 18.9% of the students classified into the group "else". For all of these non-Nepali students, Nepali is learned as a second language (2L). From the point of view of national cohesion it is important to know what the real language proficiency levels of the students are; with the low level of Nepali language competence, it is, for example, difficult to think that a student would reach any study place in further education.

The state of the art assessment of language proficiency includes two different approaches. The traditional approach is to use the *total scores* (or percentages of maximum scores) or *standard setting* in assessment (or descriptions of proficiency levels). These two approaches are not strictly comparable from several viewpoints, one of which is that the first is based on norm-referenced testing and the latter on criterion-based testing. In practice, the norm-referenced testing means that the final test score produces a *norm* with which the different groups (such as geographical areas or sexes) can be compared. Hence, one may get to know that in a certain geographical area the results are *better* than in another area. However, with norm-referenced testing one does not know how *good* the pupils in fact are, that is, what the real proficiency level is. The criterion-based testing, in contrast, uses *external criteria* based on the standard setting procedures for language proficiency. Some of the well-known standards are the Common European Framework of Reference for Languages (CEFR), TOEFL, Cambridge Exam, or IELTS. The CEFR was selected for the basis of the standard setting in Nepal because the procedures and standards are well-described in the literature (for example, in Takala, 2009; Kaftandjieva, 2004; van der Schoot, 2009; FNBE, 2004) and the levels are transformable into other standards (see, http://en.wikipedia.org/wiki/ Common_European Framework_of_Reference_for_Languages). Thus, one advantage of the CEFR classification is that there is a connection of CEFR standards with other standards such as TOEFL.

This article introduces some challenges in standard setting from the national student achievement testing viewpoint and introduces some traditional methods in standard setting in Sections 4.6.1.1 and 4.6.1.2, discusses the CEFR levels and their modifications in Section 4.6.2, introduces the method used in Toropainen (2010; Metsämuuronen, 2009; 2010) called the three-phased, theory-based, and test-centered method for the wide range of proficiency levels (3TTW, Metsämuuronen, 2013) in Section 4.6.3, describes its modification in the assessment of Nepali proficiency in Section 4.6.4, and reports on Section 4.5.5 what the Nepali students' internationally comparable proficiency levels in reading and writing are. An interested reader finds the basic results of the language proficiency in Section 3.2 of this book. The methodological part is based mainly on Metsämuuronen (2013) where the method itself is originally described. Because the readers of this article are most probably unaware of the article, relevant parts of the original article are presented also here.

4.6.1.1 Challenge in setting standard in the national student achievement test

The traditional approach and standard setting approaches can be combined though the output is not always credible. As an example, in the assessment done by the Finnish National Board of Education (FNBE) of Swedish as 2L for the native Finnish speakers (Tuokko, 2009), the classical approach was used in assessing the receptive skills (Listening and Reading) and the CEFR levels were used in assessing the productive skills (Writing and Speaking). At the post phase, the total score of the receptive test was transformed into the CEFR levels (Takala & Kaftandjieva, 2009). As an outcome, the credibly (normally) distributed skill level in the sample of over 5,000 students was distorted to be some kind of Bactrian camel type of distribution with two hunches. Obviously, the transformation was not given credibility among the reders of the text though both distributions were kept in the report; the main analysis was done on the basis of the percentages of correct answers. The lesson learnt was that when the aim is to set standards for a test of a wide range of proficiency levels, the strict transformation of the total score does not necessary lead to a credible outcome.

In a later project of Finnish as a 2L for the native Swedish speakers (Toropainen, 2010), a new method – somewhat the same as introduced in this article – was created and tested to find the proficiency levels without strictly transforming the total score to CEFR levels (Metsämuuronen, 2009; 2010). The sample of 1,700 Swedish speaking students learning Finnish as a second language showed a radically non-normal distribution. The reason for the distribution was that the achievement levels in two different areas of Swedish speaking inhabitants were radically different. However, as the method used managed to find the distributions; the classification was treated credible.

4.6.1.2. Factors discriminating the methods used in standard setting

There are tens of methods for the standard setting. Kaftandjieva (2004) compares 34 different procedures but, after compiling from different sources, she estimates that there are more than 50 methods and many of those have several modifications (Kaftandjieva, 2004, 11). These methods can be separate from each other in numerous ways. Here three criteria are handled: 1) how the proficiency levels are defined; theory- or empirically-based, 2) what is the orientation of the method; examinee- or test-centered, and 3) whether the method is suitable for classifying a wide range of proficiency levels. The last is an important question from the national assessment point of view: as it is necessary to be able to assess the proficiency levels of the students with one attempt across several levels and not only one or two levels which is the usual procedure in language testing settings.

4.6.1.2.1 Theory- or empirical-based determination of proficiency level

The first factor discriminating the methods for standard setting is how the boundaries of the proficiency levels are defined. In the older methods such as the Nedelsky method (Nedelsky, 1954; Livingstone & Zieky, 1982) or the Tucker-Angoff method (Angoff, 1971) the items can be defined to belong to certain proficiency levels without any empirical knowledge of the difficulty levels of the items. This is possible when the criteria are well formulated – as in CEFR levels they are. Then it is taken as a fact that an item can be easy or more difficult and still measure a certain kind of reading ability – such as picking a piece of information from a short text. Then the context of the assignment strictly defines the proficiency level of an item; "extracting information from a postcard" type of assignment cannot be classified into a higher level of proficiency even though it would be more difficult in the sample than some other item strictly defined at a higher level. In the more recent methods as in the Item-descriptor Matching Method (Ferrara *et al.,* 2002), the Basket Method (Alderson, 2005), and the Bookmark Method (Mitzel *et al.*, 2001) the items are classified based on the theory but, as additional information for the judges, also the difficulty level acquired from the Item Response Theory (IRT) is given. Hence in a test of gradually increasing difficulty level of items the proficiency level of the illogically classified items can be changed when the difficulty level of the item is known on the basis of the empirical data.

Primarily, the empirical-based methods are based on changing the total score into the proficiency levels on the basis of empirical evidence (see Takala 2009, 58; Kaftandjieva 2004, 1, Takala & Kaftandjieva 2009). The terms '*Cut-off Score*' or '*Cut Score*' are used to refer to the specific value of the total score which is used to classify the examinee to different levels. In the traditional methods, the judges are asked to think on the basis of 100 hypothetic '*Borderline Person*' (or '*Minimally*

Accepted Person', 'Just Barely Passing' or *'Minimally Competent Candidate'*)
how many of those at the lower proficiency level would pass an individual item.
Each panel member gives his/her estimation and each figure will be divided by
100. Hence, a panel may have given a judgment that in order to fulfill the requirement
of being at the level A2 the examinee should have an average of 16.54 points of
the total score, that is, either 16 or 17 points. If the purpose is to use the test to
measure several proficiency levels at the same time – as is the aim in the national
assessment testing, the same hypothetical question has to be asked at all levels in
the test for all the items. Another set of empirically-based – maybe more modern
– procedures in standard setting is to use Item Response Theory (IRT) modeling
in defining the boundaries for the proficiency levels. In these methods (such as in
the Bookmark Method), IRT modeling is used to define the theoretical response
probability which is required of those examinees who are expected to be at a
certain level. When the experts have classified the items at different proficiency
levels, the success in an item can be defined, for example, so that 67% success
probability is required, that is, that two thirds of the borderline examinees give the
correct answer in the item. The response probability is not unanimous: is can be
set to 50–80%.

Because the empirically-based methods are (usually) based on using the
total score as the basis of the standards, they technically embed a challenge called
"compensation" (Takala 2009, 84) even though it is not discussed often. Namely,
the total score is formed by summing up all the items – also those which are
classified to a totally different proficiency level. Hence, the examinee can
compensate sleepiness (or ignorance) in some lower level items by giving partially
(or fully) correct answer in the higher level items. The examinee may then, for
technical reasons, be classified (undeservedly) at B1 level even though (s)he has
not passed an adequate number of B1 items but has compensated for those by
guessing (or knowing) some higher level items and gained a score high enough to
be rated as being at level B1. Though in some borderline cases, the procedures
may lead to significantly wrong classifications, it may be a justified principle.
However, in some cases – as described in Section 4.6.1 – the procedure may lead
to radically non-logical transformations.

4.6.1.2.2 Test- or examinee-orientation of the method

Another factor discriminating the methods for standard setting is what the
orientation of the method is: test-centered (*Test-centered continuum methods*)
or examinee-centered (*Examinee-centered continuum models*). Jaeger (1989,
493), Kaftandjieva (2004, 12), and Takala (2009, 60) value this factor as the first
and maybe the most central discriminating factor of the standard setting methods.

In the test-centered procedures, as in the Nedelsky Method and its variations (e.g. Reckase, 2000), The Tucker-Angoff Method and its variations (e.g. Impara & Plake, 1997; 1998; Loomis & Bourque, 2001), the Item-descriptor Matching Method, the Basket Method, and the Bookmark Method the basis of the procedure are the individual items and their classification into certain proficiency levels (see Section 1.2.1). Kaftandjieva (2004, 14) estimates that approximately 70% of the methods belong to these test-based methods.

In the examinee-centered methods – as in the Contrasting Groups Method (Reckase, 2000; Brandon, 2002), the Borderline Group Method (Livingstone & Ziegy, 1982), or the Body of Work Method (Kingston *et al.*, 2001) – the basis is on how well the examinee manage to solve the individual items and on how well they rate in the total test. Usually in these methods, it is important that at least one of the judges knows the examinee well and this judge can classify the examinee to correct proficiency level because of his/her experiences.

4.6.1.2.3 Applicability to classify wide range of proficiency levels

From the national assessment viewpoint, a third relevant factor in the evaluation of the standard setting methods is whether the method is suitable for classifying the examinee to a wide range of proficiency levels. The CEFR manual (Takala 2009, 63, 65) is openly sceptical of using the same test for assessing several levels in one go: "*It is an illusion to think it is possible to build a test and to set standards for the six basic levels of the CEFR (A1 to C2) within the same test or examination by using test-centered standard setting methods*." The experiences in the national student assessment in Finland (Tuokko, 2009; Toropainen, 2010) show, however, that it actually *is* possible to use the same test for assessing several CEFR levels with one test. The experience in Tuokko's project showed that a traditional method of transforming the total score to CEFR levels was not successful in the case of wide levels of proficiency in the test. In contrast, Toropainen's project showed that the new method was successful in the task.

The general idea in the classical methods is that the test measures specific proficiency level or levels very near to each other. Hence, the methods are optimal within a *narrow* scale of proficiency levels. However, the reality in the national assessment of languages – compared with the passing and failing of tests of a certain level – is that there are usually thousands of the students who should be tested within a limited time frame and practically with one test in order to acquire information on the national distribution of the proficiency levels. Then, the national student assessment requires tests with a *wide* scale of proficiency levels. This gives a specific challenge to the final classification of the examinee; in order to

assess credibly and covering several proficiency levels, the measurement instrument should include enough items in each proficiency level though not being too long to make the students exhausted. This challenge does not differ from that in the Mathematics testing, for example; in a national test of 30 items there may be only five items from geometry and still the proficiency level in geometry is reported without hesitation. Similarly, with five or even with only four items clearly falling into the level A2 this may be enough to assess whether the test takers are at this level or not. The logic of this is discussed in Section 3 based on the CEFR classification. Because the CEFR classification is most probably not familiar to the reader it is handled first in the next section.

4.6.2. CEFR classification

4.6.2.1 Traditional CEFR

In CEFR classification, the original set of proficiency levels are fixed to six levels: Breakthrough or beginner (A1), Waystage or elementary (A2), Threshold or intermediate (B1), Vantage or upper intermediate (B2), Effective Operational Proficiency or advanced (C1), and Mastery or proficiency (C2) (Table 4.6.1). The contents differ in different areas of language.

Table 4.6.1 Traditional CEFR levels

CEFR level	Short Description	Condensed Content
A1	Breakthrough or beginner	Limited communication in the most familiar situation
A2	Waystage or elementary	Basic needs for immediate social interaction and brief narration
B1	Threshold or intermediate	Dealing with everyday life
B2	Vantage or upper intermediate	Managing regular interaction with native speaker
C1	Effective Operational Proficiency or advanced	Managing in a variety of demanding language use situations
C2	Mastery or proficiency	Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in the most complex situations.

4.6.2.2 CEFR levels in the Finnish core curriculum

In Finland, before the FNBE started to use the CEFR levels in the core curriculum of the student assessment of languages (FNBE, 2004), it was noticed that six levels were not a fruitful basis for student assessment in schools. Hence, the national

experts of CEFR levels divided the classification into more precise levels which are now used in teaching and student assessment in Finnish schools – the same levels are also used in what follows at the Nepali language proficiency levels. The levels are divided as in Table 4.6.2.

Table 4.6.2. CEFR levels used in the Finnish compulsory education (FNBE 2004)

CEFR level	Short Description
A1.1	First stage of elementary proficiency
A1.2	Developing elementary proficiency
A1.3	Functional elementary proficiency
A2.1	First stage of basic proficiency
A2.2	Developing basic proficiency
B1.1	Functional basic proficiency
B1.2	Fluent basic proficiency
B2.1	First stage of independent proficiency
B2.2	Functional independent proficiency
C1.1	First stage of fluent proficiency

The levels higher than C1.1 are not defined in the Finnish system because it is not expected for anyone to reach fluency in the foreign language acquisition within school years. The condensed, though more detailed, descriptions of the contents for the topics of reading and writing are seen on Table 4.6.4 in what follows.

In the Finnish system, the criteria for a "good" performance at the end of the compulsory education (9th grade) are set differently in different second languages and in different topics (Table 4.6.3). In English, the expected level is higher than in the other languages and more skills are required in the receptive topics (Listening and Reading) than in the productive ones (Speaking and Writing). Note that no criteria are set for native speakers – the criteria are usually applied for 2L students only.

Table 4.6.3 Levels for "good" performance in the Finnish educational system (FNBE, 2004, 142)

Language	Listening	Speaking	Reading	Writing
English	B1.1 Functional b.p.[1]	A2.2 Developing b.p.	B1.1 Functional b.p.	A2.2 Developing b.p.
Other languages	A2.2 Developing b.p.	A2.1 First stage of b.p.	A2.2 Developing b.p.	A2.1 First stage of b.p.

1) b.p. = basic proficiency

Table 4.6.4 Abridged application of the CEFR levels in the FNBE (FNBE 2004)

CEFR level	Reading comprehension	Writing comprehension
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).
A2.1	• 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 most 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 nformation 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).
B1.1	• 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 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).	Can read a few paragraphs of text about many different topics (newspaper articles, brochures, user instructions, simple literature) and can also 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.
B2.1	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 summarise information from different sources in his/her own texts.	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.
B2.2	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, analyse, reflect on and summarise information and thoughts.	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 summarise or paraphrase the main points.
C1	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 organisational means and cohesive devices.	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, summarise these and draw demanding conclusions from these.
C2	(no description in FNBE)	Can understand with ease virtually everything heard or read. Can summarise 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.

4.6.3. 3 TTW – Three-phased theory-based and test-centered method for the Wide range of proficiency levels

In the Three-phased Theory-based and Test-centered method for the Wide range of proficiency levels (3TTW, Metsämuuronen, 2013), based on 2TTW (Metsämuuronen, 2009b; 2010), the standard setting is done in three subsequent phases as the name indicates. At the first phase, the items are classified on the theoretical bases into "baskets" of proficiency levels required to solve the problem. At the second phase, the examinees are classified into theoretically sound levels on the basis of a wide range of levels in the test; this is done on the basis of theoretical classification and empirical data. At the third phase, the classification is adjusted on the basis of the IRT modeling of the proficiency levels of the examinees in the empirical data and by utilizing both the theoretical classification of the examinee and the original distribution as the reference points.

4.6.3.1 First phase – Classification of the items on the basis of theory

3TTW can be classified as one of the theory-based methods such as the classic Nedelsky-, Tucker-Angoff-, or Aldersson methods are (see Section 1.2.1). The two-phased characteristic of the method comes from the fact that, at the first phase, the experts classify the items into certain proficiency levels on the basis of theory – in this case the theory comes from CEFR levels and specifically from the CEFR levels as defined in the core curriculum of FNBE. This classification can be done without knowing anything of the proficiency levels of the examinee or of the difficulty level of the items. The logic is that when the "theory" (see Table 4) says that at the level A1.3 the examinee "*can locate specific information required in a short text (postcards, weather forecasts)*" and the item is specifically written so that there is a short postcard text where the examinee is asked to locate a simple piece of fact or information, the level of the item is A1.3 even though it would be difficult or easy to the wide group of examinee. At the first phase, the theoretical framework and the context of the task strictly guide the experts regardless of what would be the difficulty level of the item on the empirical dataset.

4.6.3.2 Second phase – Classification of the examinee on the basis of theory and the empirical data

At the second phase, the examinees are classified into certain proficiency levels on the basis of their empirical achievement in the sub-tests. Hence, 3TTW can be classified as Test-centered methods – as 70% of the methods are (see Section 4.6.1.2.2). However, what differentiates 3TTW from the classical methods is the characteristic of taking into account the wide range of proficiency levels. Namely,

it is usual in the test-centered methods for the *total scores* to be changed to proficiency levels. This works credibly when all the items are measuring the same proficiency level, but when the aim is to cover a wide range of proficiency levels in one attempt, the transformation of the total score does not necessarily lead to a credible result as discussed in Section 4.6.1.1. In 3TTW, the second phase combines the theory-based classification of items and the empirically adjusted-based difficulty level of the examinee to a unique feature.

When, at the first phase, the required achievement level in each item or the 'proficiency level of the items', is known on the basis of "theory", at the second phase the levels of the examinee are assessed.[74] In the simplest and most optimal situation the logic goes in the following three steps:

1. Calculate the sum of the items on each proficiency level. Hence, the subscore of items reflecting the level A1.3 are summed up and the result is "the score of proficiency level A1.3", for example. Similarly, when the items reflecting the level A2.1 are summed up it results in "the score of proficiency level A2.1".

2. Decide how many percentages of the maximum score of each level have to be passed in order to be classified (at least) to the specific level. Presumably, the cut-off of less than 50% should be argued for.[75] The boundary can be set to 50–80% as in the classic methods (see Chapter 2.1). In practice then, when the examinee's score is higher than 50% of the maximum score of the A1.3 level items, (s)he has shown proficiency at this level. The real achievement level can be much higher. The highest level of achievement is assessed at the third step.

3. The profile of the examinee is assessed as a whole to set the proficiency level of the examinee. The proficiency level is the highest consecutive level that the examinee has passed credibly on the basis of the bound-

---

[74] Logically, this should be called the 'difficulty level of the items'. However, the logic is the same as in the famous presentation of Wright (1968) where he showed that in the IRT modeling (or Rasch modeling) the ability level is identical with the difficulty level of the item when certain preconditions are met (of the derivation, see also Metsämuuronen, 2013, 171–173).

[75] A lower than 50% cutoff may be justified when – as a tradition in Nepal is – the markers are not willing to give the highest marks except for the exceptionally good examinees. Practically then, in one long essay type of item with 10 points as maximum (objectively marked) the markers at the pretest phase did not give the marks 10 or 9 at all. Hence for the final test, the maximum score was lowered to 8. As a consequence, at the final test the markers did not give any marks of 8 and very rarely 7. Obviously, this is a nuisance for the IRT modeling used in item calibration and test equation because each value between the minimum and maximum score has to be observed. Otherwise the estimation cannot be done. Thus, the technical maximum was lowered to seven marks and the final boundary for passing the level was lowered to just below 50%.

aries set in step 2. In the simplest and theoretically most solid case, when the examinee has passed both levels A1.3 and A2.1 with more than 50% of the scores but not levels higher than this, the proficiency level of the examinee is A2.1 even though (s)he might have some random correct answers in the items showing higher skills than A2.1.

Passing and failing at a level can be denoted by '1' and '0'. As an output, three-step procedure leads to the options seen in Table 6 assuming that the levels A1.3 to B2.2 are of interest in the test.

Table 4.6.6. Theoretically expected profiles of the examinee in different proficiency levels

A1.3	A2.1	A2.2	B1.1	B1.2	B2.1	B2.2	Conclusion
0	0	0	0	0	0	0	lower than A1.3
1	0	0	0	0	0	0	A1.3
1	1	0	0	0	0	0	A2.1
1	1	1	0	0	0	0	A2.2
1	1	1	1	0	0	0	B1.1
1	1	1	1	1	0	0	B1.2
1	1	1	1	1	1	0	B2.1
1	1	1	1	1	1	1	B2.2 or higher

The theoretical procedure with three-steps tells the logic of the method but it works optimally only when 1) the proficiency levels of the items have been set correctly or at least credibly, 2) there are a sufficient number of items at all levels used in the test, 3) the original markings are correct, 4) the cut-off boundaries are set meaningfully (> 50% in the example), and 5) the examinee does not do several carelessness mistakes at the lower level items. If there are inadequacies in the points above, it may lead to a lowered validity of the classification. In practical situations, it seems that only some – in some cases most – cases can be classified to these theoretically sound levels. These theoretically sound classifications are adjusted by empirical data.

4.6.3.3 Third phase – Adjusting the classification on the basis of empirical data and IRT modeling

One of the challenges in the theoretical classification is that the items within a proficiency level can be easy or difficult. For example, an item clearly classified into level A1.3 – "finding a single piece of information from a post card", for

instance – is easier when the information in the question is in the same format as in the text itself (numbers or text) and more difficult when they are different. Similarly, extracting a piece of information from a timetable is a much easier task when there are only a few numbers than when there are lots of numbers. This leads to a challenge in that the examinee may be able to solve several easy tasks from the upper level items but actually cannot show solid proficiency at any of the higher proficiency levels – or at any levels. The challenge is to classify these examinees into a credible proficiency level. In order to adjust the original classification, the thirst phase combines the theoretical classification of items found at the first phase, examinees profiles found in the second phase, the empirical achievement levels of the students (Theta parameter found by the IRT modeling), and the original marginal distribution of the Theta values.

In Table 4.6.7, real-life data from the version 1 of Nepali data shows three dilemmas in combining the theoretical classification and latent Theta values. First, out of 8,023 examinees 1,521 (19%) were not behaving logically in the classification and hence they are in the crosstable.[76] Second, at each theoretical CEFR level, one recognizes a Normal distribution of achievements. Especially wide is the distribution of those classified to be lower than level A1.3. This means that there are over 600 examinees who did not reach the required 50% correct in any of the CEFR levels but, on the basis of their achievement level (Theta), most of them could have been at the level A1.3 or somewhat higher.[77] Third, at each Theta level or each value of the total score there seems to be several possibilities for the best fit with the CEFR levels. For example, at the Theta level -0.5, the highest frequencies are at the CEFR levels B1.2 (244 cases) and B1.1 (207 cases). Hence, it is not unanimous how the Theta values should be combined with the theoretically formed CEFR levels.

An additional note on the adjustment comes from the original marginal distribution (Figure 4.6.1). A very strong assumption is that the language proficiency in population is distributed more or less normally – not as in the case seen in Table 4.6.7, where the distribution is skewed to the right because of the many very easy items on the test. It is essential that the distribution of proficiency levels has to approximate to the shape of the marginal distribution. Otherwise the classification has not found the main feature of the population.

In Table 4.6.7, two possible alternatives are given to transforming the Theta values to CEFR levels on the basis of the empirical data. Those with the gray

---

[76]   In Version 2, somewhat 50% of the cases showed non-theoretical patterns.
[77]   It is good to remember that Theta and the final score are strictly bounded to each other. Hence, estimating Theta, a compensative approach is used; where a mistake in the lower level item is compensated by knowing or guessing a few easy items from the higher level. This means that Theta is not necessarily comparable with the proficiency levels as it is defined in 3TTW.

shading are based mainly on the highest frequencies of the theoretical classification and those with the bold characters are based on combining the frequencies, the credibly classified highest levels (B2.1, B1.2, and B1.1), and the shape of the distribution of the marginal frequencies. From Figures 4.6.1 and 4.6.2, one notes that the suggestion based solely on the highest frequencies (the gray shading) leads to a clearly deviant distribution compared with the original marginal distribution. The latter suggestion (the bold characters) leads to a far better distribution. However, in the second option, one loses the connection between the Theta levels of the examinees and the original CEFR level of the students at the lowest achievement level. This is not a crucial deficiency because Theta – in any case – utilizes the compensation approach which is not adequate to the 3TTW method.

Table 4.6.7 A Cross-table of theoretical classification and achievement level (Theta) in the data

	CEFR levels in Reading – theoretical classification							Count	suggestion 1:[1]	suggestion 2:[2]
Theta	<A1.3	A1.3	A2.1	A2.2	B1.1	B1.2	B2.1	(n)	Gray shade	Bold font
-5.5	22	0	0	0	0	0	0	22	<A1.3	<A1.3
-4.6	42	0	0	0	0	0	0	42	<A1.3	<A1.3
-3.7	82	15	2	0	0	0	0	99	<A1.3	<A1.3
-3.1	102	59	5	0	0	0	0	171	<A1.3	A1.3
-2.7	117	93	15	2	0	0	0	239	<A1.3	A1.3
-2.3	107	176	26	21	0	0	0	374	A1.3	A2.1
-1.9	104	237	55	44	2	0	0	544	A1.3	A2.1
-1.5	60	262	87	85	4	0	0	693	A1.3	A2.2
-1.2	14	191	116	137	28	19	0	763	A2.1	A2.2
-0.8	9	70	152	189	66	73	0	923	A2.2	B1.1
-0.5	0	9	51	98	207	244	2	901	B1.1	B1.1
-0.1	0	0	15	83	207	438	4	927	B1.1	B1.1
0.3	0	0	0	4	168	518	36	790	B1.2	B1.2
0.7	0	0	0	0	68	550	79	704	B1.2	B1.2
1.3	0	0	0	0	8	411	133	552	B2.1 or higher	B2.1 or higher
2.1	0	0	0	0	0	131	97	228	B2.1 or higher	B2.1 or higher
2.9	0	0	0	0	0	0	51	51	B2.1 or higher	B2.1 or higher
total	659	1,112	524	663	758	2,384	402	8,023		

[1] based on the adjustment of the highest frequencies in the theoretical classification
[2] based on the credible classification of the highest classes and the distribution of the marginal frequencies
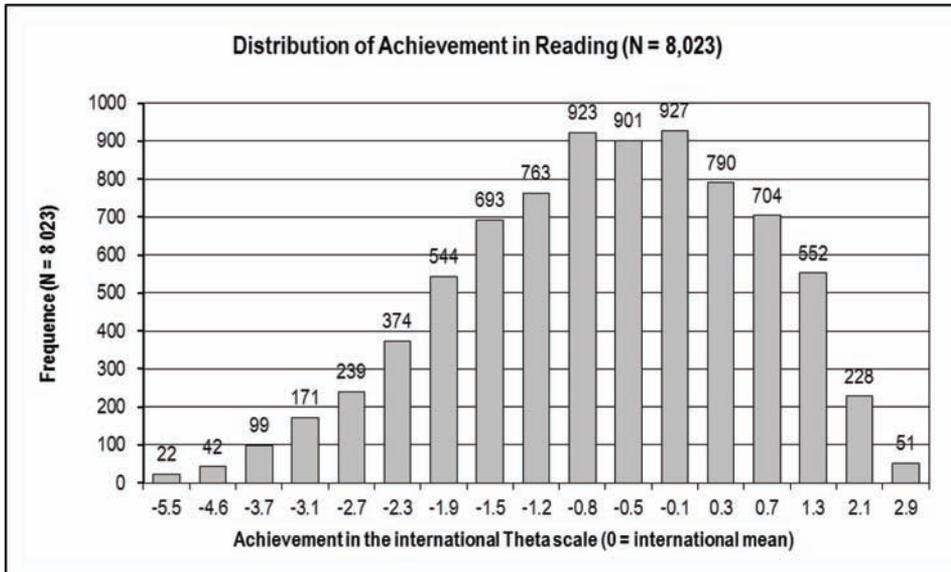
Figure 4.6.1 Marginal distribution of Theta values in the original metric
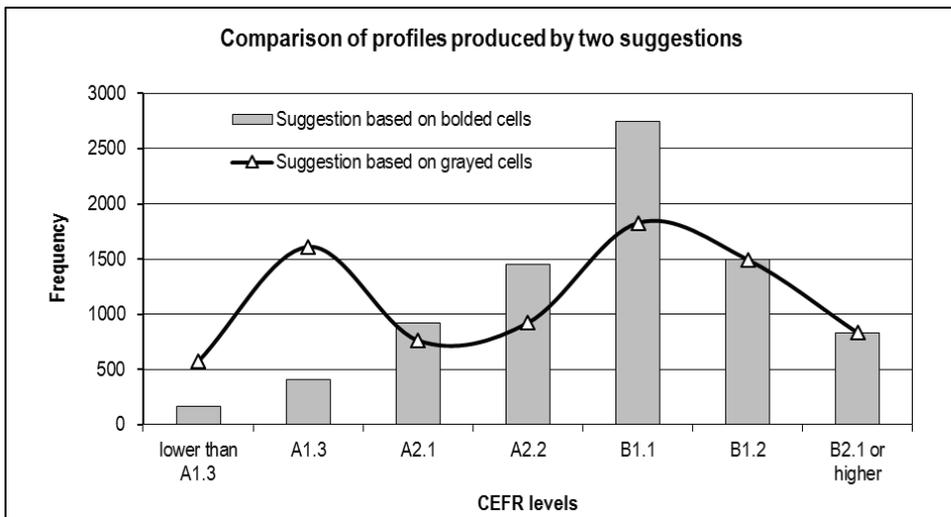
Figure 4.6.2 Two alternative transformations of the Theta values to CEFR levels on the basis of Table 4.6.7

4.6.4 Procedure of 3TTW in Nepal

4.6.4.1 Starting the process

Using the standards prepared for mastering the *foreign* language (not the native language) makes sense in Nepal because, according to the latest census in 2011, less than 60% of the students natively speaks the Nepali language though it is taught as the primary language in all schools. Another motivation for using the classification comes from the fact that in the pretest there were schools where all the students were given zero marks in the writing tasks; supposedly the achievement level in many non-Nepali speaking schools is so low that it makes sense to test how *low* the level really is. In these cases, the standard essay types of assignments are far too demanding for the students. Thus, simple and short writing assignments were created for the test.

A small scale workshop of standard setting was organized during the test construction for the selected item writers and officers in ERO. The content of the workshop was to internalize the standards developed for student assessment according to CEFR levels in FNBE – the standards are relatively explicit and they were quite easy to apply in writing and reading items (see Table 4). On the basis of the workshop, the reading and writing items were initially classified on the basis of this "theory", that is, on the basis of the description of the FNBE for certain levels. In the classification, it is noteworthy that the assignment itself restricts the level of achievement shown. For example, when needing to read a simple text in a postcard, it is not possible to show much higher achievement than A1.3 or A2.1 even though the reader's achievement level would be much higher. Also the Grammar and Vocabulary items were classified on the basis of the final IRT parameters as is the usual manner in the modern standard setting procedures known as the *Item-descriptor Maching Method* (Ferrara *et al.* 2002), the *Basket Method* (Alderson 2005) or the *Bookmark Method* (Mitzel *et al.* 2001; van der Schoot, 2009). However, these items were not used in the standard setting at the final phase; only Reading and Writing were analyzed.

The reading and writing items with only one mark were quite easy to categorize. For example, a reading item of picking a simple fact from an easy post card text is classified to level A1.3 because of the description in Table 4: "*Can locate specific information required in a short text (postcards, weather forecasts)*." On the other hand, when the assignment is constructed so that several marks can be given and the assignment is an essay type in nature, it is relevant to think that the highest marks on the item show a much higher achievement level than the lowest marks on the same item. For example, the classical essay type of assignment of a common topic may correlate to the basic categorization of B2.2 because of the description in Table 4: "*Can write an essay… [Can] express his/ her point of view, develop arguments systematically, analyse, reflect on and*

summarise information and thoughts." This kind of writing can be given 5 to 10 points. In Nepali, seven points were given as a maximum. However, when seven marks are given as the maximum point in a long essay type of item, there is no objective criterion for knowing which CEFR level is shown with 7 marks or with 1 mark, even though it is reasonable to assume that the seven-point-pupil is at the higher level than the one-point-pupil. Another challenge, typical for Nepali markers is that the highest marks are not given at all, as discussed in the footnote 75. These issues are elaborated in Section 4.2.

4.6.4.2 Categorizing the items with more than 1 mark

The following mechanical logic was used in the categorization of writing items with the maximum score of higher than 1:

1. The theoretical classification of an item was set to correlate to the *highest* marks possible in the item. Hence, when the theoretical level of the item is B1.2 and the maximum mark is 4, the original marks 3 and 4 refer to level B1.2. Similarly, when the maximum points of a B2.2 level item would be 7, the values 5, 6, and 7 indicate this level.

2. The marks lower than the highest marks indicating the actual CEFR level were classified at the lower CEFR level. Hence, with four point item of level B1.2, the original marks 1 and 2 were categorized one level lower than the marks 3 and 4, that is, at the level B1.1. The logic makes sense though it is not necessarily unambiguous. Similarly, when the long essay writing was given 7 marks as the maximum score, the classification as in Table 5 was used: marks 7, 6, and 5 represent the theoretical level (B2.2), marks 4 and 3 one level lower (B2.1) and when the test taker is given only 1 or 2 points, the level would be B1.2 at the highest. The logic is that the less marks the test taker is given the less is the ability and hence, the less probable it is that the actual level measured by the item was reached. If the test taker is given a zero score, it indicates that one does not know the ability level even though it is reasonable to think that it is lower than of those with one or two marks.

4.6.5 Results

4.6.5.1 Reading proficiency in Nepal

On the basis of the proficiency levels of over 16,000 students, the average reader of 8th graders is at the CEFR level B1.1 (Figure 4.6.3). Hence, derived from Table 4, the description of an average reader in Nepal is as follows: "[(S)he] *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*." The level is not very high. The level means, for example, that an average 8th grader reader cannot read and understand newspapers independently; this level is achieved at the next proficiency level (B1.2).

8th graders' Reading Proficiency - CEFR levels

CEFR level	Percentage of students
A1.3 or lower	10.2
A2.1	13.9
A2.2	21
B1.1	24.1
B1.2	19.3
B2.1	9.4
B2.2 or higher	2.1
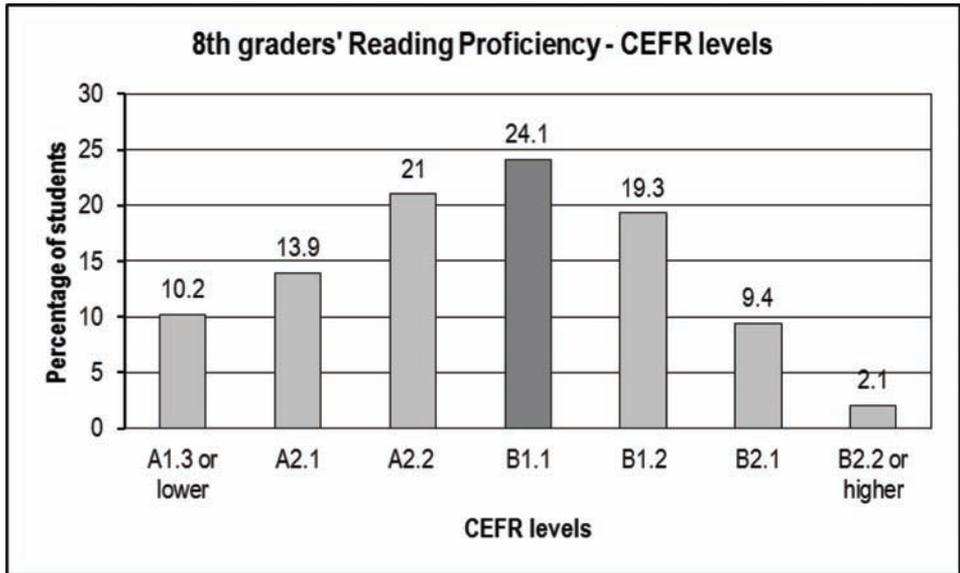
Figure 4.6.3 Distribution of reading proficiency in Nepali

In Figure 4.6.3, it is notable that 10% of the 8th graders are at the level A1.3 or lower meaning 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 low reading skill was clearly seen in the reading task called "Pokhara letter". In this indicating task 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?*". Of the alternatives of three, four, six and thirteen days, 6% of the students guessed three days, and 2% selected thirteen days. The distracting numbers were mentioned in the text, but they had nothing to do with the question.

Table 4.6.8 Percentages of 8ᵗʰ graders which has reached the specific CEFR levels in reading

B2.2	Can read independently several pages of complex text written for a variety of purposes (daily newspapers, short stories, novels).	2.1	2.1
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.	11.5	9.4
B1.2	Can read a few paragraphs of text about many different topics (newspaper articles, brochures, user instructions, simple literature)	30.8	19.3
B1.1	Can read a few pages of a wide variety of texts about familiar topics (tables, calendars, course programmes, cookery books)	54.9	24.1
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).	75.9	21.0
A2.1	Can understand simple texts containing the most common vocabulary.	89.8	13.9
A1.3	Can understand very short messages dealing with everyday life and routine events or giving simple instructions.	98.0	8.2
<A1.3			2.0

Positively thinking, 98% of the students have reached at least the level A1.3 in reading and hence they can **understand very short messages** dealing with everyday life and routine events or giving simple instructions (Table 4.6.8). From another perspective, if the reading skills will not dramatically increase within two years (grades 9 and 10), 70% or even 90% of the students are not able to apply to any higher education institutions or at least they cannot survive there because their reading skills are that low that they cannot independently read and understand the higher level textbooks (levels B1.2 and B2.1). The risk of being a weak reader is higher in Tharu, Limbu, and "other" groups (Table 4.6.9) and especially amongst the Madhesi population (Table 4.6.10).

Table 4.6.9. Reading proficiency levels in different language groups (%)

CERF	Nepali (n = 11382)	Magar (n = 525)	Tamang (n = 500)	Tharu (n = 364)	Newar (n = 178)	Gurung (n = 111)	Limbu (n = 83)	Sherpa (n = 32)	Other (n =3032)
<A1.3	1.2	0.8	1.4	3.3	1.1	0.9	1.2	0	4.7
A1.3	7.1	7.6	2.4	12.9	6.2	5.4	13.3	6.3	12.7
A2.1	12.8	21.3	11.2	19.2	3.9	21.6	27.7	12.5	16.5
A2.2	20.6	25.5	20.2	25.3	14.6	26.1	27.7	34.4	21.1
B1.1	24.8	25.3	33.0	19.5	20.8	27.0	14.5	34.4	20.7
B1.2	20.5	13.9	23.2	14.8	28.7	12.6	12.0	9.4	15.9
B2.1	10.3	5.1	8.4	4.4	20.2	6.3	3.6	3.1	7.3
B2.2 or higher	2.5	0.4	0.2	0.5	4.5	0	0	0	1.1

Table 4.6.10. Reading proficiency levels in different castes (%)

CEFR	Brahman/Cheetri (n = 6,176)	Janjati (n = 6,615)	Dalit (n = 1,523)	Madhesi (n = 1,690)	Alpasankhyak (n = 306)
<A1.3	1.2	1.3	2.6	5.9	4.2
A1.3	6.7	6.8	10.7	16.0	11.8
A2.1	12.7	14.0	14.4	17.6	12.1
A2.2	19.9	21.5	24.4	20.0	20.3
B1.1	23.8	26.4	24.6	16.3	23.2
B1.2	21.7	19.1	15.6	15.1	19.6
B2.1	11.1	9.1	6.8	7.3	7.5
B2.2 or higher	2.9	1.6	0.9	1.7	1.3

4.6.5.2 Writing proficiency in Nepal

The average writer of 8th graders is at the CEFR level A2.2 (Figure 4.6.4) though there are quite a number of very good writers too. Hence, derived from Table 4, the description of an average writer in Nepal is as follows: "[(S)he] *can manage in routine everyday situations in writing. [(S)he] 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).*" The level is somewhat lower than that of the reading even though the writing skills are highly focused in the Nepali curriculum. The level means, for example, that an average 8th grader writer cannot make lecture notes or brief summaries independently.

413

The difference between the reading and writing can be seen when comparing the reading and writing results of six assignments with short text and simple questions on the basis of the text. The same items were marked two times – one time for the reading skill (whether the contents of the answers were meaningful) and another time for the writing skill (whether the students were able to write the ideas in a correct form). When, on average, 64% of the students were able to solve the reading tasks, only 24% of them were able to mediate their answers without spelling mistakes.
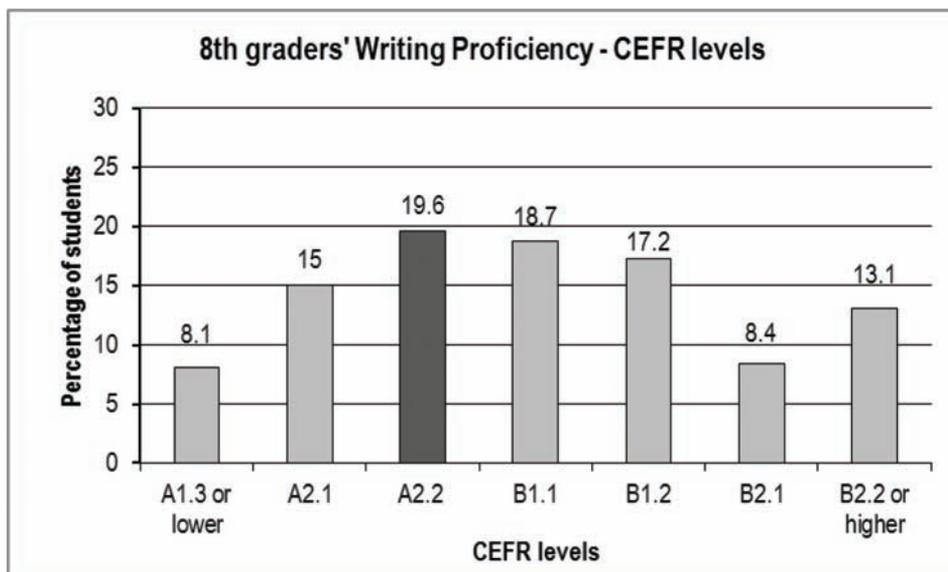
Figure 4.6.4 Distribution of Writing proficiency in Nepali

Achievement level A2.2 of the students means that while 77% of the students could write a simple application for a higher education institution, less than 40% of them could survive there because they cannot write lecture notes or summarize information (Table 4.6.11). 21% of the 8th grader students could summarize information from different sources. It is notable that 8% of the students are very weak in writing – the risk is very high in the language groups of Tharu, Limbu, and Gurung (Table 4.6.12) and especially amongst Madhesi societies (Table 4.6.13).

A positive signal in writing proficiency is that there are a notable number of very good writers among the 8th graders. According to Tables 4.6.11 and 4.6.4, somewhat 13% of the students "*can write clear, detailed, formal and informal texts about complex real or imaginary events and experiences, mostly for*

familiar and sometimes unfamiliar readers. [They] *can write an essay, a formal or informal report, take notes for future reference and produce summaries.* [They] *can write a clear and well-structured text, express his/her point of view, develop arguments systematically, analyse, reflect on and summarise information and thoughts.*" These high level writers can be found most probably from Newar and Tamang language groups (Table 4.6.12) and from the Janjati and Brahman/Cheetri societies (Table 4.6.13). It is notable that the distribution of the writing proficiency in the language groups of Nepali, Tamang, Newar, and Sherpa is not normal; in their groups there are more very good readers than what is expected. The reason is not discussed here.

Table 4.6.11 Percentages of 8th graders which has reached the specific CEFR levels in Writing

CEFR	Brief description of ability	% reaching the level	% in each level
B2.2	or higher Can write a clear and well-structured text, express his/her point of view, develop arguments systematically, analyse, reflect on and summarise information and thoughts.	13.1	13.1
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.	21.4	8.4
B1.2	Can write a few paragraphs of structured text (lecture notes, brief summaries and accounts based on a clear discussion or presentation).38.6		17.2
B1.1	Can write a clearly formulated cohesive text by connecting isolated phrases to create longer sequences (letters, descriptions, stories, telephone messages).	57.3	18.7
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).	77	19.6
A2.1	Can write brief, simple messages (personal letters, notes), which are related to everyday needs.	91.9	15
A1.3 or lower	Can write simple messages (simple postcards, personal details, simple dictation).		8.1

Table 4.6.12 Writing proficiency levels in different language groups (%)

CEFR	Nepali (n = 11,382)	Magar (n = 525)	Tamang (n = 500)	Tharu (n = 364)	Newar (n = 178)	Gurung (n = 111)	Limbu (n = 83)	Sherpa (n = 32)
A1.3	5.2	4.8	3.2	11.5	1.7	9.9	10.8	0
A2.1	12.7	16.2	8.0	24.7	9.0	7.2	28.9	12.5
A2.2	19[1]	20.2	13.4	29.1	15.2	23.4	31.3	18.8
B1.1	19.2	21.3	20.8	18.1	15.2	27	16.9	28.1
B1.2	18.9	21.1	17.4	12.6	12.4	17.1	7.2	15.6
B2.1	9.8	7.0	13.4	3.3	14.0	9.0	2.4	12.5
B2.2	11.4	8.0	18.6	0.5	21.3	3.6	2.4	12.5
> B2.2	3.8	1.3	5.2	0	11.2	2.7	0	0

1) the highest frequencies in each language group are highlighted

Table 4.6.13 Writing proficiency levels in different castes (%)

CEFR	Brahman/Cheetri (n = 6,176)	Janjati (n = 6,615)	Dalit (n = 1,523)	Madhesi (n = 1,690)	Alpasankhyak (n = 306)
A1.3	4.8	5.5	10.4	26.8	11.8
A2.1	6.7	14.5	16.5	24.4	17.3
A2.2	12.7	19.9	22.3	16.7	23.2
B1.1	19.9	20.2	17.7	13.6	16.3
B1.2	23.8[1]	17.4	16.5	9.6	13.4
B2.1	21.7	8.6	6.9	3.0	6.9
B2.2	11.1	10.3	8.1	5.3	8.5
> B2.2	2.9	3.7	1.6	0.6	2.6

1) the highest frequencies in each language group are highlighted

4.6.6 Discussion

This article introduced two methodological solutions to objectively assess the language proficiency in Nepali: the *standards* according to CEFR and *standard setting* according to a new 3TTW method. Internationally comparable proficiency levels in Reading and Writing show that the reading skills are somewhat higher than the writing skills as expected because the receptive skills are usually at the higher level than the productive skills. There are notable differences between the language groups and castes, though. The risk of being a weak reader is high in Tamang, Tharu, Newar, and Limbu groups and especially in the Madhesi population. The risk of being a weak writer is very high in the language groups of Tharu, Limbu, and Gurung and especially in Madhesi societies. In the Madhesi population, more than 50% of the students are at the lowest levels of the writing skills measured with the test (that is, A1.3 and A2.1). Practically speaking, *a notable percentage of the students in these groups have not reached the required level of reading and writing Nepali for being able to continue in the higher education in the Nepali educational system*. When comparing the reading and writing proficiency levels, it is evident that differences are much wider between the language and ethnic groups in writing skills. In many groups, especially in the language group of Newars and Tamangs, there are many exceptionally good writers. It seems that writing skills are highly appreciated in these societies.

From the equality viewpoint, the *educational system in Nepal is not able to offer equal opportunities for all students to reach the same level at Nepali*. This has a strong predictable implication for the enrollment rate in the higher education of certain minorities in the society; from certain minority populations it is practically impossible to rise to the level of reaching any study place in higher education institutions because of a low level of the school's instructional language. The reason for the low performance may be caused because of social reasons; for example, the Madhesi population has historically had a very strong connection to the Indian Madhesi population and they may think that learning Nepali is not useful when most connections are with Indians. However, there are many other small populations which evidently need more support in their language learning.

Learning a common language, the *lingua franca*, in society may be a valuable goal from the national cohesion viewpoint. However, there is another philosophy, and scientific literature supporting the idea, that the children should use their first language for creating the main concepts and hence their intellectual capacity. According to this idea, it is important that children start their school life

with their own language in order to create the basic concepts and vocabulary needed in their everyday life and later they would learn the second language. Otherwise there is a risk that they do not learn properly either their mother tongue as spoken at home and the second language spoken only in school. [78] Especially challenging is if the Nepali teachers at school are not either native speakers of Nepali or properly trained for fluent Nepali. Extremely harmful for the children's mental growth and the proficiency of the language skills is that the Nepali language is started at an early age in school but the teacher is not able to teach proper Nepali. In this scenario, the risk is high that the children's mental capacities are not optimally developed because language development corresponds with the development of our understanding. In Finland, these sad stories were noticed during the Second World War in the 1940's when small children, before they learned proper Finnish, were sent to the neighboring country of Sweden to be safe. After some years, the children were sent back and it was noticed that they did not learn the Swedish language properly and that their mental capacities were limited.

Another way of thinking this issue is that there are good examples for supporting the local languages to flourish. In Finland, for example, there are three official languages: Finnish, Swedish, and Sami (for indigenous tribes at the Northern Lapland). According to Governmental policy in Finland it is taken as richness that there are such old traditions still alive in the country – and these traditions are appreciated and supported. The Government is then bound to produce material for those languages. The wide variety of the languages and dialects in Nepal is a challenge as well as the willingness that parents and students have to see that the Nepali language is taught in schools. The best solution for the issue is that the *government takes care that all the Nepali teachers are really competent in teaching proper Nepali. This gives a challenge for the teacher's training and for the monitoring of the teachers' own proficiency level in Nepali.*

---

[78] A (degoratory) term used in the cases is *semilingualism* or *double semilingualism* (see Hassinen, 2002, 33; Baker & Prys Jones, 1998, 14; Skutnabb-Kangas, 1981, 248). It is defined as a phenomenon that one learns the native language only so that it can be used only in the most common situations at home and the main language in the society is not learned properly. It was first reported in North Scandinavia (Hansegård, 1968) where the Sami-speaking children were not allowed to use their own language at schools and they did not properly learn the official language at schools. Hence, the children did not learn either of the languages. Typical for semilingualism is that the children mix the languages arbitrary in the same sentence. Semilingualism should not be messed with *double lingualism* where the children learn both the languages properly. There are lots of good examples of this phenomenon too. It requires, however, from (two) teachers consistently the fluent speaking of both languages (or at home, one parent consistently speaking his/her own language and the other speaking her/his own language to children).

References for Section 4.6

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency*. London: Continuum.

Angoff, W. H. (1971). Scales, Norms, and Equivalent Scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd Edition), Washington, D.C.: American Council of Education. 508–600.

Baker, C. & Prys Jones, S. (1998). Encyclopedia of Bilingualism and Bilingual Education. Multilingual Matters Ltd., Clevedon.

Brandon, P. (2002). Two versions of Contrasting-Groups Standard-Setting Method: A Review. *Measurement and Evaluation in Counseling and Development,* **35**, 167–181.

Ferrara, S., Perie, M. & Johnson, E. (2002). Matching the Judgmental Task with Standard Setting Panelist Expertise: the item-descriptor (ID) matching procedure. Washington DC: American institutes for Research.

FNBE (2004). *National Core Curriculum*. Yliopistopaino: Helsinki.

Hansegård, N.-E. (1968). Tvåspråkighet eller halvspråkighet? / med ett tillägg "De språkliga minoriteterna". Stockholm: Aldus/Bonnier. Serie: Aldusserien, 0346-5454 [In Swedish]

Hassinen, S. (2002). *Simultaaninen kaksikielisyys. Läheiset sukukielet viro ja suomi rinnakkain*. Doctoral Thesis. Humanistinen tiedekunta, Suomen ja saamen kielen ja logopedian laitos, Oulun yliopisto. [In Finnish]

Impara, J. C & Plake, B. S. (1997). Standard Setting: An Alternative Approach. *Journal of Educational Measurement*, **34**(*4*), 35–56.

Impara, J. C. & Plake, B. S. (1998). Teachers' Ability to Estimate Item Difficulty: A Test of the Assumptions in the Angoff Standard Setting Method. *Journal of Educational Measurement*, **35**(*1*), 69–81.

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational Measurement*. (3rd Edition), Washington, D.C.: Americal Council of Education. 485–511.

Kaftandjieva, F. (2004). Standard Setting. Reference Suplement, Section B in S. Takala (Ed.) (2009). Relating Language Examinations to the CommonEuropean Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Language Policy Division, Strasbourg.

Kingston. N., Kahl, S. R., Sweeney, K. & Bay, L. (2001). Setting performance standards using the body of work method. In **G. J. Cizeck** (ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum. 219–248.

Livingstone, S. & Zieky, M. (1982). Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests. Princeton, NJ: ETS.

Loomis. S. C. & Bourque, M. L. (2001). From tradition to innovation: Standard-setting on the National Assessment of Educational Progress. In **G. J. Cizeck** (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum. 175–218.

Metsämuuronen, J. (2009). *Ratkaisuprosentin muuntaminen taitotasoksi – A-Finska-oppiaineen aineiston erityiskysymyksiä [Changing the percentage of correct answers into proficiency level – specific questions of A-Finnish assessment].* Unpublished memorandum 28.12.2009. [In Finnish]

Metsämuuronen, J. (2010). Omvandlingen av poängtal till färdighetsnivåer inom det receptive delområdet av provet i A-lärokursen och i den modersmålsinriktade lärokursen. [Changing the percentage of correct answers into proficiency level in Finnish language for Swedish speakers A-course and native speakers' course test]. Appendix VII in O. Toropainen (2010). *Utvärdering av läroämnet finska i den grundläggande utbildningen. Inlärningsresultaten i finska enligt A-lärokursen och den modersmålsinriktade lärokursen i årskurs 9 våren 2009*. Uppföljningsrapporter 2010:1. Utbildningsstyrelsen. Helsinki. Retrieved from http://www.oph.fi/publikationer/2010/utvardering_av _laroamnet_ finska_i_den_grundlaggande _utbildningen. [In Swedish] pp. 164–168.

Metsämuuronen, J. (2013). A new method to setting standard for the wide range of language proficiency levels. *International Education Research* **1**(1). Available at http://www.todayscience.org/IER/v1/Jari % 20 Metsamuuronen.pdf.

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*. Mahwah, NJ: Lawrence Erlbaum. 249–281.

Nedelsky, L. (1954). Absolute Grading Standards for Objective Tests. *Educational and Psychological Measurement*, **14**(1), 3–19.

Reckase, M. D. (2000). A Survey and Evaluation of Recently Developed Procedures for Setting Standards on Educational Tests. In **M. L. Bourqey & Sh. Byrd** (Eds), *Student Performance Standards on the National Assessment of Educational Progress: Affirmations and Improvement.* Washington, DC: NAEP. 41–70.

Skutnabb-Kangas, T. (1981). *Bilingualism or Not: The Education of Minorities.* Multilingual Matters, Clevedon.

Takala, S. (ed.) **(2009).** Relating Language Examinations to the CommonEuropean Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Language Policy Division, Strasbourg.

Takala, S. & Kaftandjieva, F. (2009). Reseptiivisen kielitaidon kokeen muuntaminen taitotasoiksi. [Changing the receptive language test to proficiency levels]. In **E Tuokko**, *Miten Ruotsia osataan peruskoulussa? Perusopetuksen päättövaiheen ruotsin kielen B-oppimäärän oppimistulosten kansallinen arviointi 2008.* [How Swedish language is spoken in general education? National assessment 2008 of the student achievement in Swedish language B-course at the end of basic education]. Oppimistulosten arviointi 2/2009. Opetushallitus. Helsinki: Edita Prima Oy. 118. Retrieved from http://www.oph.fi/julkaisut/2009/miten_ruotsia_osataan_peruskoulussa [in Finnish].

Toropainen, O. (2010). Utvärdering av läroämnet finska i den grundläggande utbildningen. Inlärningsresultaten i finska enligt A-lärokursen och den modersmålsinriktade lärokursen i årskurs 9 våren 2009. Uppföljningsrapporter 2010:1. Utbildningsstyrelsen. Helsinki. Retrieved from http://www.oph.fi/publikationer/2010/utvardering_av_laroamnet_finska_i_den_grundlaggande_utbildningen. [In Swedish]

Tuokko, E. (2009). Miten Ruotsia osataan peruskoulussa? Perusopetuksen päättövaiheen ruotsin kielen B-oppimäärän oppimistulosten kansallinen arviointi 2008. [How Swedish language is spoken in general education? National assessment 2008 of the student achievement in Swedish language B-course at the end of basic education]. Oppimistulosten arviointi 2/2009. Opetushallitus. Helsinki: Edita Prima Oy. Retrieved from http://www.oph.fi/julkaisut/2009/miten_ruotsia_osataan_peruskoulussa [in Finnish]

van der Schoot, F. (2009). Cito variation of the Bookmark Method. Reference Supplement, Section I. In S. Takala (ed.), Relating Language Examinations to the CommonEuropean Framework of Reference for Languages: Learning, Teaching, Assessment (CIFR). A Manual. Language Policy Division, Strasbourg.

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.

Section **5**

Conclusions and Implications of NASA 2011

Dr. Bhoj Raj Kafle[1]
&
Dr. Jari Metsämuuronen[2]

1. NASA unit, Educational Review Office (ERO), Ministry of Education, Nepal
2. Helsinki University, Faculty of Behavioral Sciences

5.1 Objectives of the Study

The main objective of this study was to present a national scenario of grade 8 student achievement by providing evidence on the student learning in the education system. More specific objectives were nuanced in Chapter 1.1. The main objectives of the study were:

1. To determine the current national level of achievement of eighth-grade students in Mathematics, Nepali, and Social Studies (MFs 3 and 7);

2. To determine variations in student achievement between different ecological zones, Developmental regions, districts, school location (rural/urban), ethnicity, gender, language groups, socioeconomic conditions as well as school type (community/institutional schools) (Major Findings [MF] 1, 2, 4, 5, and 6);

3. To examine extent to which the school, teacher, home background, and other pupil factors influence learning achievement (MFs 4, 5, and 6);

4. To compare student learning achievement in the current study with that of the 2008 study (MF 8);

5. To compare the student learning achievement in Nepal with that of international studies: PISA (Reading) and TIMSS (Mathematics and Geography) (MF 8).

Each finding is discussed in relation to relevant suggestions for educational development in order to generate recommendations for policy makers to improve educational quality as well as for other stakeholders to improve their respective working areas. The recommendations are summarized in Section 5.3.

5.2 Major Findings and Discussion Areas for Implementation

This section comprises the main findings from NASA 2011.

Major Findings and Discussion Areas for Implementation

1. Differences: There are great differences in achievement between the students, schools, districts, and developmental regions.

The study confirms that there are great differences between the achievement levels across students and schools in Nepal. Some students in the survey did not give a correct answer to any single test item (0%) while the best students gained more than 95%. The average achievement of students in the lowest performing schools was below 15% while the students in some of the highest performing schools gained, on average, over 90%. When knowing that the average score in a school is below 15%, it means that in those schools many students fall below that level. From an equal opportunities viewpoint, it is not a good sign that 'low' and 'high' performing students are concentrated in certain schools.

Not only is the variation great amongst students and schools, but it is also significant across the 25 districts covered by the study. In the five lowest performing districts (Bhojpur, Jajarkot, Morang, Kanchanpur, and Ilam), the average performance was below 40%, while in the three highest performing districts (Lalitpur, Kathmandu, and Bhaktapur) it exceeds 55%. The students in the Kathmandu Valley schools (59%) exceed the average students in the other regions (41–49%) across numerous indicators. The achievement level of the students from the Eastern (41%), Far-Western (42%), and Mid-Western (43%) development regions is far behind that of the students in the Kathmandu Valley. Once again from an equal opportunities standpoint, this is not a positive sign.

Further analysis of the highest and lowest performing community schools gives some hints as to the possible reasons behind the deviating results. (1) The number of study days in the highest performing schools is higher than in the lowest performing schools. (2) The higher performing schools are smaller (388 students or less) than the lowest performing schools (461 students or more). (3) The student/teacher ratio is smaller in the highest performing schools. (4) The highest performing

students spent more time on homework, they were less absent from school, more of them did not need to work while studying, they were younger, and their attitude towards the subject was positive. (5) Parents' illiteracy – especially that of fathers – is prevalent within the lowest performing students.

Discussion and suggestions: Such a significant difference in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. Though lots of work has been done to reduce the gap, there still would appear to be a substantial amount of work to do. As a minimum, the confirmed state of educational inequity across the country across most indicators should at least initiate discussions, within and between the stakeholders, as to how the difference in average achievement among the schools can be successfully reduced.

The possible reasons behind the deviation indicated above, provide a starting point for such discussions. The lower number of study days challenges the stakeholders to consider whether the main reason behind this indicator could be the number of strikes and disturbances in these regions due to which the schools remained closed for several days. If this was the reason for the low number of study days, it would urge teachers' unions and other partners to find more positive ways to seek justice without harming the opportunities of the children to gain constant access to the education system. It also challenges all district education offices to detect those schools which are at danger of becoming one of the poorest performing schools. The possible reasons for the low performance should be identified and actions should be taken to address this. If more resources are needed for raising the standard in these lowest performing schools, it might be recommendable to create a system within the Ministry of Education to distribute grants to these schools, in a spirit of positive discrimination, for example.

2. Institutional/Community Schools: The students in institutional schools perform well. Within the community schools, there is a wide range in average student achievement. The difference between the low and high performing community schools is remarkable.

The average achievement in institutional schools is remarkably higher (63%) than in the community schools (44%). The comparison is, nevertheless, unfair because the average socio-economic status (SES) is more than twice as high in the private schools (74%) as compared with the students at the community schools (34%). This means that the educational, social and economic capacity in these families to provide private tuition, for example, is much higher in the private schools. Hence, the higher results can be explained by a strict selection of the students and by

tuition given by some external source outside of the schools, and not necessarily by the better processes within the school itself.

What is important is that there are a significant number of community schools where the average results are at the same level as in the private schools even though the SES is radically lower. In these schools, either the processes are more effective than in the private schools or the students are of the same ability as those in the private schools and are not adversely affected by the processes within the school or their socioeconomic status.

The analysis based on comparing the highest and lowest performing community schools showed that in the highest performing schools teachers are using more homework evaluation and classroom evaluation, they have more interaction with other teaching staff, and they have a more realistic perception of their students learning capacity. In the highest performing schools, there was more interaction with the community, the student/teacher ratio was smaller, there were fewer incidents of being late in schools as well as destroying school property, and there was more interaction with teaching staff with regard to improving teaching and learning processes. The students in the highest performing schools spend more time on homework, they are less absent, they do not spend too much time playing and doing sports, or on home chores, and they do not undertake paid work, nor are they bullied as much as their counterparts in the lowest performing schools.

An analysis of all community schools shows that the better results were achieved when the student/teacher ratio was lower than 31 students, the number of study days was 200 or more, when there are computers for mathematics teaching and staff for maintenance and technical assistance, when there are few or no incidents of being late for school, or of the destruction of school premises, when there are interactions with the teaching staff on improving teaching and learning processes more than seven times a year, and when there are classroom evaluations more than twice a week.

Discussion and suggestions: A major lesson which can be learnt from the findings is that many of the community schools are capable of maintaining a high level of achievement despite the poor socioeconomic status of their students. These schools seem to use more effective processes than the private schools because they can reach similar levels of achievement with significantly fewer resources. On the other hand, there are many community schools where the average performance is very low. These schools with a similarly low socioeconomic status need to learn from these other comparable schools about the processes they are utilizing to achieve these high results.

A comparison of the highest and lowest performing schools gives some hints as to the areas to focus on: the low number of study days, absenteeism in schools, students' paid work, less effective practices in home work and classroom evaluation, as well as low communication between the community and teaching staff.

These kinds of activities and checklists provided in the report may be used by the district officers when monitoring the schools. Not only the schools but also all those concerned with educational development and management including policy makers, implementing and monitoring agencies, curriculum developers, training providers and managers, should, at least, be aware of these differences between the highest and lowest performing community schools.

What activities are carried out and how learning takes place in such schools is a matter of discussion at the policy making level. It is urgent that such issues be discussed at the policy making level and solutions to the problems sought. All possible efforts should be made to reduce the differences between these schools.

3. Cognitive skills: Students are apt in tasks related to memorization and recall, but are not effective in skills requiring application or ability at a higher cognitive level.

Students are comparatively poorer in the ability to solve problems, to analyze, deduce logic, generalize, justify an argument or viewpoint, and in the ability to transfer learning from one context to another. A remarkably high number of students were able to solve only 15% or less of the practical problems, (17% of the students in Mathematics, 13% in Nepali, and 8% in Social Studies).

In Mathematics, students are able to do basic calculations, but are weak in reasoning, problem solving, plotting, 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.

In Nepali, students performed well when called upon to recognize the correct answer, in recalling simple facts from the texts, fundamental thinking, the basic interpretation of a paragraph, tables, charts, and in logical thinking that required only a few steps. However they are much weaker in producing fluent texts or letters, and in preparing synthesis and abstracts from a text. In Nepali, the students did attempt open-ended tasks but the skills were not high enough for the highest marks.

In Social Studies, students are good in recognizing the correct answer, in very fundamental knowledge/content, in true or false questions, matching texts,

and in the selection of words for gap fill activities. They are much weaker in reasoning, problem solving, and in constructing arguments.

Contemporary thinking is that constructive and child-centered methods may help with higher cognitive tasks. The comparison of Nepali and Finnish teachers' perceptions of good teaching shows that the teachers in Nepal are aware, at least in theory, of the same constructivist principles of teaching and learning as the Finnish teachers. However, in Nepal, the classroom reality of large class numbers, sometimes in excess of 100 students in a single class, and excessive classroom sessions per day do not encourage teachers to use student centered teaching methods, which may in fact be a necessity for students to develop their higher ability skills'. The reality of the classrooms in many cases is that teachers lecture, although interestingly the teachers themselves do not necessarily see it this way, suggesting a possible lack of understanding of the constructivist principles despite their claims to be aware of them.

Discussion and suggestions: The educational system in Nepal would appear to be geared toward remembering items rather than solving problems creatively. This begs the following questions: What are the reasons for this culture? How is it nurtured? What could be done to change this culture? What kind of citizen is the education system producing?

Traditionally, in South Asia, wisdom has been transmitted orally with the onus on remembering by heart in order to pass knowledge on. Once introduced, the written text was not taken to be as reliable as that of the spoken word, and hence, the tradition continued to favour remembering teachings by heart. This long-lived oral tradition connected with the deep-rooted culture of idolism, has given the teacher the position of a 'trusted' guru. The students are taught at a very early age to trust the teacher without question. Students are expected to receive this knowledge passively without opportunity for inquiry or exploration on their own account (in the modern constructive psychological manner: children are not given possibilities to create their own realities).

Consistently large classroom sizes of 40–100 students, as well as the heavy-loaded curricula, nurture teacher-led teaching methods. To complete the whole curriculum is not an easy task if the teacher is expected to cover all areas.

The analysis of effective class size in Nepal shows that the optimum student/teacher ratio, from an achievement point of view, is 22 students or less per teacher (mean score 55%), the lowest results come from schools where the student/teacher ratio is more than 31 students (43%).

As a comparison, in Finland, the student/teacher ratio is 13.5 and the maximum class size is usually 25 students. A median classroom size for 1st–6th graders is 60 m² for 25 students, which equals 2.5 m² for each student to work and move in. Practically speaking, all Finnish students have their own desk with half a meter of space around their work area in which to move. To change the size of classroom is not an easy task. It requires more teachers, new school buildings, new ways of thinking of the students' learning environment, as well as the proper maintenance of premises.

Similarly, though the teachers claim to know the meaning of continuous evaluation in class it is doubtful to what extent it could really be applied when there are so many students within one classroom.

The curricula, also seems to be loaded with detailed tasks to fulfill. If this is feeding the teacher-centered methods in schools, it may be worth rethinking the role and form of the curricula. Perhaps the curricula could be shorter with more guidelines as it is in Finland rather than an exhaustive teachers' guide as it is in Nepal. Or training on how to employ the curricula in a creative way may be necessary. Either way, the pedagogy and its practical application in the classroom should be changed to allow for more creative thinking, reasoning, and application.

4. Socio-economic background: The low educational and social background is directly and strictly related to low results.

The dataset gives a strong indication that Socio Economic Status plays a significant role in the educational processes in Nepal. The difference in the achievement level of the students from the lowest (41%) and highest SES group (64%) is remarkable. One of the indicators for the SES was parents' educational level and specifically parents' illiteracy. In the dataset, out of 1200 schools, there are 73 schools where more than 40% of grade 8 students reported to have an illiterate father and 609 schools where more than 40% had an illiterate mother. 82 schools have a mothers' illiteracy rate of more than 80%, and there were 18 schools where all the mothers were reported to be illiterate. Many students from illiterate families can reach a high achievement level, however, when the parents are, or the father alone is illiterate, the probability of being within the highest performing students is quite low (39% or the highest level students) compared with the situation when both the parents are literate and they are working outside of agriculture (64% of the students).

When there are more than 10% of students with an illiterate father in the school, the average result of the school is significantly lower (46%) than if there

was less than 3% illiterate fathers (62%). Correspondingly, when there are more than 54% of the students with an illiterate mother in the school, the average result of the school is significantly lower (43%) than if there was less than 16% illiterate mothers (60%). When both parents were literate, the achievement level was 63%.

Another aspect for the impact of the illiteracy is that by using the illiteracy rate in the school, it is quite easy to provide a rough indicator for assessing whether the school belongs to the lowest level schools (the lowest quintile). If there were 8% of the students with illiterate fathers in a school, the probability of it belonging to the lowest quintile is 65%. In parallel, if 40% of the students had an illiterate mother, the probability of belong to the lowest level schools was 38%. If more than 60% of the students had an illiterate mother, the probability would be 60%. That is, by simply asking the students how many of them have an illiterate father, we have a clear indicator of whether the school is one of the lowest performing. If one out of ten students raise their hand, the school most probably is one of those with very low results. The same can be inferred by asking the number of illiterate mothers in the school; if six out of ten students raise their hand, the school most probably is one of those with very low results.

Poverty is seen in another way in the data. The students who need to carry out paid work during their school time perform statistically lower (40–41%) than those without a need for paid work (50%).

Discussion and suggestions: The educational and social capacity of the family may impact on the child in several ways. First, the parents' education and especially their literary skills lay the ground for the child's early development, in particular on their range of vocabulary, and on the child's motivation to learn. Second, more educated parents are most probably more motivated to push their child to study harder in order to improve, or at least to maintain their social position. Third, educated parents understand the value of education in raising the standard of their child's future. Fourth, educated parents can give tuition to their children in the early years of schooling which may help the children radically in these years. Several other advantages could also be found, however, it is not easy to change the social structure in a society without realistic possibilities of finding employment or a place of study after primary education.

Parental illiteracy is most probably malignantly tied to the occupational background of the parents, a low motivation to educate their children, poverty, or, in some cases, even to the caste of the family. All possible efforts should be made to raise the level of literacy among the parents of the young children.

Another question related to illiteracy, is how to reduce the negative effect of poverty on educational achievement. It is assumed that families with very low incomes are not motivated to send their children to school because the children are needed in making a living. If the children were taken care of, by offering them, for example, one warm lunch and a small snack during the school day, it may reduce the need for paid work and so motivate the parents to send the children to school. All the children in all the schools would need this support regardless of the achievement level of the schools. However, it may be reasonable to start the aid in those schools which are detected as the lowest level schools. The district offices could detect these schools and national grants could be addressed to them.

5. Caste/ethnicity/home language: Achievement level still depends on caste/ethnicity as well as on the home language of the student.

The achievement level of the Dalit students is found to be lower in all 3 subjects. Dalit girls especially seem to be low-achievers in Mathematics (35%), Nepali (35%), and Social Studies (45%), as compared with the Dalit boys (39%, 40%, and 48% respectively). It is worth noting that in the Valley, Dalits' results are much higher than the average for the country.

Madhesi students' performance in Mathematics is higher than the average (49%), but the results in Nepali (45%) and Social Studies (47%) are lower than the average. In the Madhesi population, the girls seem to be behind the boys especially in the Mountain zone (44% against 50%), in the Hill zone (42% against 50%), and to some extent in the Tarai zone (42% against 45%).

With regard to the Nepali language, using the standard setting of the CEFR levels, reading proficiency was remarkably lower in the Madhesi population than in all other castes. 40% of the Madhesi' students fell into the category of "lower than A2.2". This is a remarkably high value when compared with the other castes with 21–28% of students in this category. This indicates that they cannot even "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)" as defined in the nuanced CEFR classification. The result means that almost half of the Madhesi students are at the level where their Nepali reading ability is limited to "simple texts containing the most common vocabulary". Also, in the language groups of Limbus and Tharus there are far too many students at this low level (42% and 35% respectively) and a high proportion in the Magar (30%) and Gurung (28%) groups too.

With regard to writing skills, it is alarming that 51% of Madhesi students are classified at the level A2.1 or lower. This means that they can (only) write "brief, simple messages (personal letters, notes), which are related to everyday needs" in Nepali, but they could not, for example, "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) (level A2.2). At this low level there are an astonishing 100% of Yadavs, 40% of Limbus, 36% of Tharus, and 33% of Rais. The number of Yadavs and Rais are, however, very small in the data and, hence, no deep analysis or interpretation can be done.

Discussion and suggestions: The skills of reading and writing are essential for reaching a study place in Higher Education. In all language groups, especially among Newars, Tamangs and Urdus, there are many exceptionally good writers. However, it is evident that a notable percentage of students have not reached the required level of reading and writing of Nepali to be able to continue in Higher Education in the Nepali educational system. From an equality viewpoint, the educational system in Nepal is not able to offer equal opportunities for all students to reach the same level in Nepali as already discussed above.

There may be natural reasons for this, as historically the Madhesi population have had a very strong connection to the Indian Madhesi population and they may think that learning Nepali is not useful when most connections are to be with Indians. However, there are many other small populations which evidently need more support in their language learning. In any case, from a national cohesion viewpoint, it is important to think of possibilities of integrating the strong Madhesi population into Nepali society. One option is to offer the minority populations possibilities to study mainly with their own language and to study the Nepali language as an obligatory second language. This suggestion is not easy to apply because it will lead to a situation where there should also be study places in the Higher/tertiary education with the minority languages, too. Hence, this would prove quite a radical change in society. The advances and risks of the change should be carefully surveyed by the National Planning Commission, as the number of non-Nepali speaking groups is so large that providing mother tongue language instruction in schools to some or one of them and not others will have potentially massive political repercussions.

6. Similarities: There are no remarkable differences between boys and girls, rural and urban schools, and across ecological zones.

Though the boys are slightly out-performing girls in all content areas of Mathematics the differences are not significant. Generally speaking, in all subjects, there seems to not be any remarkable difference in achievement between the boys (48%) and

girls (46%) who have reached grade 8. Although the results are more positive in cities (52%, without Kathmandu Valley 47%) than in the rural areas (45%) the difference again is not remarkable. With the exception of Kathmandu Valley, there are no differences between the average achievement levels of the students from the Mountain (46%), Hill (45%), and Tarai (45%) ecological zones. In fact they are remarkably consistent. From an equity point of view the signal is positive, although there is work to be done to reduce the gap between Kathmandu Valley and the other areas.

Discussion and suggestions: The results indicate positive signs with regard to educational equity between sexes, the locality of the school, and ecological zones. The results do not, however, mean that there is no need to follow-up on these indicators in the future. Another aspect is that there might be remarkable deficiencies in equity when it comes to certain subgroups in ecological zones and school location as well as between sexes. One of these issues is that of the ethnic groups; for example, among Madhesi students, the achievement of the girls is noticeably lower than that of the boys. It is also known that many non-Nepali speaking children drop out in early grades. The main suggestion is that, in the future, gender equity is stressed more at the lower grades rather than at the higher grades. However, it is still important for the Curriculum Development Centre to review and analyze its educational materials from a gender equity viewpoint. Likewise, current efforts for teacher training also require a review from the gender perspective. Each school needs to analyze the annual achievement level from a gender perspective and to adopt measures for improvement. Teachers need to be sensitive towards this issue and carry out appropriate learning activities.

The educational opportunities are much greater in the Kathmandu Valley than in the remaining part of the country. It would appear that the main intellectual capacity of the country is concentrated in the Valley, mainly because of the wide possibilities to both study in the universities and work in high level business. Hence, it is wise to think how the intellectual capacity could be protected and enhanced all over the country and not only in the Valley. If educated parents in other areas find out that their children are able to get the same high scores as are received in the Valley, it may halt the uncontrolled shift of the population to the Valley.

7. Three content areas: In Nepali, achievement in reading and writing is low in absolute terms – an average student of grade 8 cannot read and write sufficiently well to manage higher studies for example. Achievement in Mathematics is not distributed normally, and in Geometry and Sets it is remarkably low when compared with the other content areas. In Social Studies, the achievement levels in Politics, History, and Civics is lower than the other content areas.

433

In Nepali, the Common European Framework of Reference for Languages was employed to identify the absolute achievement of students in the language skills of reading and writing. It was found that, although there are many good readers and writers in the sample, an average 8th grader reader of Nepali cannot read and understand newspapers independently. The risk of being a weak reader is higher in the language groups of Tamangs, Tharus, Newars and Limbus, and especially so amongst the Madhesi population. An average 8th grader writer of Nepali cannot make, for example, lecture notes or brief summaries independently.

8% of the students are exceptionally weak in writing with the risk being very high once again in the language groups of Tharus, Limbus, and Gurungs, and especially so in Madhesi societies.

In Mathematics, the population is not normally distributed as it is in the Nepali and Social Studies assessments. The distribution suggests that there are three distinctive student populations: 'low', 'mediocre', and 'high' performing students. The low-performing students where the majority of the students lie, achieve on average, a score of 20–25%, the medium-performing students 40–50%, and the high-performing students as high as 70–80%. Of the content areas, the achievement is lower in Geometry (37%) and Sets (38%) than in the other areas (48–49%). In Social Studies, the achievement is lower in Politics (46%), History (50%), and Civil Society (51%) as when compared to Economics (58%) and Geography (57%).

Discussion and suggestions: Grade 8 students have not developed their reading and writing skills in Nepali to a level required to access Higher Studies. In particular, they are weak in writing essays, reports, and summaries, and in the ability to give an opinion, and to express their own views and those of others. It is clear that many of these students may not be able to access Higher-level studies due to this lack of skills. From an equality viewpoint, the educational system is not able to offer equal opportunities for all students to reach the same level in Nepali. This has a strong predictable implication for the enrollment rate in Higher Education, as at present from certain minority populations it is practically impossible to rise to the level of studying in a Nepali Higher Education institution.

On the basis of the pretests for the reading test, students were not, in general, used to reading longer texts or in expressing their opinions on the basis of the read text. At present only 10% of the time allocation at grade 8 is given to reading whereas 65% is given to writing, and 25% for grammar and vocabulary combined. The validity of this weighting therefore needs review.

As Nepali is a second language for more than 30% of the students in the sample, it seems appropriate for the pedagogy to be developed accordingly, such that Nepali is not taught as a mother tongue to these students but as a second language, else these language minorities will continue to suffer from poor academic performance. Learning Nepali may in many cases be too demanding for a non-Nepali student if the teaching methods used are directed at Nepali speakers, and do not cater for second language learners. In many cases, this is the reality. In addition, there is the issue of the capability of the teacher as this challenges those teachers who themselves are not speaking Nepali as their native language as it is significantly more challenging to teach children to speak, read, and write Nepali if the teacher is not himself fluent.

In the Mathematics test, 5% of the students scored less than 15%. These students will have severe challenges in their possibilities to enter Higher Education, not to mention their everyday life numeracy. The same students almost certainly scored very low at the lower grades too, suggesting a need to identify low-achievers early on and to make additional arrangements to cater for their particular needs.

8. 2008/2011: The results in Mathematics have slightly declined from those of 2008 while the results in Nepali have increased. Reading skills in Nepal are lower than at the international level.

The average achievement level in Nepali was 49%, which is 2% higher than in the year 2008. In Mathematics, the score of 43% was 4% lower than in 2008. The Social Studies score of 49% is not comparable with the 2008 results because of obvious differences in the marking schemes of the linking items.

In all content areas, Mathematical skills are slightly lower in the year 2011 as compared with 2008. This is clearly seen in most linking items, where the average performance in 2011 is 8% lower than in 2008. The Nepali results have increased, on average by 2% within the four years. The level has especially risen in Writing (+16) and Reading (+6) but it has dropped dramatically in Vocabulary (-22). On the basis of the dataset it is not possible to say anything firm with regard to development in Social Studies between the years 2008 and 2011, due to the apparently different marking schemes.

The dataset gives a signal that the average reading proficiency in Nepal is much lower (-0.88) than the international average (0.00). This is supported by the fact that the CEFR levels show a low achievement in reading. According to the dataset, the average achievement of Mathematics and Geography in Nepal seems to be slightly better than the international average (+0.27 and +0.31 respectively). Though the results are obtained by using a good number of linking items to the

international item bank, there are mitigating factors, such as the poor level of India in the TIMSS 2011 survey, so more rigorous studies are needed to confirm the results.

Finally, a comparison of the same test items for the 2008 and 2011 samples shows that the students' ability to solve these problems has reduced. This is possibly an indication that there are more weak students enrolling in studies because of a successful "Education for All" program.

Discussion and suggestions: The change between 2008 and 2011 does not necessarily need any action. It is, though, a good sign that the reading and writing proficiency levels seem to have been increased within these three years. It is more difficult to explain why the results in Mathematics have decreased. It may be possible to explain part of this by a stricter marking scheme in 2011, but it is too early to give any recommendations concerning the result. NASA 2013 testing will give more comparable evidence of the change; as then the 2013 results can be compared with both the 2011 and 2008 results.

It is worth mentioning that the equating procedures used in Nepali, Mathematics, and Geography was identical. Hence, there is no reason related to the equating process which would explain the deviating results between the subjects. However, one technical aspect may explain the incredibly high results in Mathematics and Geography as compared with Nepali. Namely, hindsight shows that the selection of the items from the international item bank differed, maybe crucially, between the subjects. In Nepali, the reading items were selected without any connection to curriculum while the Mathematics and Geography items were selected on the basis of the local curriculum. The consequence seems to be that, in the Nepali reading content area, the equated achievement level corresponds better to the international level while in Mathematics and Geography, the obtained results are better than if items were randomly selected from the international bank. This may be a remarkable lesson to learn for the next NASA rounds.

5.3   Suggestions and Recommendations Classified by Stakeholder

This section comprises the suggestions and recommendations from Section 5.2 and presents them as classified by the relevant stakeholders.

National planning commission

- The educational system in Nepal is geared towards the art of recall rather than higher cognitive processes such as problem solving. *What kind of citizens is the Education System producing?*

- The large class sizes of 40–100 students, as well as the heavy-loaded curricula, nurture a teacher-led or teacher-centered teaching methodology. *What kind of teachers is the Education System producing?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

- The educational opportunities are much higher in the Valley than in the remaining part of the country. The main intellectual capacity of the country is at present concentrated in the Valley. *How to address this imbalance?*

- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

- Parents' education and especially their literary skills would appear to be a significant factor in achievement levels in Nepal. *How to raise the literacy levels among the parents of the poor performing children, as well as the children themselves?*

- It is assumed that families with very low incomes are not motivated to send their children to school because the children are needed to help make a living. *If the children were taken care of by offering them, for example, one warm lunch and a small snack during the school day, it may reduce the need for paid work and so keep these children in school?*

- From the national cohesion viewpoint, it is important to discuss the possibilities of integrating the strong Madhesi population as well as the large Tharu society into common Nepali society. *One option is to offer these minority populations the possibility to study mainly in their own language at the lower grades and to study the Nepali language as an obligatory second language?*

Department of education (DOE)

- Low achievement level of many students in Mathematics. *Need for a supportive national program for the slow learners?*

- Teaching pedagogy and practices in the classroom focus on learning by rote teacher-centred activites. *How to allow for a more student-centred approach focusing on areas such as creative thinking, problem solving and deductive reasoning?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

- Low achievement level of many community schools. *Need for a supportive national program to positively discriminate in favour of these schools?*

- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

Teacher educators

- Low achievement level of many students in Mathematics. *Need for a supportive national program for the slow learners with a new pedagogy?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

Curriculum developers

- Low reading and writing skills on average in Nepali, and specifically in certain language minorities. *How to reduce the gap between different groups in language proficiency?*

- Students are not used to reading longer texts or to expressing their opinions on the basis of the read text. *Should the allocation for reading be higher than the 10% currently specified?*

- Very low results of the main student population. *Is the curriculum so loaded with material it is causing difficulties for the teachers?*

- The educational system in Nepal is geared toward remembering things rather than solving novel problems. *Should the curricula be shorter, with more guidelines rather than an exhaustive teachers' guide?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

- No remarkable differences in achievement between boys and girls. *Still important to review and analyze educational materials from the gender equity viewpoint to maintain this?*

- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

Teacher trainers

- Low reading and writing skills on average in Nepali, and specifically in certain language minorities. *How to reduce the gap between different groups in language proficiency?*

- Nepali is a second language for more than 30% of the students in the sample. *Should the pedagogy be developed, such that Nepali is not taught as a mother tongue to these students but as a second language?*

- Very low results of the main student population. *Should there be more emphasis on support with remedial teaching for those who are in danger of dropping out of Mathematics in the early grades?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

- No remarkable differences in achievement between boys and girls. *Still important to review and analyze educational materials from the gender equity viewpoint to maintain this?*

- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

School leave certificate (SLC) board

- The educational system in Nepal is geared towards the art of recall rather than higher cognitive processes such as problem solving. *Should the SLC Examination be rewritten to promote higher cognitive processes?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

Non-formal education centre

- Parents' education and especially their literary skills would appear to be a significant factor in achievement levels in Nepal. *How to raise the literacy levels among the parents of the poor performing children, as well as the children themselves?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

District education offices (DEOs)

- Low reading and writing skills on average in Nepali, and specifically in certain language minorities. *How to reduce the gap between different groups in language proficiency?*

- Low level of Nepali in certain language minorities. *How to support teachers with low levels of Nepali to teach in the medium of Nepali?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

- Low achievement level of many community schools. *Need for a supportive national program to positively discriminate in favour of these schools?*

- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

- It is assumed that families with very low incomes are not motivated to send their children to school because the children are needed to help make a living. *If the children were taken care of by offering them, for example, one warm lunch and a small snack during the school day, it may reduce the need for paid work and so keep these children in school?*

Teachers, teachers' unions, and head teachers

- Low achievement levels in Nepali in certain language minorities. *How to reduce the gap between different groups in language proficiency?*

- Very low results of the main student population. *Should there be more emphasis on support with remedial teaching for those who are in danger of dropping out of Mathematics in the early grades?*

- The significant differences in the achievement level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. *How to bridge the significant gap in educational achievement?*

- Low number of study days. *Are bandhs useful, how to seek justice without harming the opportunities of the children to access educational provision?*

- No remarkable differences in achievement between boys and girls. *Still important to review and analyze educational materials from the gender equity viewpoint to maintain this?*

- There are many community schools where the average performance is very low. Clear indicators have been identified as to why some community schools perform poorly and others perform exceptionally well. *How to use these indicators to bridge the gap in educational achievement?*

5.4 Final Conclusions

The findings of NASA 2011 found, and confirmed that there are serious deficiencies in the current educational system of Nepal. The great differences in the achievement

level of students, schools, districts and regions indicates the weak capacity of the current education system to produce equal educational opportunities for all students. Many of the issues raised in the report can be tackled at the grass-roots level by teachers, head teachers and District Education Offices when there is a will to make the necessary changes. Some necessary changes can be made by middle level actors such as Teacher trainers, Curriculum developers, Special education experts, the SLC board or at the Department of Education without any great policy level action. However, some of the required changes are quite fundamental, such as tackling the issue of the classroom sizes; changes in the system to raise the level of literacy among the parents of the young children, and not only among the children themselves; of offering one warm lunch and a small snack during the school day for those children in most need and from the poorest families; or to offer the minority populations, such as the children from Madhesi and Tharu communities, the possibility to study mainly in their own language at the lower grades and to study Nepali language as an obligatory second language.

The NASA 2011 results offer rich data and analysis in raising the standard of Education in Nepal. The articles within the report give numerous insights into the variables connected to low and high results. This information can be used as a basis for planning the necessary changes in the country. The first task is to find the potentially lowest level schools and to try to find out how to help them to raise their standards. Several check lists are introduced in the report to detect these schools. More research is needed, not only into the processes in the schools but also how the achievement levels are changing over time. The NASA 2011 dataset can be used as a basis for further studies when comparing, for example, the institutional schools with the best performing community schools. The next dataset, NASA 2012, concentrates on grade 3 and 5 students. New and interesting knowledge can be acquired of the beginning of the students' academic path. NASA 2013 will give comparable results to those of grade 8 students. Hence, NASA 2011 results can be taken as a preliminary, or as a base line against which to measure changes in the country.

Based on the findings of the study every stakeholder has to identify their roles themselves. The roles and responsibilities developed from one place or at a central level could just be symbolic. The answers to the queries like "*What can I do with the results of the study?*" and "*What is the solution?*" can be sought in the report. Persons, organizations and activities have to be listed and an action plan has to be prepared according to these, then that plan can be brought in to implementation. Other agencies can play the role of supporting or monitoring and evaluation. In this way, expected improvement can be made gradually.

References for the whole book

Acharya, S. (2007). *Social inclusion: Gender and equity in education SWAPs in South Asia, Nepal case study.* Kathmandu: UNICEF, Regional Office for South Asia.

Acharya, S. (2011). Interview 16.12.2011

Acharya, S. (2012). Email communication 13.3.2012

Acharya, S. & Giri, D. (2009). *Measuring learning achievement of Nepali and non Nepali speaking students of Dang district.* Kathmandu: Save the Children Japan.

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

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency.* London: Continuum.

Angoff, W. H. (1971). Scales, Norms, and Equivalent Scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd Edition), Washington, D.C.: American Council of Education. 508–600.

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

Aryal, H. (2012). Email communication 13.3.2012

Awasthi, L. (2011). Interview 23.12.2011

Baker, C. & Prys Jones, S. (1998). *Encyclopedia of Bilingualism and Bilingual Education.* Multilingual Matters Ltd., Clevedon.

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

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

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

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

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

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

Brandon, P. (2002). Two versions of Contrasting-Groups Standard-Setting Method: A Review. *Measurement and Evaluation in Counseling and Development*, **35**, 167–181.

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

Bruner, J. S. (1996). The Culture of Education.Cambridge, London: Harvard University Press.

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

Caddle, M. (2009). Private schools as battlefields: Contested visions of learning and livelihood in Nepal. In Promod B., *Education in Nepal: Problems, reforms and social change*. Kathmandu: Martin Chautari.

Cambridge Education Ltd and METCON Consultants (2009). *Joint Evaluation of Nepal′s Education for All 2004-2009 Sector Program*. Oslo: Norad (Norwegian Agency for Development Cooperation).

CBS (2002). *National population and housing census 2001*. Kathmandu: CBS/ National Planning Commission.

CBS (2011). Population. Retrieved from http://www.cbs.gov.np/ population_projection_nepal.php.

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

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

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

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

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

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., Weinfeld, F. D., & York R. L. (1966). Equality of Educational Opportunity . Washington, DC: US Department of Health, Education & Welfare. Office of Education. U.S. Government Printing Office.

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

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*(10), 15–24.

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

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

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

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

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

Embassy of Finland, Kathmandu (2011). Unpublished background documents of School Sector Reform Project.

Embassy of Finland, Kathmandu (2012). Information on Nepal. Retrieved from http://www.finland.org.np/Public/default.aspx?nodeid=44214&culture=fi-FI&contentlan=1&displayall=1.

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.

Ferrara, S., Perie, M. & Johnson, E. (2002). Matching the Judgmental Task with Standard Setting Panelist Expertise: the item-descriptor (ID) matching procedure. Washington DC: American institutes for Research.

Fink, A. (1995). *The survey handbook*. Thousand Oaks, CA: Sage Publications.

FNBE (2004). *National Core Curriculum for Basic Education 2004*. The Finnish National Board of Education. Helsinki: Yliopistopaino. Retrieved from http://www.oph.fi/english/sources_of_information/core_curricula_and_qualification_requirements/basic_education.

Fullbright (2008). A *final report on national assessment of grade-V students*, Fulbright Consultancy & Cooperative Hands in restoration (CHIRAG).

Kathmandu: Government of Nepal, Ministry of Education and Sports, Department of Education.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43–56.

Greaney, V. & Kellaghan, T. (2008a). *Implementing a national assessments of educational achievement*, Vol 3, the World Bank.

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

Hair, J. F. Jr. , Anderson, R. E., Tatham, R. L., & , W. C. (1998). *Multivariate Data Analysis*, (5th Ed.). Upper Saddle River, NJ: Prentice Hall.

Hakkarainen, K., Lonka, K. & Lipponen, L. (2001). *Tutkiva oppiminen. Älykkään toiminnan rajat ja niiden ylittäminen*. Porvoo: WSOY. [In Finnish]

Hansegård, N.-E. (1968). *Tvåspråkighet eller halvspråkighet? / med ett tillägg "De språkliga minoriteterna"*. Stockholm: Aldus/Bonnier. Serie: Aldusserien, 0346-5454 [In Swedish]

Hassinen, S. (2002). *Simultaaninen kaksikielisyys. Läheiset sukukielet viro ja suomi rinnakkain*. Doctoral Thesis. Humanistinen tiedekunta, Suomen ja saamen kielen ja logopedian laitos, Oulun yliopisto. [In Finnish]

Hautamäki, J. &Kuusela, J. (2005). Measuring the children, infering the school.In H.K. Lyytinen & A. Räisänen (eds), *Kehittämissuuntaa arvioinnista*. [Directions to develop in Evaluation] Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä: Jyväskylänyliopistopaino. pp. 229–242. [In Finnish]

Heikkinen, H. L. T. (2000). Opettajan ammatin olemusta etsimässä. In K. Harra (Ed.), *Opettajan professiosta*. Artikkelisarja. OKKA -säätiö. Saarijärvi: Saarijärvi Offset Oy. pp. 8−19. [In Finnish]

Hirsjärvi, S. & Hurme, H. (2004). *Tutkimushaastattelu: teemahaastattelun teoria ja käytäntö*. Helsinki : Helsinki University Press. [In Finnish]

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.

Impara, J. C & Plake, B. S. (1997). Standard Setting: An Alternative Approach. *Journal of Educational Measurement*, **34**(4), 35–56.

Impara, J. C. & Plake, B. S. (1998). Teachers' Ability to Estimate Item Difficulty: A Test of the Assumptions in the Angoff Standard Setting Method. *Journal of Educational Measurement*, **35**(1), 69–81.

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational Measurement*. (3rd Ed.), Washington, D.C.: Americal Council of Education. 485–511.

Jencks, C., Smith, M., Ackland, H., Bane, M.J., Cohen, D., Ginter, H., Heyns, B. & Michelson, S. (1972). *Inequity: A reassessment of the effect of the family and schooling in America*. New York: Basic Books.

Jyrhämä, R. (2002). Ei kysyvä tieltä eksy. Pedagogisen ajattelun kehitysvaiheita. In P. Kansanen & K. Uusikylä (Eds.), *Luovuutta, motivaatiota, tunteita. Opetuksen tutkimuksen uusia suuntia*. Jyväskylä: PS-kustannus. (pp. 73–95). [In Finnish]

Kadijevich, Dj. (2006). Developing trustworthy TIMSS background measures: A case study on Mathematics attitude. *The Teaching of Mathematics*, **9**(2), 41−51. Retrieved from http://elib.mi.sanu.ac.yu/journals/tm/17/tm924.pdf.

Kadijevich, Dj. (2008). *TIMSS 2003: Relating Dimensions of Mathematics Attitude to Mathematics Achievement*. Retrieved from http://www.ipisr.org.rs/Dokumenta/zbornici%20PDF/Zbornik%202008%2040-2/Zbornik-08-2-006.pdf.

Kaftandjieva, F. (2004). Standard Setting. Reference Suplement, Section B in S. Takala (Ed.) (2009). Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Language Policy Division, Strasbourg.

Kansanen, P. (1995a). Kuorolaulu, lausunta, kansantanhut ja tanssi. Entä pedagoginen ajattelu. In A. Vaahtokari & A. Vähäpassi (Eds.) *Tutki, vertaile ja arvioi. Näkökulmia opetuksen suunnitteluun ja tutkimukseen*. Helsingin yliopiston Lahden tutkimus- ja koulutuskeskus. (pp. 9-23). [In Finnish]

Kansanen, P. (1995). Teachers' pedagogical thinking – What is it about? In C. Stensmo & L. Isberg (eds.) *Omsorg och engagemang. Pedagogiska institutionen*, Uppsala universitetet.

Kansanen, P. (2003). Teacher education on Finland: current models and new developments. In M. Moon, L. Vlãsceanu, & C. Barrows (Eds.), *Institutional Approaches to teacher Education Within Higher Education in Europe: Current Models and New Developments*. Bucharest: Unesco-Cepes. (pp. 85–108).

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, **29**(2), 119–127.

Kingston. N., Kahl, S. R., Sweeney, K. & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizeck (Ed.), *Setting*

Performance Standards: Concepts, Methods, and Perspectives. Mahwah, NJ: Lawrence Erlbaum. 219–248.

Knapp, T. R. & Brown, J. K. (1995). Ten Measurement Commandments That Often Should Be Broken. *Research in Nursing & Health*, **18**, 465–469.

Koirala, B. N. & Acharya, S. (2005). Girls in science and technology education: A study on access, participation, and performance of girls in Nepal. UNESCO: Kathmandu.

Krieger, N., Williams, D. R & Moss, H. W. (1997). Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu. Rev. Public Health* **18**, 341–78.

Kumpulainen T. (ed,) (2011). *Opettajat Suomessa 2010.* Koulutuksen seurantaraportit 2011:6. Opetushallitus. Retrieved from www.oph.fi/ download/131532_Opettajat_Suomessa_2010.pdf. [In Finnish]

Lampinen, A. & Korhonen, H. (2010). Suomessa opitaan matematiikkaa Varga-Neményi-menetelmän mukaan. *Dimensio* **2**/2010, 24–28. [In Finnish]

Landau, B. (2009). Classroom management. In L. J. Saha & A. G. Dworkin (Eds.), *International Handbook of Research on Teachers and teaching.* New York: Springer. (pp.739–753).

Lauriala, A. (2000). Opettajan ammatillinen uudistuminen: sociokulttuurinen näkökulma opettajan oppimiseen. In K. Harra (Ed.), *Opettajan professiosta.* Artikkelisarja. OKKA -säätiö. Saarijärvi: Saarijärvi Offset Oy. (pp. 8–19). [In Finnish]

Liberatos, P., Link, B. G., Kelsey, J. I. (1988). The measurement of social class in epidemiology. *Epidemiol. Rev.* **10**, 87–121.

Liu S (2006). *School effectiveness research in China.* An Unpublished Doctoral dissertation at Louisiana State University. Retrieved from http://etd.lsu.edu/ docs/available/etd-10222006-172424/unrestricted/Liu_dis.pdf.

Livingstone, S. & Zieky, M. (1982). Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests. Princeton, NJ: ETS.

Lockheed, M. E., Verspoor, A. M., and associates (1996). *Improving Primary Education in Developing Countries.* New York. Oxford University Press.

Loomis. S. C. & Bourque, M. L. (2001). From tradition to innovation: Standard-setting on the National Assessment of Educational Progress. In G. J. Cizeck (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives.* Mahwah, NJ: Lawrence Erlbaum. (pp. 175–218).

Lord, F.M. (1952). The relationship of the reliability of multiple-choice test to the distribution of item difficulties. *Psychometrika,* **18**, 181–194.

McLoyd, V. C. (1997). The impact of poverty and low socioeconomic status on the socioemotional functioning of African-American children and adolescents: mediating effects. In R. D. Taylor & M. **Wang** (Eds), S*ocial and Emotional Adjustment and Family Relations in Ethnic Minority Families*. Mahwah, NJ: Erlbaum. (pp. 7–34).

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology*. 4th edition. Harcourt Brace College Publishers, Fort Worth.

Metfessel, N., Michael, W. B., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the Writing of behavioral Objectives. *Psychology in the Schools*. **6**, 227–231

Metsämuuronen, J. (2002). *Small book of test theory and test construction*. Helsinki: International Methelp Oy.

Metsämuuronen, J. (2009a). *Methods assisting the assessment*. Oppimistulosten arviointi 1/2009. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

Metsämuuronen, J. (2009b). *Ratkaisuprosentin muuntaminen taitotasoksi – A-Finska-oppiaineen aineiston erityiskysymyksiä* [Changing the percentage of correct answers into proficiency level – specific questions of A-Finnish assessment]. Unpublished memorandum 28.12.2009. [In Finnish]

Metsämuuronen, J. (2010a). Changes in Mathematical achievement and attitudes during the first years in compulsory education.In E. K. Niemi & J. Metsämuuronen (Eds.), *Learning Outcomes and Changes in Achievement in Mathematics*. Koulutuksen seurantaraportit 2010:2. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

Metsämuuronen, J. (2010b). Omvandlingen av poängtal till färdighetsnivåer inom det receptive delområdet av provet i A-lärokursen och i den modersmålsinriktade lärokursen. [Changing the percentage of correct answers into proficiency level in Finnish language for Swedish speakers A-course and native speakers' course test]. Appendix VII in O. Toropainen (2010). *Utvärdering av läroämnet finska i den grundläggande utbildningen. Inlärningsresultaten i finska enligt A-lärokursen och den modersmålsinriktade lärokursen i årskurs 9 våren 2009*. Uppföljningsrapporter 2010:1. Utbildningsstyrelsen. Helsinki. Retrieved from http://www.oph.fi/publikationer/2010/utvardering_av_laroamnet_finska_i_den_grundlaggande_utbildningen. [In Swedish]. (pp. 164–168).

Metsämuuronen J (2012a). Challenges of the Fennema-Sherman Test in the International Comparisons. *International Journal of Psychological Studies,* **4**(3) September, 1–22. http://www.ccsenet.org/journal/index.php/

ijps/article/view/16904/12480.

Metsämuuronen J (2012b). Comparison of Mental Structures of Eighth-Graders in Different Countries on the basis of Fennema-Sherman test**.** *International Journal of Psychological Studies,* 4(4) December, 1–17. Available at http://www.ccsenet.org/journal/index.php/ijps/article/view/17305/14331.

Metsämuuronen, J. (2013a). A new method to setting standard for the wide range of language proficiency levels. *International Education Research* 1(1). Available at http://www.todayscience.org/IER/v1/Jari%20Metsamuuronen.pdf.

Metsämuuronen, J. (2013b). *Handbook of Basics of Research Methods in Human Sciences*. International Methelp Oy. Available as E-book at www.booky.fi.

Metsämuuronen, J. (2013c). *Changes in student achievement from grade 3 to grade 9*. Koulutuksen seurantaraportit 2013:4. Opetushallitus. Helsinki: Edita Prima. [In Finnish].

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).

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*. Mahwah, NJ: Lawrence Erlbaum. (pp. 249–281).

MLE Research & Review team (2009). Report on MLE policy and strategy based on the Analysis of Mother tongue Speaking Children's Learning Environment in Kanchanpur, Palpa, Rasuwa, Dhankuta, Sunsari and Jhapa. Sanothimi: Inclusive Education Section/Department of Education.

MOE (2009). *The school sector reform plan, 2009–2015*. Kathmandu: The government of Nepal, Ministry of Education MOE.

MOE (2012). *Flash I report, 2012*. Kathmandu: Ministry of Education.

MOES (2002). *Secondary education support programme 2003*. Kathmandu: Ministry of Education and Sports, DANDDA/DOE

MOES (2003). *Education for all 2004–2009: Core document*. Kathmandu: Ministry of Education and Sports/DOE

Monks, C. P., Robinson, S,. & Worlidge, P. (2012). The emergence of cyberbullying: a survey of primary school pupils' perceptions and experiences. *School Psychology International*, **33** (5), 477–491. http://dx.doi.org/10.1177/0143034312445242

Muijs, D. & Reynolds, D. (2005). *Effective teaching – evidence and practice*. Gateshead: Athenaeum Press Ltd.

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.

Murphy, J. (2010). *The educator's handbook for understanding and closing achievement gaps.* . Thousand Oaks, CA: Corwin Press. Retrieved from http://knowledge.sagepub.com/view/the-educators-handbook-for-understanding-and-closing-achievement-gaps/SAGE.xml

Nedelsky, L. (1954). Absolute Grading Standards for Objective Tests. *Educational and Psychological Measurement*, **14**(1), 3–19.

Nepal in Figures (2008). Retrieved from http://www.cbs.gov.np/Nepal%20in%20figure/Nepal%20in%20Figures%20%20eng%202008.pdf.

Niemi, E. K. & Metsämuuronen, J. (Eds.) (2010). *Learning Outcomes and Changes in Achievement in Mathematics*. Oppimistulosten arviointi 1/2010. Opetushallitus. Helsinki: Yliopistopaino. [In Finnish].

Niemi, H. & Jakku-Sihvonen, R. (2006). Research-based teacher education in Finland. In R. Jakku-Sihvonen & H. Niemi (Eds.), *Research-Based Teacher Education in Finland – Reflections by Finnish Teacher Educators*. Turku: Finnish Educational Research Association. (pp. 31–51).

Niemi, H. & Jakku-Sihvonen, R. (2011). Teacher education in Finland. In M. Valenèiè Zuljan & J. Vogrinc (Eds.), *European Dimensions of Teacher Education: Similarities and Differences*. Slovenia: University of Ljubljana & The National School of Leadership in Education. (pp. 33–51).

Niemi, H. (2010). Teachers as high level professionals – What does it mean in teacher education? Perspectives from the Finnish teacher education. In K. G. Karras & C. C. Wolhuter (Eds.), *International Handbook of Teacher Education: Issues and Challenges* (Vol. l & ll). Athens Greece: Atrapos. (pp. 237–254).

Niemi, H. (2011). Educating student teachers to become high quality professionals – A Finnish case. *Center for Educational Policy Studies Journal*, **1**(1), 43–66.

Niemi, H. (2012). The societal factors contributing to education and schooling in Finland. In H. Niemi, A. Toom, & A. Kallioniemi (Eds.), *The Miracle of Education: The Principles and Practices of Teaching and Learning in Finnish Schools*, Rotterdam: Sense Publishers. (pp. 19–38).

Nunnally, J. C. (1978). *Psychometric theory*. (2nd ed.) New York: McGraw-Hill.

Oaks M. (2011). *Measuring Socioeconomic status*. In e-Source. Behavioral & Social Science Research. Retrieved from http://www.esourceresearch.org/Portals/0/Uploads/ Documents/Public/Oakes_FullChapter.pdf.

OECD (2003). *Learning for tomorrow's world*: First results from PISA. OECD, Paris.

OECD (2007). The Program for International Student Assessment (PISA) 2006: Science Competencies for Tomorrow's World Executive Summary. OECD, Paris.

OECD (2008, January). *Ten steps to equity. Policy Brief*. Retrieved from http://www.oecd.org/publications/Policybriefs

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.) (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://pirls.bc.edu/timss2007/mathreport.html.

Palonen, J., Kurnitski, J., Haverinen-Shaughnessy, U., & Shaughnessy, R. (2009). Preliminary results from Finnish primary schools' ventilation system performance study. In *Proceedings of Healthy Buildings*, Syracuse, USA.

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.

Parajuli, K. (2008). The Valley's Relentless Growth. *Himal South Asian*, **21**(10/11), 34–41.

Parajuli, M & Acharya, S. (2008). Measuring gender equality in education. A Micro-study of learning environment at home and school: Through the perspective of gender equality. UNESCO: Kathmandu.

Patrikainen, R. (2000). *Opettajuuden laatu ja opettajan ammatin professiomuutos*. In K. Harra (Ed.), Opettajan professiosta. Artikkelisarja. OKKA -säätiö. Saarijärvi: Saarijärvi Offset Oy. (pp. 20¯31). [In Finnish]

PEDP (1998). *Primary Education Development Project: An assessment*. 1998. Kathmandu.

PISA (2001). Knowledge and Skills for Life.First results from PISA 2000. Paris: OECD.

PISA (2003). Education at a Glance. OECD Indicators 2003. Paris: OECD;

PISA (2003b). *PISA School Questionnaire*. Retrieved from http:// pisa2003.acer.edu.au/downloads.php.

PISA (2005). *Equity in Education.Thematic review.Finland, Country note*. Retrieved from http://www.oecd.org/dataoecd/49/40/36376641.pdf

PISA (2006a). *PISA Student Questionnaire*. Retrieved from http:// pisa2006.acer.edu.au/downloads.php.

PISA (2006b). *PISA School Questionnaire*. Retrieved from http:// pisa2006.acer.edu.au/downloads.php.

PISA (2006c). *PISA Released Items – Reading*. OECD

PISA (2007). *PISA 2006 Results*. Retrieved from http://www.oecd.org/ document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html

PISA (2009). *PISA 2009 Results*. Retrieved from http://stats.oecd.org/ PISA2009Profiles/#

PISA (2009). Take the Test. Sample Questions from OECD's PISA Assessments. OECD.

PISA (2010a).PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Vol. I). Paris: OECD.

PISA (2010b) *Finland: Slow and Steady Reform for Consistently High Results*. Retrieved from www.oecd.org/dataoecd/34/44/46581035.pdf

Rana, N. (2006). *The Experience of Bing Victims of School Bullying: A Phenomenological Study* [Unpublished Ph.D Thesis]. Kathmandu: Kathmandu University

Reckase, M. D. (2000). A Survey and Evaluation of Recently Developed Procedures for Setting Standards on Educational Tests. In M. L. Bourqey & Sh. Byrd (Eds), *Student Performance Standards on the National Assessment of Educational Progress: Affirmations and Improvement*. Washington, DC: NAEP. (pp. 41−70).

Reynolds, D., Teddlie, C., Creemers, B., Scheerens, J., & Townsend, T. (2000). An introduction to school effectiveness research. In C. Teddlie & D. Reynolds (eds.), *The international handbook of school effectiveness research.* London: Falmer. (pp. 3−25).

Rhem, J. (2010). *Quotations on Teaching, Learning, and Education*. Retrieved from www.ntlf.com/html/lib/quotes.htm.

Sahlberg, P. (2011a). The Professional Educator: Lessons from Finland. *American Educator,* **35**(2), 34−38.

Sahlberg, P. (2011b). Lessons from Finland: Where the Country's Education System Rose to the Top in Just a Couple Decades. *Education Digest*, **77**(3), 18−24.

Salkind, N. J. (1997). *Exploring research.* (3rd Ed.) Upper Saddle River, NJ: Prentice Hall.

Santona Memorial Academy (2011). Role of resource center for improving quality education in schools. Kathmandu.

Scheerens, J. (2000). *Improving school effectiveness*. In International Institute for Educational Planning, Fundamentals of Educational Planning, 68. Paris: IIEP.

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.

Schleicher, A. (2011). Is the Sky the Limit to Education Improvement? *Phi Delta Kappan*, **93**(2), 58−63.

Shen, C. & Tam, H. P. (2008). The Paradoxical Relationship between Student Achievement and Self-Perception: A Cross-National Analysis Based on Three Waves of TIMSS Data. *Educational Research and Evaluation*, **14**(1), 87−100.

Shen, C. (2002). Revisiting the relationship between students' achievement and their self-perceptions: A cross-national analysis based on the TIMSS 1999 data. *Assessment in Education*, **9**(2), 161−184.

Skar, H. O. & Cederroth, S. (1997). *Development aid to Nepal: issue and option in energy, health, education, democracy and human right*. Copenhagen: Nordic Institute of Asian Studies (NIAS).

Skutnabb-Kangas, T. (1981). *Bilingualism or Not: The Education of Minorities*. Multilingual Matters, Clevedon.

Statistics Finland (2011). *Internet tables*. Retrieved from http://pxweb2.stat.fi/ Dialog/varval.asp?ma=020_vaerak_tau_101_fi&ti=Kansalaisuus+i%E4n+ ja+sukupuolen+mukaan+maakunnittain+1990+%2D+2011&path=../ Database/StatFin/vrm/vaerak/&lang=3&multilang=fi. [In Finnish]

Statistics Finland (2012a). [*Internet issue***].** Retrieved from http://pxweb2.stat.fi/ Database/StatFin/kou/kjarj/kjarj_fi.asp

Statistics Finland (2012b) *Esi- ja peruskouluopetus* [internet issue]. ISSN=1799-3709. Helsinki: Tilastokeskus Retrieved from http://www.stat.fi/til/pop/ index.html. [In Finnish]

Stevenson, H. W. (1998). A Study of Cultures: Germany, Japan and the United

States - An Overview of the TIMSS Case Study Project. *Phi Delta Kappan*, **79**(7), 524−529.

Syrjäläinen, E., Jyrhämä, R. & Haverinen, L. (2009). *Praktikumikäsikirja*. Helsinki: Studia Pædagogica 33. Original 2004, Reneved 2009. Retrieved from http://www.helsinki.fi/behav/praktikumikasikirja/lukuIII/index.htm. [In Finnish]

Takala, S. & Kaftandjieva, F. (2009). Reseptiivisen kielitaidon kokeen muuntaminen taitotasoiksi. [Changing the receptive language test to proficiency levels]. In E Tuokko, *Miten Ruotsia osataan peruskoulussa? Perusopetuksen päättövaiheen ruotsin kielen B-oppimäärän oppimistulosten kansallinen arviointi 2008*. [How Swedish language is spoken in general education? National assessment 2008 of the student achievement in Swedish language B-course at the end of basic education]. Oppimistulosten arviointi 2/2009. Opetushallitus. Helsinki: Edita Prima Oy. 118. Retrieved from http://www.oph.fi/julkaisut/2009/ miten_ruotsia_osataan_peruskoulussa [in Finnish].

Takala, S. (Ed.) (2009). Relating Language Examinations to the CommonEuropean Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Language Policy Division, Strasbourg.

Teddlie, C., Reynolds, D. & Sharon, P. (2000). Current topics and approaches in School Effectiveness Research: The contemporary field. In C. Teddlie & D. Reynolds (eds.), *The International Handbook of School Effectiveness Research*. London: Falmer. (pp. 26-51).

Thapaliya, T. P. (2011). *Leadership Process and School effectiveness*. An unpublished Ph. D. Dissertation submitted to the School of Education, Kathmandu University, Nepal.

Tikkanen, P. (2008). "*Helpompaa ja hauskempaa kuin luulin": matematiikka suomalaisten ja unkarilaisten perusopetuksen neljäsluokkalaisten kokemana*. Jyväskylä studies in education, psychology and social research, 337. Jyväskylä University. Retrieved from http://urn.fi/URN:ISBN:978-951-39-3247-3. [In Finnish]

TIMMS & PIRLS (2008). Appendix C: *The test-curriculum matching analysis*. Trends in Mathematics and Science Study (TIMMS) & Progress in International Reading Literacy Study (PIRLS) International Study Center. Retrieved from: http://timss.bc.edu/timss2007/PDF/ T07_M_IR_AppendixC.pdf

TIMSS (1999). Mullis, I,V.S., Martin, M.O, Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J. & Smith, T. A. (1999). *TIMSS 1999 International Mathematics Report. Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth*

Grade. Boston College, Lynch School of Education. Retrieved from http://www.timss.org/timss1999i/math_achievement_report.html.

TIMSS (2003). *TIMSS Contextual Background Questionnaires*. Retrieved from http://timss.bc.edu/timss2003i/context.html.

TIMSS (2006). *Questionnaires*. Retrieved from http://nces.ed.gov/timss/questionnaire.asp

TIMSS (2007). TIMSS *2003 Science Items. Released Set, Eight grade*. IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2009a). *TIMSS 2007 User Guide for the international Database. Released Items, Mathematics – Eight grade*. IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2009b). *TIMSS 2007 User Guide for the international Database. Released Items, Science – Eight grade*. IEA, TIMSS & Pirls International Study Center. Lynch School of Education, Boston College.

TIMSS (2011). Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Toropainen, O. (2010). Utvärdering av läroämnet finska i den grundläggande utbildningen. Inlärningsresultaten i finska enligt A-lärokursen och den modersmålsinriktade lärokursen i årskurs 9 våren 2009. Uppföljningsrapporter 2010:1. Utbildningsstyrelsen. Helsinki. Retrieved from http://www.oph.fi/publikationer/2010/utvardering_av_laroamnet_finska_i_den_grundlaggande_utbildningen. [In Swedish]

Tuokko, E. (2009). Miten Ruotsia osataan peruskoulussa? Perusopetuksen päättövaiheen ruotsin kielen B-oppimäärän oppimistulosten kansallinen arviointi 2008. [How Swedish language is spoken in general education? National assessment 2008 of the student achievement in Swedish language B-course at the end of basic education]. Oppimistulosten arviointi 2/2009. Opetsuhallitus. Helsinki: Edita Prima Oy. Retrieved from http://www.oph.fi/julkaisut/2009/miten_ruotsia_osataan_peruskoulussa [in Finnish]

Tynjälä, P. (2002). Oppiminen tiedon rakentamisena. Konstruktivistisen oppimiskäsityksen perusteita. Tampere: Tammi. [In Finnish]

UNDP (2010). *Nepal Millennium Development Goals*. Retrieved from http://www.beta.undp.org/content/dam/aplaws/publication/en/publications/MDG/nepal-mdg-report-2010/Nepal%20MDG%20report%202010.pdf.

van der Schoot, F. (2009). Cito variation of the Bookmark Method. Reference Supplement, Section I. In S. Takala (ed.), Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Language Policy Division, Strasbourg.

Varga-Nemenyi Association (2012). *Internet material*. Retrieved from http://www.varganemenyi.fi/includes/menetelma.php. [In Finnish]

Verhelst, N.D., Glas, C.A.W., Verstralen, H.H.F.M. (1995). *One-Parameter Logistic Model OPLM*. Cito, Arnhem.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychol. Bull.*, **91**, 461–81.

Wilkins, J. L. M. (2004). Mathematics and science Self-Efficacy: An international investigation. *The Journal of Experimental Education*, **72**(4), 331−346.

World Bank (1996). National assessments: Testing the system. Washington: The International Bank for Reconstruction and Development. The world Bank.

Wright, B. D. (1968). *Sample-free test calibration and person measurement*. Proceedings of the 1967 Invitational Conference on testing Problems. Educational Testing Service, Princeton, NJ.

Yonjan-Tamang, A. (2009). *Linguistic Heritage of Nepal*. Multilingualism for Social Justice, Harmony and Peace. [Presentation] Kathmandu University, Kathmandu. MPhil/PhD in Education.

Index for the whole book

E

F

G

M