National Assessment of Student Achievement 2020

Main Report

Report on National Assessment of Student Achievement in Mathematics, Science, Nepali and English for Grade 8

Government of Nepal
Ministry of Education, Science and Technology
Education Review Office (ERO)
Sanothimi, Bhaktapur

National Assessment of Student Achievement 2020

Main Report

Report on National Assessment of Student Achievement in Mathematics, Science, Nepali and English for Grade 8

Writers

Dr. Bishnu Khanal,
Dr. Aita Bishwakarma
Mr. Shyam Prasad Acharya
Mr. Deviram Acharya
Mr. Lav Dev Bhatta

Data Analysis

Mr. Shyam Prasad Acharya Mr. Deviram Acharya,
Mr. Lav Dev Bhatta and Mr. Parsuram Tiwari

Language Edition

Mr. Chandra Kanta Bhusal
Mr. Lesley Rhodes
Mr. Parsuram Tiwari

Advanced Sampling

Mr. Shyam Prasad Acharya

Government of Nepal
Ministry of Education, Science and Technology
Education Review Office (ERO)
Sanothimi, Bhaktapur

National Assessment of Student Achievement 2020: Main Report
Report on National Assessment of Student Achievement in Mathematics, Science, Nepali and English for Grade 8

Writers
Dr. Bishnu Khanal, Dr. Aita Bishwakarma, Mr. Shyam Prasad Acharya, Mr. Deviram Acharya, Mr. Lav Dev Bhatta

Data Analysis
Mr. Shyam Prasad Acharya, Mr. Deviram Acharya, Mr. Lav Dev Bhatta and Mr. Parsuram Tiwari

Language Edited
Mr. Chandra Kanta Bhusal, Mr. Parsuram Tiwari, Mr. Lesley Rhodes

Advance Sampling
Mr. Shyam Prasad Acharya

Year of Publication: 2022 (2079 BS)

Published by:
Education Review Office
Sanothimi, Bhaktapur, Nepal
Phone: +977-1-6636518, 6634362
Email: eronasa@gmail.com
Website: www.ero.gov.np

Printed at:
Omega Zone Services Pvt.Ltd.
Bagbzar, Kathmandu, Nepal

FOREWORD

National Assessment of Student Achievement (NASA) communicates the status of learning achievement of students and suggests measures for improving learning. The assessment provides evidence to the policymakers to formulate practical and implementable educational policies at the national and sub-national level for the needed educational reforms. The NASA is a curriculum-based systematic evaluation of student learning outcomes by using standardized tools.

This is the fourth cycle report of grade 8 among eight large-scale national assessments conducted by the Education Review Office. The previous rounds of assessment for grade 8 were carried out in 2011, 2013 and 2017. In this assessment, Mathematics, Nepali, Science and English subjects have been assessed, based on the standardized test booklets. Those test booklets were developed based on the subject-specific assessment frameworks developed in line with the respective subject National Curriculum of Nepal. This report of NASA 2020 stands for Grade 8 in Mathematics, Science, Nepali and English subjects based on the response's of a national representative sample of 43497 students from 1800 schools in Nepal with an almost equal number of schools and students in each of the four subjects, considering seven provinces as the explicit strata. Five versions of standardized tests together with the background information questionnaire to the sample students, teacher questionnaire to subject teachers, and school survey questionnaire to the headteachers were administered in each school. Data were analyzed to present both overall mean score and proficiency levels, and the relation between the achievement scores and various influencing factors with the use of the background information questionnaire. Analysis and comparison of the results were produced using the Item Response Theory (IRT) and the parameters of linking items. This report is the first report in the School Sector Development Program 2016-2022 period which presents the comparative results of NASA 2020 with NASA 2017 in two subjects Mathematics and Science by using the linking items. However, the results of NASA 2020 for Nepali subject are not comparable because of the lack of sufficient linking items between NASA 2017 and NASA 2020. In case of English, large-scale assessment in English was not carried out before. The results presented in this report are generalized results over the defined population and they provide evidence of the level of learning.

I would like to acknowledge the contribution of teachers, experts, members of subject committees, and researchers throughout the process of framework and tool

development, test administration, data analysis and report writing. My sincere thanks goes to previous Director Generals; consulting firm Summunat Nepal for administrating the test administration. NASA unit chief director Narayan Prasad Jha, and other directors as well as other staffs of ERO for their direct and indirect involvement in various phases of this assessment. My sincere thanks also goes to World Bank for providing experts to write the report. Similarly, I would like to thank British Council Nepal for providing technical and financial support to develop and carry out listening and speaking test as wall as technical support in developing reading and writing test in English subject. I also acknowledge the supportive and coordinating role of Dr. Uttam Sharma and Karthika Radhakrishnanin from the World Bank in various phases of training and report writing. I would like to thank to Dr. Hari Lamsal and Deepak Sharma and Shreeprasad Bhattarai for the valuable comments on the report. I highly appreciate the contribution of previous director generals of ERO for initiating NASA activities, Central Level Agencies and the Ministry of Education, Science and Technology for regular support in budgeting, monitoring of test administration and tool development for the program. I express my sincere gratitude to Mr. Dvendra Poudel (hon.minster) and Mr. Ramkrishna Subedi (secretary) for providing us valuable suggestions and guidance. It would be impossible to develop items without the restless efforts of subject committees, subject experts and teachers. They are the priorities of my gratitude.

I believe that this report will be a milestone for policymakers, program designers, teachers, educators, researchers and other stakeholder for their role in improving students' learning. I hope this report will be a foundation for bringing about a change in the quality of education at the school level in Nepal.

Mr. Chandra Kanta Bhusal
Director General

ACRONYMS/ABBREVIATIONS

CDC	:	Curriculum Development Centre
CEFR	:	Common European Framework of References
CEHRD	:	Centre for Education and Human Resource Development
CI	:	Confidence Interval
CR	:	Constructed Response
CRT	:	Criteria-Referenced Test
CSS	:	Clustered Sample Size
CTT	:	Classical Test Theory
DEO	:	District Education Office
DOE	:	Department of Education
DPL	:	Defining Proficiency Level
EAP	:	Expected a Posteriori
EDCU	:	Education Development and Coordination Unit
EMIS	:	Education Management Information System
ERO	:	Education Review Office
GPCM	:	Generalised Partial Credit Model
ICC	:	Item Characteristic Curve
ID	:	Identification
IEA	:	International Association for the Evaluation of Education
IRT	:	Item Response Theory
LAF	:	Language Assessment Framework
MC	:	Multiple Choice
MCQ	:	Multiple-Choice Questions
MISD	:	Main Ideas and Supporting Details
MLE	:	Maximum Likelihood Estimation
MOEST	:	Ministry of Education, Science and Technology
MOS	:	Measure of Size
N cases	:	Number of Cases/Students in the Population
NASA	:	National Assessment of Student Achievement
NEB	:	National Examinations Board
NRT	:	Norm-Referenced Test
NU cases	:	Number of Cases/Students in the Sample
NU psu	:	Number of Primary Sample Units (Schools)
OECD	:	Organisation for Economic Co-operation and Development
OMR	:	Optical Mark Recognition

One PM : One Parametric Logistic Model
PCAP : Pan-Canadian Assessment Program
PCM : Partial Credit Model
PISA : Programme for International Student Assessment
PPS : Probability Proportionate to Size
PRC : Printing-Ready Copy
PSU : Primary Sample Unit
PV : Plausible Value
RWGT : Replicable Weight
SAQ : Short-Answer Questions
SE : Standard Error
SES : Socio-Economic Status
SIID : Specific Information and Important Details
SPSS : Statistical Package for Social Science
SR : Selected Response
SRS : Simple Random Sampling
SSDP : School Sector Development Plan
SSRP : School Sector Reform Plan
TAM : Test Analysis Module
TIF : Test Information Function
TIMSS : Trends in International Mathematics and Science Study
TPD : Teacher Professional Development
WLE : Weighted Likelihood Estimation

EXECUTIVE SUMMARY

Context

At the start of 2020, the Education Review Office (ERO) assessed the learning outcomes of grade 8 students in mathematics, science, Nepali and English. The primary objective of this assessment was to prepare baseline data for the School Sector Development Plan (SSDP) as well as compare the learning achievement of 2020 with the previous cycle of National Assessment of Student Achievement (NASA) in 2017 to ensure quality school education. Altogether 43,886 students, 1,800 teachers and 1,800 head teachers from 1,800 schools participated in this assessment. National assessment has been well accepted as a means of measuring the quality of education (ERO, 2019)) that provides both quantitative and descriptive forms of information about student achievement. This is considered as an output of the teaching–learning process and its quality (World Bank, 1996). National Assessment provides basic information for policy makers, politicians and the broader educational community and informs policy makers about the key aspects of the system (Greaney & Kellaghan, 2008b; ERO, 2013). In this context, ERO has its roadmap to conduct two rounds of NASA for grades 5, 8 and 10, to assess the quality of education and trends of learning achievement within the SSDP period. NASA 2020 is the second cycle of assessment for grade 8 in mathematics, Nepali, science and English in the SSDP period.

Objectives of NASA 2020

The main aim of NASA is to provide policy feedback through the assessment of learning and identify the trends of learning over time. NASA 2020 has the following specific objectives:

a) to identify the current level of grade 8 students' achievement in mathematics, science, Nepali and English,

b) to explore variations in students' achievement by gender, province, types of schools, ethnicity, home language and socio-economic status

c) to identify factors that influence student achievement

d) to identify trends in student learning and produce baseline data for comparison in the future

e) to strengthen the capacity of the education system in conducting national assessments

f) to provide the Ministry of Education, Science and Technologywith recommendations for policy formulation to improve quality and ensure equity, particularly in school education.

Methodology

The ERO has emphasised that the methodological procedures used should make results more scientific and valid in both national and global contexts. Three sets of questions with background information were asked in each subject. All sets were linked with anchor items. The ERO has used Item Response Theory (IRT) to analyse the latent ability of students, using various contextual variables to explain the latent traits of the students. NASA 2020 has used advanced procedures to bring rigour to data analysis by generalising the results at national and provincial levels, through seven explicit strata and various other implicit strata. Student learning outcomes were tested in four subjects – mathematics, science, Nepali and English – on a national representative sample of 43,886 students from 1,800 schools of Nepal, with an almost equal number of schools and students in each of the four subjects, considering seven provinces as explicit strata. The multistage sampling strategy – probability proportional to size (PPS) – sampling method was used to draw this large sample. Three versions of standardised tests, together with the background information in the questionnaire to the sample students, the teacher questionnaire to subject teachers and the school survey questionnaire to the head teachers, were administered in each school. Data was analysed to present overall mean score and proficiency levels and to demonstrate the relationship between the achievement scores and various influencing factors identified from the background information questionnaire. Analysis and comparison of the results were carried out using IRT and the parameters of linking or anchor items. Results are presented in a transformed scale of plausible values (PVs), with a mean of 500 and standard deviation of 50. The results presented in this report are the generalised results over the defined population and they provide evidence of the level of learning.

Key features of data analysis procedures are described in the italics in the box below.

The main reporting procedure was quite similar to NASA 2019. Some of the test items used in the 2017 assessment were also used as anchor items in the 2020 assessment. All the items of 2020 have been calibrated by fixing the 2017 parameters of those anchor items. The 2020 results obtained in this way are comparable to the results of 2017. Also, the results from 2020 will either increase to, equal or decrease from the results of 2017. In 2017, the average score of students was 500, but in 2020, it may be different from 500. Item Response Theory has been used to analyse all this. Thus, when using Item Response Theory, 2PL (Two-parameter Logistic Model) was used to analyse items of one point, and GPCM (Generalised Partial Credit Model) was used for questions with two or more points. Test Analysis Module (TAM) was used to analyse this result, where TAM is a package that runs on the R Programming Language, an open-source software.

The latent ability (theta) of each student is obtained from the analysis as stated above. Ten plausible values were calculated to find out the ability of the population on the basis of the latent ability and the dummy variables made on the basis of the variables asked in the students' background questionnaire. After transforming the average of those plausible values to be zero (0), the following formula was used to present the results.

Student Achievement Score = 500 + Average PV ×50

Thus, readers are requested to study results by considering the above-stated methodologies. However, comparison of the results from 2020 with 2017 was carried out in mathematics and science only. Since the number of anchored items available between 2017 and 2020 for Nepali was low, and available anchor items behaved differently in differential item function analysis, comparison of 2020 results with 2017 was not possible. Therefore, the results are fixed at a mean of 500 in Nepali. Similarly, English was assessed for the first time in grade 8 in 2020. Therefore, English achievement was also not compared to any other assessment years, so its mean score is also fixed at 500 in 2020.

Although the assessment results have shown the national average achievement to be 500 in all four subjects, it does not mean that all subjects have been equally learned. This report therefore presents the results in terms of what the students can and cannot perform, the existing gap between the written curriculum and the achieved curriculum, and the number of students who have developed their ability to a minimum competency level. Moreover, the student proficiency level is defined at six levels, namely: below basic, basic, proficient 1, proficient 2, proficient 3 and advanced level.

Findings, Conclusions and Recommendations

Students are struggling to acquire even minimum learning. The majority of students are not able to learn what is taught in all subjects. In fact, the majority of students have achieved or mastered less than 50% of the curriculum in all subjects. Most of students could not solve higher-order thinking items. Conclusions of the NASA 2020 study were similar to the conclusions drawn from previous NASA grade 5 and grade 10 assessments. As a result, it is argued that there are problems in the teaching–learning strategies, remedial actions and the role of head teachers. On average, students in institutional schools have massively outperformed students in community schools. However, it is worth noting that the average scores for students in some community schools were the highest among all schools in all subjects. Deeper analyses of the reasons behind their success should be considered, as they can provide valuable insights and lessons for other community schools and policy makers alike.

A summary of results disaggregated from various categories and influencing variables are presented below:

Province

The comparative study of province-wise achievement of students in mathematics shows variations in the achievement level of the students. Overall, only 37.7% of students have shown performance above the minimum level. The national average of NASA 2017 was 500, whereas the national average score of NASA 2020 is 483, which is a decrease of 17 scale score. The achievement of students in Bagmati (509), Gandaki (497) and Lumbini (485) was, on average, better than other provinces and was above the national average (483) in 2020. Similarly, the national mean of science was 470, which is 30 scale score below the national mean from the NASA 2017 study. Bagmati (492), Gandaki (485), Lumbini (475) and Province 1 (472) were high-performing provinces, whereas as other provinces scored below the national mean in science. The achievements in Nepali of Province 1 (505), Bagmati (511), Gandaki (521) and Lumbini (498) students were distinctly above the national average. The disparity in achievement by province was much wider in English, though. The achievement of Bagmati (529), Gandaki (522) and Province 1 (501) students was above the national average. The performance of Madhesh, Karnali and Sudur Paschim students was lower in all four subjects than the national average.

Age

A distinct variation in achievement was seen by age group as well. Almost all of the students who participated in the assessment were aged between 12 and 16. Students aged between 12 and 16 years participated in the assessments. Among them, students aged 13 years were the highest scorers in all four subjects assessed. Achievement scores for students aged 15 years or more was lower, on average. This result was consistent in all four subjects.

Home language

There was a significant difference in the achievement of the students who use Nepali as their home language compared to the achievement of students who use other languages as their home languages. The gaps between the achievements of students who used Nepali as a home language and students with other languages as their home language were in scale scores of 15 in Nepali, and 13 in English.

School type

The comparative study of achievement showed a vast gap between community schools and institutional schools. The institutional schools topped the community schools in their achievement in all subjects. For instance, the achievement of students in institutional schools was 500 in science, whereas students of community schools scored 463, a significant difference of 37 scale scores between them. Overall, the achievement of community schools was below the national average, whereas the achievement of institutional schools was distinctly above the national average.

Career aspirations

Based on students' future goals, the study showed that students wishing to be doctors/engineers, civil servants or to work abroad, where in-depth learning is required, had higher achievement in subjects like mathematics and science. than students wishing to be farmers, teachers or employees in the private sector.

Parental education

Parents' educational level has a direct positive association with children's achievement in all subjects assessed. Based on achievement, it can be said confidently that the higher the educational qualifications of the father or mother, the greater the scores of the children on average. Educated fathers and mothers contributed significantly to their children's learning achievement, whereas children whose father or mother was illiterate performed comparatively worse. The achievement differs significantly from illiterate to literate parents and less-qualified to higher-qualified parents. This result is consistent with the study carried out by Kainuwa & Yusuf (2013), who stated that children of fathers or mothers with university degrees perform considerably well and get the highest degree in education.

Parental Occupation

While analysing the relationship between parental occupation and student learning, students' performance was highest for those whose parents were teachers. Students whose parents were involved in government jobs, and business also had higher scores. Children whose fathers and mothers were involved in agriculture and households, working in others' homes, and handling only the households, had, on average, lower scores.

Family Size

Family size was also seen to be an important predictor in the learning achievement of students. Students residing in households where the family size was four to five members had higher achievement scores. Beyond that, achievement decreased with additional family members.

Teacher's regularity

The regularity of a teacher in the classroom demonstrates both dedication and awareness of the importance of delivering quality education to shape the bright future of students. Teachers can give an in-depth knowledge of the subject matter and completing the curriculum in good time assists the teacher and is an important predictor of students' achievement. Thus, considering the findings above, teachers who were dedicating all their time to the classroom were successful in improving students' achievement. Meanwhile, students of teachers who would come late and go early, or who did not come to class at all, had disappointing performances.

Interest in subjects assessed

Developing a strong interest in a subject encourages students to work harder in the subject, which helps boost their achievement in that subject. The findings show that the majority of students who enjoyed the different subjects mentioned here wanted to learn and excel in those subjects.

Homework and feedback

Based on the analysis of data, any feedback after homework has boosted students' performance. In addition, feedback given on a regular basis was found to be more helpful. The performance of students who received regular feedback on their homework was higher than those who never received feedback. There is a significant difference in achievement between students who received regular homework and feedback and those who got neither homework nor regular feedback. This difference was 15 scale score in science and 34 scale score in Nepali. The difference in science and Nepali was statistically significant in the mean score, which indicates the importance of receiving homework and feedback regularly.

Findings for mathematics and science

Students in institutional schools perform, on average, much better than community school students in mathematics. Although this is not a causal relationship, there are many who believe that institutional schools are more effective than community schools in improving student learning. Similarly, the relationship between socio-economic status and mathematics scores is positively correlated.

Female students are, on average, faring worse in mathematics than boys, and the difference is both substantial in magnitude and statistically significant. Similarly, student age and mathematics scores are negatively correlated. Compared to Brahman and Chhetri students, Dalit students are performing significantly worse in mathematics. There is an expected positive relationship between the father's education level and the child's achievement in mathematics.

There are some school-level variables that are also important. For example, students in schools where the head teacher is permanent have, on average, higher scores in mathematics. Similarly, students in schools where the mathematics teachers are permanent are also doing, on average, better than students where mathematics teachers are not permanent. This is perhaps an indication that these teachers and head teachers can focus more on teaching or administrative duties and not worry about other aspects related to their tenure status.

Findings for Nepali and English

Students in schools where the head teacher is a secondary-level appointee are performing better than others, and the difference is statistically significant. Similarly, students in schools that have implemented initiatives to reward teachers have also performed better in Nepali. Similarly, English-medium community schools are performing better than community schools where the medium of instruction is Nepali. There is a grey area which deserves to be explored, even though English-medium instruction does not mean a quality education.

With regard to child-level characteristics, female students are performing worse than male students in Nepali, but the magnitude of the difference is substantially lower than in mathematics. Similarly, age of the student and Nepali scores are negatively correlated, a finding consistent with mathematics. There is a positive relationship with regard to having a dictionary and other educational reference books at home. The positive coefficient for having access to a dictionary and other educational reference books may be a proxy for these households prioritising education.

Conclusion

An educational system covers input, process and output in education. Curriculum, pedagogy, teaching and learning practices and assessment are at centre-stage of attention for the formation, implementation, and monitoring and evaluation of educational policies. Rigorous research and evidence-based findings are the pillars for assessing the overall system of education. NASA has been endeavouring to assess the educational output of school education since its establishment as one of its core activities in Nepal.

The main objective of this assessment was to prepare baseline data for the SSDP as well as compare the learning achievement of 2020 with the previous cycle of NASA (2017), to analyse how quality education in the school system has evolved over time. The study, as before, shows variation in the performance of province-level achievement in mathematics, science, Nepali and English. Bagmati, Gandaki and Lumbini are high-performing provinces, whereas Province 1, Madhesh, Karnali and Sudur Paschim are low-performing ones. The disparity seems deeper in gender-based achievement, as boys have performed higher than girls.

The most appropriate age for learning grade 8 seems to be 12 or 13 years (starting grade 1 at age 5 or 6), as students in this age group, on average, achieved higher scores than other age groups. Students older than 14 years score lower, perhaps a reflection that these children are repeating grades or that children, presumably with less conducive learning environments at home, are starting school later.

A substantial difference in achievement has been observed based on the home language. Children whose home language is Nepali scored higher than those whose home languages were other than Nepali. This important finding has a notable influence on the use of classroom pedagogy and the achievement of students, even in earlier grades.

The achievement of institutional schools is comparatively far better than community schools. Despite the investment of huge resources from the government, the achievement of community-school students remained below the average level. Raising the quality of community schools has been one of the greatest challenges.

There is a difference in achievement based on the future goal of children. Students who wished to be teachers, farmers or to work in private businesses have lower levels of achievement compared to those who aspire to be doctors, engineers, civil servants or to work abroad. One could argue that this is partly a reflection of occupations such as doctor, engineer and the civil service being valued by society at the cost of other civilian professions. There is a need for occupations such as farming, teaching and private business to be made respected professional areas.

There is a remarkable difference in the achievement of children from illiterate and literate parents. There is a positive relationship between student achievement and parents with at least grade 8 of education. Similarly, parental profession has a positive influence on the achievement of students. Scores were lower for students whose parents were involved in agriculture, household work and working for other households.

Children from a nuclear family have achieved, on average, higher scores than those from a joint family. Data shows that the greater the number of family members, the lower the achievement of students. Similarly, students with positive attitudes have succeeded in excelling academically by scoring good grades in various subjects. Likewise, teachers who were dedicating all their time to the classroom were successful in improving students' achievements.

Similarly, providing feedback on homework is leading to the improved achievement of students. The availability of a table for study, a separate study room, a computer for schoolwork, internet access, children's magazines, stories/poetry, pictures, a dictionary, reference books, and so on at home contributes to boosting their learning performance. Lastly, a permanent head teacher and teachers are associated with higher achievement scores. Similarly, permanent school buildings and infrastructures also positively influence learning, as shown by the data.

Recommendations

1. **A large number of students are at below-grade level, and an alarming gap exists between the intended and achieved curriculum.**

   While considering the proficiency levels of students in achievement, the results show their low level of ability as: 32.1% in mathematics, 37.7% in science, 58.8% in Nepali and 51.5% in English of students have passed the basic proficiency levels, whereas

67.9% in mathematics, 62.3% in science, 41.2% in Nepali and 48.5% in English have achieved below the basic proficiency level. These achievement levels indicate students' poor competence levels, and only a small number of students have the higher level of proficiency. The majority of students have achieved or mastered less than 50% of the curriculum in Mathematics and Science subjects. This evidence indicates an alarming gap between the intended and achieved curriculum.

Recommendation: The overall gaps of intended and achieved curriculum demand a radical change in the policy, resource management, curricular design and implementation process, and monitoring and evaluation strategies. Policy reformation, allocation of the necessary budget, an activity-based curriculum, emphasis on pedagogical delivery and resource management are some of the strategies the government should implement instantly for removing the gaps between intended and achieved curriculum. Moreover, given that below-grade-level learning is already evident by grade 5, as the previous administration of NASA at grade 5 has amply demonstrated, remedial education should be seriously considered in earlier grades. Furthermore, training curricula for Teacher Professional Development (TPD) should be reoriented to better equip teachers to identify, and provide tailored instruction to, students entering a particular grade with knowledge below that grade level (Schaffner, Glewwe and Sharma, 2020). More specifically, a campaign of ***No child is left below the minimum level of learning*** is highly recommended at the school level. In this campaign, the Curriculum Development Centre (CDC) is advised to start to define the minimum level of learning (learning standards) with technical co-ordination with ERO; the Centre for Education and Human Resource Development (CEHRD) is advised to prepare teacher training guidelines in focus with this campaign; and the National Examinations Board (NEB) is advised to prepare guidelines to evaluate such learning.

2. **There are wide gaps in achievement between provinces.**

The study shows variations in the performance of province-level achievement in mathematics, science, Nepali and English. A huge gap between the high-performing and low-performing provinces in achievement has a scale of 49 in mathematics, 48 in science, 55 in Nepali and 28 in English. Bagmati, Gandaki and Lumbini are high-performing provinces, whereas Province 1, Madhesh, Karnali and Sudur Paschim are low-performing ones.

Recommendation: To address the wide gap between high-performing and low-performing provinces, justified distribution of resources is a necessity. In Province 1, Madhesh, Karnali and Sudur Paschim, policy reformation, special emphasis on budget allocation, development of human resources, contextualisation of the curriculum and close monitoring and evaluation of educational programmes are suggested areas of primary intervention by the government. A minimum standard of infrastructure, learning opportunities, resources, incentives, and retention of good teachers and identification of learning difficulties along with remedial teaching, are supportive activities to enhance learning and increase students' achievement. Specific curricula and instruction methods that can be embodied in daily teaching guides and related instructional materials can be developed, and distribution of these guides and materials and the teacher training can be packaged together to improve student learning (Schaffner, Glewwe and Sharma, 2020). In addition, small-scale policy experiments should be designed and analysed to help improve the implementation aspects so that programmes have a high success probability.

3. **here is a huge disparity in achievement by type of school.**

A huge disparity in achievement between community and institutional schools may create a two-tiered society in the future. A huge gap in achievement is seen between institutional and community schools, with a range of scale score of 37 in science, as an example, and the range of differences are similar in the other subjects too.

Recommendation: The gap should be eliminated by upgrading community schools through strategic interventions in school education. It is imperative to identify faults in the input, process and output of community school mechanisms and reform policy to improve the current situation. A comprehensive analysis of better-performing institutional and community schools is sorely needed to explore how poor-performing community schools can be improved. Local governments also have an important role to play in improving the quality of public education.

4. **The home language used also brings about a remarkable gap in achievement.**

A remarkable gap in achievement has been revealed due to the home language used, with ranges in scale score of 15 in Nepali and 22 in English between high achievers and low achievers.

Recommendation: This gap can be narrowed by teachers using the home language of children, even in the earlier grades. Teachers need at least a basic-level language learning package in their students' language or the language of the community surrounding the school. Teachers should be able to communicate in the community language, and they have to teach translating, code-switching and using trans-language strategies to empower those children who use languages other than Nepali at home. A comprehensive language learning package for teachers for their professional development deserves incorporation into TPD.

5. **here is a noticeable gap in the learning achievement of boys and girls.**

The results revealed that boys scored higher than girls. The study shows a noticeable disparity between boys and girls in their achievement. The gap ranges in scale scores of 12 in mathematics, 5 in science and 10 in English, though normally there is no gap in Nepali. This indicates that work is needed on gender equity in learning achievement.

Recommendation: The reasons behind such disparity in learning between boys and girls are worth exploring further so that effective interventions to reduce gender differences in learning can be devised. Suggested interventions include teachers paying attention to student-friendly behaviour (more focused on girls) and teaching and learning activities in the classroom, including remedial education. Affirmative action such as scholarships and additional incentives to girls may reduce gender disparity in achievement. Regular interactions with female role models may also help. Apart from these, teachers should create a suitable learning environment for girls by being sensitive in terms of their needs, interest, voices and providing equal opportunity for classroom participation. Parents are to be encouraged for their roles in supporting equality in their children's education.

6. **Students at the appropriate age performed better.**

Students studying in grade 8 at the ages of 12 and 13 scored higher than underage and overage students studying at the same level. The similarity in the age group among students may have encouraged them to share and discuss their education-related problems, thereby enabling them to excel academically. The gap in the achievement of students aged 12 or 13 compared to other age groups has been in scale scores of 25 in mathematics, 19 in science and 24 in Nepali.

Recommendation: If the student is below age 14 while in grade 8, the child was in grade 1 at or before age 4. Similarly, if the child is aged 15 or above in grade 8, it is most likely an indication that they have repeated grades or started grade 1 in a less conducive environment. In addition to encouraging children to enrol on time, teachers should be trained in formative assessments in earlier grades and remedial education so that children do not fall behind in their studies and repeat grades.

7. **The relationship between students' academic performance and socio-economic status is substantial, but its magnitude varies by subjects.**

The socio-economic status of a student's family has varying effects on their achievement. Many students have performed better in Nepali language, with satisfactory performance in mathematics and science, despite their low socio-economic status. This situation was reversed in English language. This depicts that the socio-economic background of the students does not entirely decide their academic performance.

Recommendation: Although the socio-economic status of students has varying effects on their achievement, it is not only the major deciding factor. Students can excel and achieve better if they focus more on their studies and practise hard, despite the minimum resources available to them. Regardless of the different levels of socio-economic status among students, if schools provide, for example, sufficient learning materials and library facilities, manage student clubs and offer study programmes, students can perform well irrespective of their socio-economic status.

8. **The achievement of Janajati and Dalit children is lower than that of other ethnicities.**

   Ethnicity has influenced the achievement of students in Nepali, English and Science. Generally, Brahman/Chhetri scored higher than Janajati and Dalit students. Students from Brahman/Chhetri communities are, on average, high achievers, whereas students from Dalit communities are achieving lower. There is significant difference in the achievement of Hill Brahman and Madhesi Dalit in scale score of 26 in science, 26 in English and 30 in Nepali.

   Recommendation: The achievement scores of students from Janajati and Dalit communities are below the national average compared to students from Brahman and Chhetris communities. The differences may have been caused by the medium of instruction, language background, content of the curriculum, teachers and cultural background. To reduce these gaps, an inclusive curriculum, remedial teaching, the incorporation of local ideologies into the curriculum, inclusiveness in the teaching profession and a change of learning culture in Dalit – more importantly, Madhesi Dalit – students need to be seriously considered.

9. **Teacher regularity and the availability of study resources have positive correlations with learning achievement.**

   Teachers who were dedicating all their time in the classroom were successful in improving students' achievement. Meanwhile, students of teachers who would come late to class and leave early, or did not come to class at all, had negative performances. Similarly, the availability of study resources such as textbooks, question banks, guides, reference materials and other support has a positive influence on learning achievement.

   Recommendation: School administration should maintain a strict code of conduct for teachers to be at school regularly, and it should be made one of the criteria for their performance evaluation. Regular teachers should be rewarded with incentives. Similarly, government or non-government agencies supporting students through scholarships or any other incentives should consider the availability of basic study resources to the students. Parents should also consider making these essential resources available to meet the primary needs of their children.

10. **The NASA result show declining pattern of achievement consistently.**

    One-third of students in mathematics and science, and nearly half of the students in English, scored below the national average. The consistently weak performance of students in NASA 2012, 2015, 2018, 2019 and 2020 indicates a low return on the investment in education made by the government. The recurring trend underlines the need for ensuring sufficient government intervention to enhance quality education.

Recommendation: The time has already come to carry out a diagnostic study to identify the challenges in the educational system, with a focus on the teaching–learning process. The critical factors that hinder achievement and quality education should be investigated, and immediate steps have to be undertaken to recover the educational loss. Pedagogical intervention in the delivery system deserves investigation, and the adoption by teachers of activity-based, learner-centred, research-based learning approaches – including developing problem-solving, critical-thinking and other 21st-century skills – which are closely monitored and evaluated has now become a necessity. The involvement of parents and community members in holding schools accountable for their students' low achievement should be ensured.

11. **There is a positive correlation between students' academic performance and the use of their leisure time in school.**

    Students who were engaged in classwork or homework during their leisure time achieved higher scores than those who spent their time in playing games or going home. There was a significant difference in achievement, with scale score ranging 32 in mathematics, 42 in science and 54 in Nepali.

    Recommendation: School administration should maintain gap periods appropriately at school level. Meanwhile the identification of weak students and provision of remedial classes for them during leisure time remain milestones for the recovery of learning lost due to various factors. School administration should manage students' leisure time, engaging them either in classwork or doing homework under the guidance of teachers. When organising teacher training or other works, the head teacher needs to pay due consideration to not creating gaps in teaching hours. The relevant educational authority can also manage learning environments and remedial classes with budgets to recover lost learning.

12. **Access to social media also has a positive effect on student achievement.**

    The NASA 2020 study showed the importance of access to social media. The achievement of students who had access to social media achieved higher scores than those without it. There is a significant difference between achievement of students who had access to devices such as mobile phones, the internet and TV and other means of social media, with a difference of 21 scale score in mathematics and 10 scale score in science compared to those who don't.

    Recommendation: Students require access to social media and communication devices to obtain information that may assist their learning and to keep them apprised of information which ultimately enhances their learning. However, these devices need to be for the purpose of facilitating their learning. Unauthorised programs and uncensored programs that are not age-appropriate may hamper them psychologically. Any social media and devices provided to students must be censored and loaded with educational programs that enhance learning opportunities for better achievement.

13. The medium of instruction impacts on achievement in community schools.

The medium of instruction plays an important role in education and pedagogical processes. Basically, institutional schools follow English-medium instruction, whereas community schools are applying Nepali as the medium of instruction. In NASA 2020, the achievement level of community schools applying English-medium instruction scored higher than for students receiving Nepali-medium instruction. There is a significant difference between community schools where English has been applied compared to those applying Nepali medium of instruction. For instance, the difference in the achievement of students receiving English-medium and Nepali-medium instruction ranges 17 scale score in science.

Recommendation: The medium of instruction is a fundamental process for communication and comprehension of the content, as well as pedagogical processes in schools. The language that students feel easiest with must be used as the medium of instruction. In NASA 2020, the language used and the achievement of students have a direct correlation, but the major factor is the comprehension of content delivered in the classroom pedagogy. This is a grey area for further research, rather than just prescribing a particular medium of instruction, as the world becomes more multilingual and our languages are vast resources of knowledge.

14. There are noticeable gaps in school governance.

Some identified factors of school governance – like teacher irregularity, bullying in school, a lack of improvement in classroom practices, the unavailability of school facilities and conducive learning environments, the untimely availability of textbooks, students' perceptions of school/teachers, the large numbers of students lacking appropriate learning opportunities, a lack of remedial action to improve learning – are some of the factors causing the detrimental situation in the achievement outcome. The difference in achievement between community schools and institutional schools is also evidence of difference in school governance.

Recommendation: These issues are basically the concern of the school governance. The issues need to be solved to improve achievement in the school sector. Monitoring and evaluating school governance, empowering head teachers and implementing reward and punishment systems are necessary steps to improve the situation. Local government should play a crucial role in the improvement of governance in the schools.

Note: Sepecfic finding and suggestions of reading, writing, listening and speaking are mentioned in the section of result of English (See pp. 196-200, 214-217).

Contents

List of Figures

List of Tables

CHAPTER I

AN OVERVIEW OF THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

Introduction

In this report, Chapter 1 presents a general introduction of the National Assessment of Student Achievement (NASA), its historical overview and objectives. Chapter 2 deals with methodological procedures of the study to select the sample, item analysis, sample weight calculation and data analysis. Moreover, to explore contextual variables, tools and technologies used during the overall study, including an explanation of the contextual variables like geography, ethnicity, gender, language and economic status, are also mentioned in this chapter. Chapter 3 provides the basic results by contextual variables, Chapter 4 provides the basic descriptive statistics by using Classical Test Theory (CTT), and the last chapter, Chapter 5, presents a summary of the findings, conclusions and recommendations.

The national assessment of grade 8 looked at the achievement of students in mathematics, science, Nepali and English and was conducted by the Education Review Office (ERO) in 2020. The report of the assessment is based on the curriculum-based standardised test. Results have been presented comparatively in all the sub-chapters, focusing on province-wise results as explicit strata and other variable-specific results as implicit strata, for example results by type of school, gender, ethnicity and language in a disaggregated form.

The assessment was conducted in 75 sample districts (out of 77; two districts Manang and Mustang were not included as the number of students was lower there), 1,800 schools and 43,497 students. The main aim of NASA is to provide valid and reliable information on the learning achievement of students at grade 8, with policy feedback to the Ministry of Education, Science and Technology (MOEST). Specifically, NASA provides feedback to the teachers, schools, curriculum developers, policy makers and programme executing agencies for the necessary reformation. A repeated cycle of NASA provides evidence-based information on the trend of student learning and other contextual variables that provide pathways for the review and design of policy and programme.

More specifically, the assessment answers questions like: How well are the students learning ? Is there any evidence of particular strengths and weaknesses in students' learning ? Do certain sub-groups of students perform poorly? What factors are associated with student achievement ? Does the achievement of students change over time? (Greaney & Kellaghan, 2007). This report has emphasised related issues and problems with some recommendations to the policy makers and other stakeholders.

National Assessment of Student Achievement

Globally, it has been well accepted that the means of measuring the quality of education is students' achievement (TIMSS & PIRLS, 2008). The national assessment provides both quantitative and descriptive forms of information on student achievement, which is considered as an output of the teaching–learning process and its quality (World Bank, 1996). National assessment thus provides basic information for policy makers, politicians and the wider educational community (ERO, 2013). Students' assessment provides data to inform policy makers about the key aspects of the education system (Greaney & Kellaghan, 2008b; ERO, 2013). Studies showed that the achievement of students in a curriculum area can be aggregated to provide an estimate of the achievement level in the education system as a whole at a particular age or grade level but not the individual level (Greaney & Kellaghan, 2008b; NASA, 2013). NASA is also a standard means of determining the achievement of the curriculum and finding gaps between the written curriculum and the taught curriculum. Therefore, it is useful for making policy decisions, especially when decisions are to be made in relation to the optimum utilisation of resources (EDSC, 2008). The NASA report provides evidence for policy makers on the availability of textbooks, class size and the number of years of teacher training. Therefore, it is globally accepted as a systematic and regular measure of learning achievement in a country, which is designed to assist policy making (EDSC, 2008; ERO, 2013; ERO, 2019).

Evolution of NASA in Nepal

Assessment practice is found to have started from the last years of the 1980s in Nepal. However, the Ministry of Education formally started the National Assessment in 1995 and continued it up to 2010 on a small scale. The large-scale NASA has been administered under the Ministry of Education since 2011. Four NASA cycles were completed during the School Sector Reform Plan (SSRP) and three, including NASA 2020, have been completed during the School Sector Development Plan (SSDP). In both plans, NASA is considered as a tool to measure quality of education, for holding educational institutions accountable for achieving educational goals.

NASA studies are conducted for the purpose of both reflection and prediction. The reflective purpose is concerned mainly with building a database to analyse both the strengths and the weaknesses of educational policies and practices that affect students' learning achievement (ERO, 2018, 2019).

The assessments completed so far and the forthcoming assessments as per the designed NASA roadmap are presented in Table 1.

Table 1 Roadmap of NASA cycles and progress

SSRP				SSDP						
2011	2012	2013	2015	2017	2018	2019	2020	2020	2021	2022
Grade 8	Grades 3 and 5	Grade 8	Grades 3 and 5	Grade 8	Grades 3 and 5	Grade 10	Grade 8	-	-	Grade 5 …
✓	✓	✓	✓	✓	✓	✓	✓	Not assessed due to COVID-19 impact		In progress

A complete NASA cycle covers a period of three years. In the first year, tasks related to test items (item development, pre-test and analysis) are completed. In the second year, the final NASA assessment is administered. Finally, in the third year, activities like data analysis, report writing, result dissemination of the assessment and policy feedbacks are performed.

The ERO follows globally accepted practices for conducting national assessments. Although the context of each country is different, there are some common practices to national assessments in most countries (ERO, 2019). Building on the comprehensive review of national assessments from various countries, ERO has adopted the following procedures:

- The ERO, within the MOEST system, is solely responsible for conducting the national assessment.
- The ERO has developed, and revised as necessary, policies and frameworks for assessment in consultation with (and with the participation of) key stakeholders such as subject experts, teachers and policy makers.
- The MOEST determines and approves the grade level and determines the area (e.g. literacy or numeracy, science or mathematics or English) to be assessed.
- The implementing agency (ERO in Nepal) defines and describes the areas of achievement testing in terms of both content and cognitive skills and develops test items along with supporting questionnaires and manuals for test administration.

ERO:

- Pilots the test items to develop best assessment items for final assessment. In the piloting process the support of external experts and reviews their validity, appropriateness and sensitivity in terms of gender, ethnicity and culture.
- Ensures that the assessment instruments are reliable and valid through pilot item analysis.
- Selects the sample schools, arranges for printing the test papers and other relevant materials, and communicates with the schools and teachers for test administration.

- Orients the test administrators (focal persons, test administrators, supervisors) administers the test and survey questionnaires in the selected schools.
- Checks the answer sheets and scrutinizes them after training the evaluators.
- Collects test scores and other necessary information, cleans the data as per the requirement and analyses them.
- Prepares draft report/s which is/are reviewed by relevant subject committees and external experts.
- Prepares and disseminates final report/s through various means such as publication and the mass media.
- Finally, the MOEST, implementing agency and relevant stakeholders study the report/s of national assessment and identify major areas for policy reforms (ERO, 2017, 2018).

NASA cycle

ERO has adopted the following cycle to conduct the national assessment of grade 8 students in mathematics, Nepali, science and English.

Figure 1 NASA process cycle

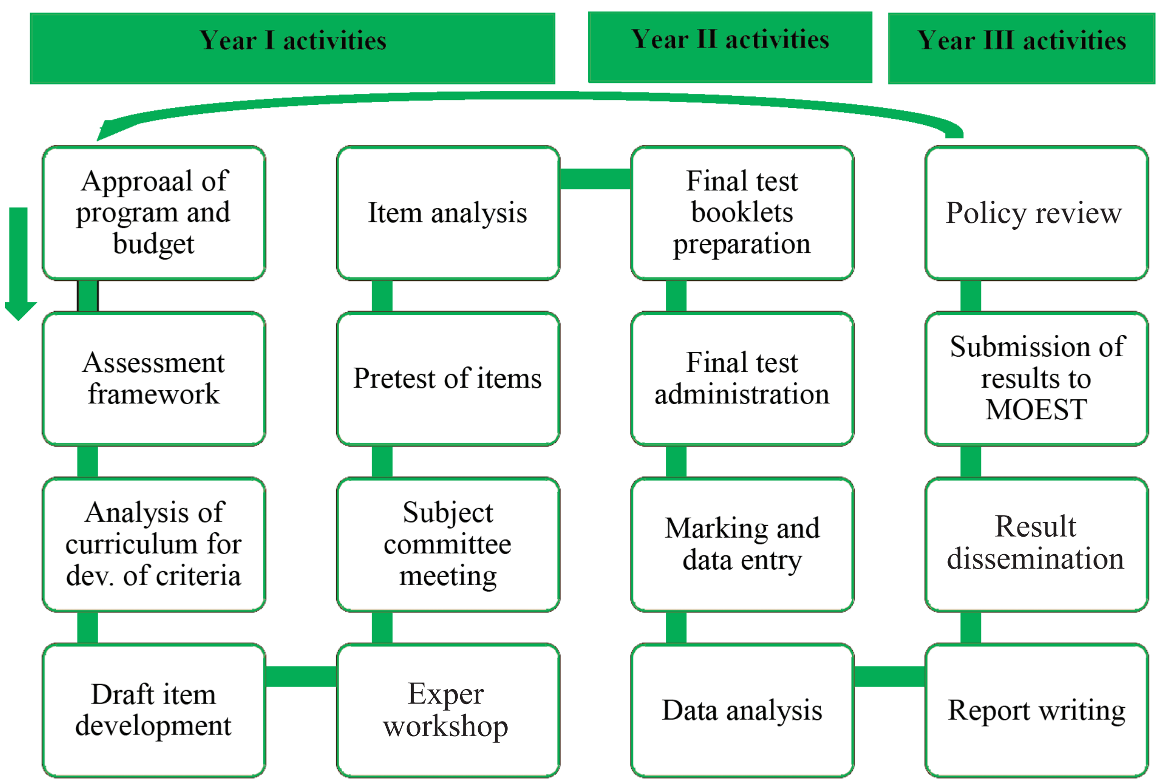


Figure 1 presents the major steps taken in planning, designing, administering and reporting the assessment. The NASA process cycle begins with approval of the required budget and programme and proceeds through a series of assessment procedures: development of the assessment framework, criteria and standards; development of test items and questionnaires; piloting, item analysis and selection of the items; designing the test booklets; administering the test; scoring and preparing data; data cleaning, calibrating items and equating the tests; analysing and setting proficiency levels; and reporting and disseminating the results.

Objectives of NASA 2020

The purpose of this assessment is to provide feedback to the Ministry of Education, Science and Technology to improve the quality of school education. This assessment does not report individual students' performance, nor does it compare the proficiencies of each individual student and school. Rather, it provides national and provincial-level results, as well as the differences in the achievement scores in relation to various influencing factors such as socio-economic status, home language and identity with geographical region, student attitudes towards subjects in schools, bullying and so on. More specifically, NASA 2020 has the following objectives:

a) To identify the current level of grade 8 students' achievement in mathematics, science, Nepali and English
b) To identify variations in student achievement by aspects such as gender, province, type of school, ethnicity, home language and socio-economic status
c) To explore factors that influence student achievement
d) To identify trends in student learning and produce baseline data for future comparisons
e) To strengthen the capacity of the education system in conducting national assessment
f) To provide the Ministry of Education, Science and Technology with recommendations for policy making to improve quality and equity, particularly in school education.

Distinct features of NASA 2020

The ERO has used Item Response Theory (IRT) to assess the latent ability of students, using various contextual variables to explain those latent traits. This assessment has used advanced procedures to bring rigour to the data analysis by generalising the results to national and provincial levels through seven explicit strata and various other implicit strata. Use of the Replicate Model for estimating the population parameters and weighted likelihood estimation (WLE) for analysis of individual student levels and reporting are examples of its advancement. Furthermore, the advancement of procedures has also been noticed in sampling methods. A probability proportional to size (PPS) sampling procedure has been used in selecting the schools as the primary sampling unit (PSU), the school clusters. Reporting of student achievement at province level and national level is done in a transformed scale with mean 500 and standard deviation 50 by using the formula:

Average scale score = 500 + plausible value × 50

or *Average scale score* = 500 + logit × 50

The distinct features of this report are:

1. learning-level descriptors prepared through a rigorous analysis
2. gaps in learning between the written curriculum and the taught curriculum in the form of achieved curriculum are presented by using the defining proficiency level (DPL) method.
3. to increase the strength of the result, the sample size answering an item has been doubled compared to previous years by combining two subject test papers to be given to a student. To accommodate this change, there are fewer items in a subject's test item set, but the number of test booklets is increased.

CHAPTER 2

Methodology

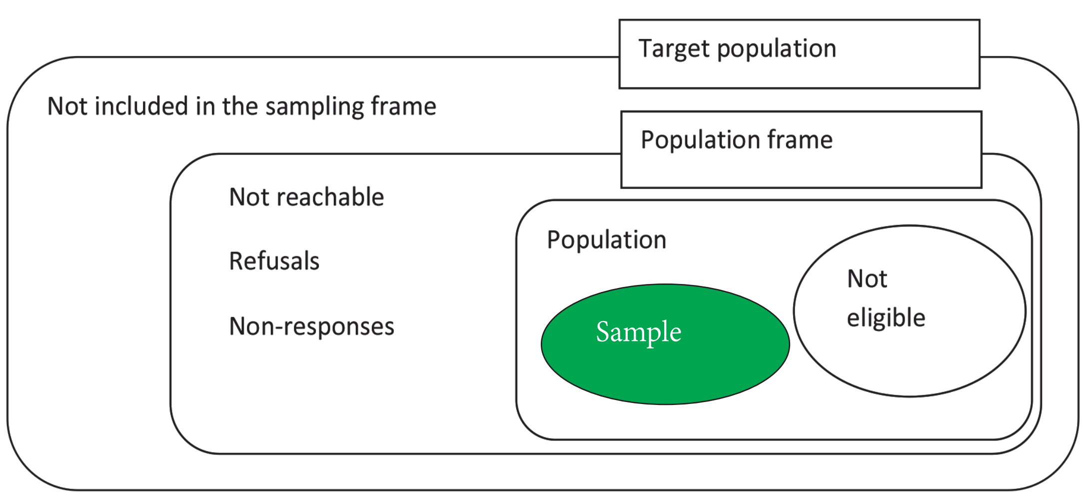
This chapter presents the process adopted for sampling, assessment framework, tools development, setting contextual variables and determining the reliability and validity of the tools. It also presents the statistical tools and techniques used in the data analysis of NASA 2020. Moreover, various formulas, symbols and techniques used in the data analysis and reporting are described in detail in this chapter.

Sampling

Target sampling frame

Sampling is a process of selecting a set of data from the population by using a defined procedure (Rahi, 2017). In this assessment, the multistage sampling process was adopted. In the first step, a list of all 12,861 schools to be included in the assessment, with their unique ID (school EMIS code) provided by the Department of Education (DOE; now Centre for Education and Human Resource Development – CEHRD) was listed. This list was the target population for developing the sampling frame. In addition to the name, location (provincial, district, geography and municipality) and ID (code) of each school, public and private categories, the total number of students, with gender categories, in each school was taken as the sampling frame. This data is available from the Education Management Information System (EMIS) of CEHRD, and is collected through the national census of schools every year. The target sampling frame for this assessment was thus prepared on the basis of the school data of 2020, with 512,865 students as the target population.

Figure 2 Conceptual diagram of population for sampling frame



Population

The population of the study is the schools running classes up to Grade 8 in the academic year 2020 (2076BS). However, some of the schools did not report the number of students (zero students) and such schools were excluded from the population frame. After the exclusion of non-student schools, from the 12,860 schools and 512,865 students from those schools who reported at least one student, schools with fewer than ten students were also deemed ineligible and excluded. The population of this assessment then reached 12,043 schools with student numbers above or equal to ten students. From those schools, the student population was estimated to be 462,668 at maximum. Sample cluster schools were selected from those schools by using the probability proportional to size (PPS) sampling method. Thus, the population for this assessment covered all students enrolled at grade 8, taken randomly from primary sampling units (PSUs). Exclusion of schools was defined by the following criteria:

- schools having fewer than ten students
- students who did not respond to the test items (during data cleaning)
- very remote schools or schools which were unreachable at the time of assessment – these were replaced during the assessment from the list of replacement schools from the sample frame
- schools which do not have students in grade 8.

Table 2 displays the province-wise population of grade 8 students.

Table 2 Province-wise population of grade 8 students

Province	Population of students
Province 1	85607
Madhes	80366
Bagmati	105400
Gandaki	43438
Lumbini	94595
Karnali	41604
Sudurpaschim	61855
Total	**512865**

Sample size

The educational survey research studies suggest that the sampling precision requirements should be satisfied by a simple random sample (SRS) of 384 students for the main criterion variable (Cohen et al. 2007). This size of simple random sample yields a 95% confidence interval for the student-level estimate, with a 3% confidence interval for margin of error (Krejcie & Morgan, 1970). However, a perfect random sampling is not an easy task in such a large-scale national assessment. The sampling design includes a combination of different sampling techniques at different stages, including stratification, clustering and random selection of students. For this, the design effect due to the multistage sampling has to be calculated and adjusted while selecting the sample size.

In this assessment, actual sample size was calculated using multistage sampling methods. Intra-class correlation was taken from the recently administered survey of grade 3 (for reference, grade 3 data was used for calculating r). Taking intra-class correlation (r or ICC) = 0.5, and school cluster size (C) equal to 25, the design effect (*Deff*) was calculated by using the formula:

$$Deff = 1 + (C - 1) \times r$$

where: *Deff* = design effect

C = the size of the cluster (number of students within the school who will be assessed in a subject)

r = intra-class correlation (ICC)

To calculate the clustered sample size (CSS), the following formula was used:

$$CSS = ESS \times Deff$$

Where ESS is the effective sample size.

- Hence, the outputs of sampling are intra-cluster correlation (ICC) = 0.23, *Deff* = 6.52, ESS = 384, CSS = 2,503, non-response of students is assumed at 4%, by the rate of 25 students per PSU, total cluster per province becomes 108.66 per subject (e.g. *mathematics* + *science* group or *English* + *Nepali* group), by adjusting school non-participation by 4%, school participation is 96%, hence a cluster of province for two subjects becomes 217 (rounded). However, when there are fewer students in any province, the number of schools or students decreases, and vice versa. Now, 7*108.66 = 763 schools stand for a subject. For sufficient sampling and better precision, 900 schools per subject were sampled. However, there are four subjects, and sampling individually for each subject, there would have to be 3,600 schools. Therefore, to keep the number of schools within 1,800, two subjects were combined in a test paper. Hence, two test papers (combined *science* + *mathematics* and *Nepali* + *English*) were administered in 1,800 schools from the whole population. The sample is therefore sufficient to generalise the results over the population. Based on the number of schools and students, more schools were selected so that a sufficient number of schools were selected randomly to make the sample more powerful. Additionally, to make the sample more valuable, an additional 6% of students were sampled so that any kind of risk of non-participation or non-response can be recovered.

Selection of a valid sample of 502,634 students from the cluster sample of 512,865 meant exclusion of 10,231, which was 2% of the total population. Thus, the exclusion of the school cluster in the sample frame was 2% in total. However, after assessment, the final population was received as 462,668 based on the weight calculated because some students did not participate or left the answer sheet blank.

Stratification of the country

Table 3 The final valid list of schools and students

S.N.	Province	Sum of G8_total (N>9)	% of total	Of 1,800 schools	Final sample schools
1	Province 1	83563	17	299.2503492	266
2	Madhes	79495	16	284.6822937	280
3	Bagmati	102045	20	365.4368785	305
4	Gandaki	41681	8	149.2652706	217
5	Lumbini	93475	19	334.7465551	295
6	Karnali	41282	8	147.8363979	217
7	Sudurpaschim	61093	12	218.7822551	220
	Grand total	**502634**			**1800**

The slight variation in the number of additional schools was decided based on the availability of schools by geographic locations.

Random number for the selection of first school

To start sampling, random numbers were generated province-wise because the sampling strategy was PPS.

Table 4 Random number generated for selection of first school in each of the seven provinces

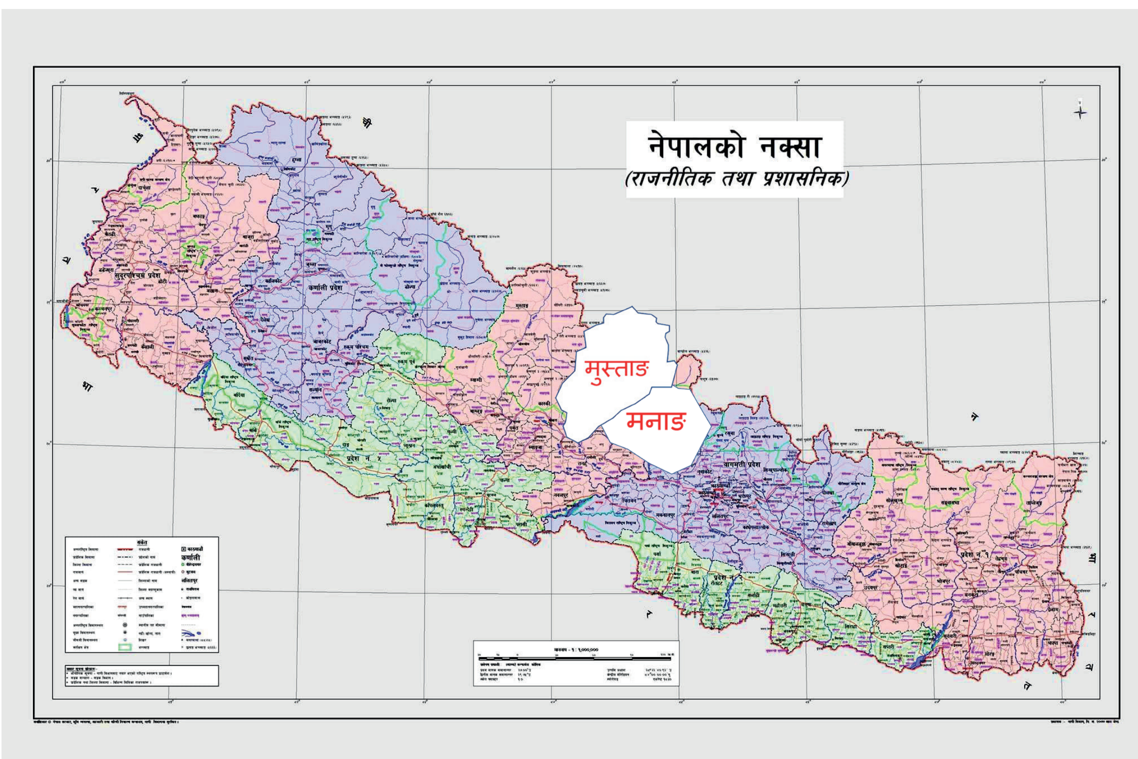
Provinces	Interval	Random number	Start at (N)
Province 1	314.15	0.3889	122.17
Madhes	283.91	0.6868	194.99
Bagmati	334.57	0.6218	208.05
Gandaki	192.08	0.8721	167.52
Lumbini	316.86	0.2368	75.021
Karnali	190.24	0.4145	78.847
Sudurpaschim	277.7	0.0801	22.242

In Table 4, the interval is the quotient of population of any stratum to the total number of students required to be sampled from the strata. 'Start' represents the random starting point for students from the cumulative number of students list, sorted from highest number of students to the lowest number of students of the schools within the province.

Sample design and stratification

The sample design for the NASA 2020 grade 8 assessment was multistage sampling by the selection of schools from each explicit stratum (province). In Nepal, seven provinces are politically divided entities of the country, and they govern educational administration within their region autonomously. Sufficient samples taken from the provinces will ensure generalisability of the results. Districts were selected randomly from each geographical location to incorporate Mountain, Hill and Terai areas. The PSU schools (clusters) were selected within the district using a PPS method. The 75 selected districts from all seven provinces are presented in Figure 3.

Figure 3 Sample shown on map of Nepal (Manang and Mustang are excluded)



Selection of schools and students

From the population, a total of 512,865 students was estimated to be taken as the sample. However, the number of students in the EMIS database did not match with the real test administration. Thus, the number of participating students was less than the estimated sample. Viewing the different sizes of schools, the maximum student sample size was fixed at 25 per school, known as measure of size (MOS).

Where a sample school had more than 25 students, the students were selected using a random sampling method; otherwise, all students were taken as the sample with a defined number of students. More specifically, the number of students sampled from each of the selected schools was decided in two different ways: (i) if the number of students was less than or equal to the expected sample size (MOS), all students were sampled; (ii) If the number of students was greater than the expected size, the required number of students were selected randomly. The probability of a particular student being selected was always the same.

Process of school selection

Table 5 provides an example of how the above information was implemented while selecting the school clusters.

Table 5 Part of the sampling frame as an example

District	Local Level	Sch_code	School Name	Sch_type	G8_Sch_size	Cumu-lative	Selection	Sample
Morang (13)	Biratnagar Mahanagarpalika	50140027	Pokhariya Ma. Vi.	1	268	268	122	1
Morang (13)	Sundarharaicha Nagarpalika	50270013	Sukuna Ma. Vi.	1	263	531	436	1
Morang (13)	Biratnagar Mahanagarpalika	50140059	Satya Narayan Ma. Vi.	1	262	793	750	1
Morang (13)	Biratnagar Mahanagarpalika	50140060	Adarsha Ma. Vi.	1	240	1033		
Morang (13)	Uralabari Nagarpalika	50640012	Radhika Ma. Vi.	1	239	1272	1064	1
Jhapa (04)	Birtamod Nagarpalika	40010018	Mahendra Ratna Ma. Vi.	1	219	1491	1378	1
Morang (13)	Sundarharaicha Nagarpalika	50580009	Saji Lal Ma. Vi.	1	206	1697	1692	1
Jhapa (04)	Gauriganj Gaunpalika	40190003	Gauri Ganj Ma. Vi.	1	198	1895		
Jhapa (04)	Mechinagar Nagarpalika	40370022	Adarsha Ma. Vi.	1	180	2075	2006	1
Udayapur (14)	Triyuga Nagarpalika	140430031	Triyuga Ma. Vi.	1	179	2254		

District	Local Level	Sch_ code	School Name	Sch_ type	G8_Sch _size	Cumu-lative	Selection	Sample
Jhapa (04)	Birtamod Nagarpalika	40010017	Debi Ma. Vi.	1	176	2430	2320	1
Sunsari (06)	Ithari Upamahanagarpalika	60270009	Janata Ma. Vi.	1	174	2604		
Jhapa (04)	Mechinagar Nagarpalika	40370021	Dhulabari Ma. Vi.i	1	173	2777	2634	1
Jhapa (04)	Arjundhara Nagarpalika	40290011	Janata Ma. Vi.	1	170	2947		
Taplejung (01)	Phungling Nagarpalika	10320004	Bhanu Jana Ma. Vi.	1	165	3112	2948	1
Morang (13)	Rangeli Nagarpalika	50500009	Public Ma. Vi.	1	164	3276	3262	1
Jhapa (04)	Bhadrapur Nagarpalika	40100013	Birendra Ma. Vi.	1	156	3432		
Sunsari (06)	Harinagara Gaunpalika	60240004	Harinagara Ma. Vi. Harinagra	1	155	3587	3576	1
Sunsari (06)	Barah Nagarpalika	60360008	Mahendra Madhyamic Bidhyalaya	1	153	3740		
Sunsari (06)	Barah Nagarpalika	60110010	Jagan Nath Dedraj Janta Madhyamik Vidyalaya	1	150	3890	3890	1
Morang (13)	Kanepokhari Gaunpalika	50090006	Janasewa Secondary School	1	149	4039		
Jhapa (04)	Buddhashanti Gaunpalika	40080013	Buddha Adarsha Ma. Vi.	1	146	4185		
Sunsari (06)	Dharan Upamahanagarpalika	60180008	Sikshya Sadan Ma. Vi.	1	144	4329	4204	1
so on ….	…..	…..	…..	…..	…..	…..	…..	…..

Because of school replacement and student non-response adjustments, a calculation of sample weight by PPS sampling methods was completed. In the raw data, some records were for background information only and some were subjective test item responses only, with unidentified unique ID or school deleted from the database. Therefore, the finalised clean data, after removal of duplicate cases, outliers and invalid entries, was as given in Table 6.

Table 6 Number of participating students from the sample in the four subjects

Province	Number of participating students in each subject			
	Mathematics	**Nepali**	**Science**	**English**
Province 1	3114	3089	2954	3112
Madhesh	3607	3590	3567	3590
Bagmati	3632	3641	3611	3643
Gandaki	2553	2536	2491	2557
Lumbini	3739	3445	3705	3456
Karnali	2570	2619	2436	2641
Sudurpaschim	2599	2658	2382	2684
Total	21814	21578	21146	21683

Table 6 shows a subject wise difference between the estimated population and the actual number of participating students subject-wise. This is because of differences in the number of students present on the day of test administration, the number of students reported in the EMIS database, student non-participation and school replacement.

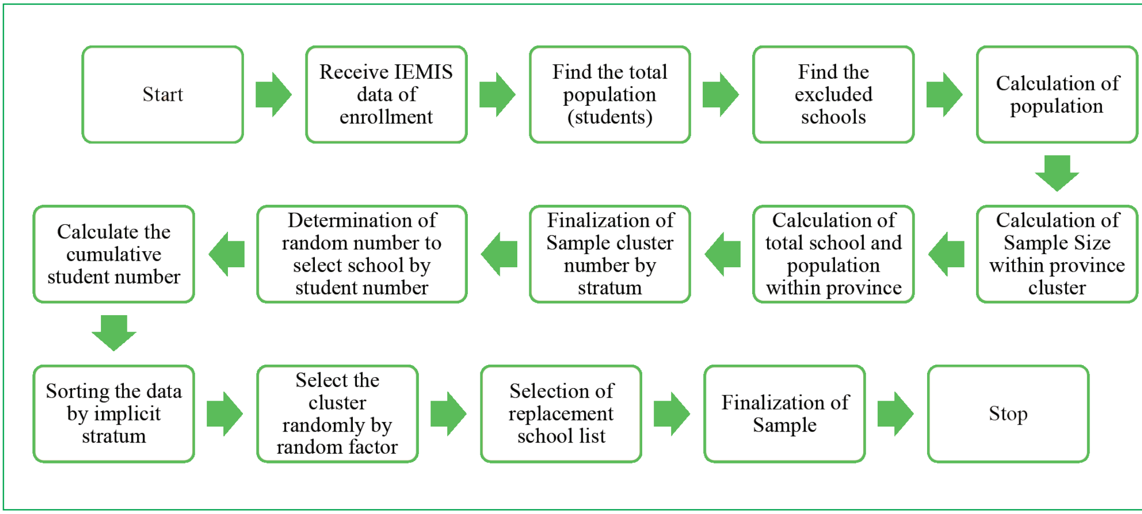
Table 7 Number of participating schools by type in the four subjects

Province	Nepali and English		Mathematics and Science	
	Community	**Institutional**	**Community**	**Institutional**
Province 1	105	26	103	26
Madhesh	128	12	131	9
Bagmati	94	57	96	66
Gandaki	86	25	88	21
Lumbini	121	26	122	26
Karnali	96	4	99	4
Sudurpaschim	99	10	98	11
Total	726	174	737	163
Grand total	900		900	

In the sample, the type of school (community or institutional) was an implicit stratum, whereas province was an explicit stratum.

The sampling procedure adopted in the study is presented in Figure 4.

Figure 4 Sampling procedure



School replacement

In normal conditions, replacing schools was unnecessary. However, when the school was closed, the school had no students available or the school was unable to conduct testing owing to other difficult circumstances, such schools were replaced from the predetermined list of replacement schools, which are either previous or successive schools. The company responsible for conducting the tests was provided this list, which was decided by both ERO and the company. Thus, only a few schools (less than 1%) were replaced from the list.

Selection of the students

Selection of the students within the cluster was truly random, by using the lottery method. However, when there were 25 students or fewer in a school, all students were included in the sample.

School weighting

School-level base weights were calculated using the formula:

$$BW_{sc}^{i} = \frac{N_{pop}}{n_{sc} \times N_{mos}^{i}}$$

where N_{pop} was the population size (students), n_{sc} was the total number of schools sampled within each explicit stratum, and N^{i}_{mos} was the measure of size assigned to the school

(i) School-level base weights were calculated for all sampled schools that satisfied the condition for eligible students actually participating in the study. For example, in mathematics, altogether 900 schools were sampled, out of which one school did not participate in testing due to some unavoidable circumstances. For this, a school-level non-response adjustment was calculated separately for each explicit stratum, using the formula:

$$Sc_{adj} = \frac{n_{sc}}{n_{psc}}$$

where n_{sc} is the total number of originally sampled schools, and n_{psc} was the number of schools that actually participated.

The final school weight was then calculated with non-participation adjustment to the base school weight. The final school weight was then equal to the product of the school base weight and non-participation adjustment,

$$W_{sc} = BW^{i}_{sc} \times Sc_{adj} .$$

Student weighting

For schools with 25 students of grade 8, base weight was assumed as 1; for schools with more or fewer than 25 students, the base weight was calculated using the formula:

$$BW_{st} = \frac{N_{st}}{n_{st}}$$

where N_{st} was the total number of students at grade 8 in the sampled school, and n_{st} was the number of sample students from the class.

A student non-participation adjustment was calculated for any school that had at least one student who was sampled and was eligible to do the test but did not participate for some reason. This was calculated with the formula:

$$St_{adj} = \frac{n_{st}}{n_{pst}}$$

where n_{st} was the number of sample students, and n_{pst} was the number of students who participated in that particular school.

The final student weight of a particular school (say, i^{th} school) was then equal to the product of the student base weight and non-participation adjustment: $W^{i}st = BW^{i}_{st} \times St_{adj}$.

The final weight was thus the adjustment between the product of the school and student final weights: $W_i = W^{i}_{sc} \times W^{i}_{st.}$.

2.2 Test administration and supervision

Test administrators for NASA 2020 were appointed from resource personnel, school supervisors and head teachers. The appointed test administrators were trained to administer standardised National Assessment as per the NASA test administration guidelines. For support and inspection of the test administration, a teacher from the school who was not teaching the assessed subject in that particular school was also appointed. Two other support staff were also assigned for test administration in a school.

For monitoring and supervision of the NASA test administration, three types of monitors were used. Some civil servants at central-level agencies from the Ministry were appointed by ERO and some by EDCU (Education Development Coordination Unit). A team of supervisors was used for immediate support and monitoring of the process in every sample district. Adoption of the test administration process is summarised as follows:

- One school participated in two subjects.
- Subject teachers were not allowed in the test administration hall; rather, they were assigned to provide responses on the Teacher's Background Information Questionnaire.
- The test administration centre head oriented the students, support staff and invigilator to ensure smooth test administration.
- Clear instructions were given to students to try their hardest, in a low-stake environment.
- After the test, the head teachers also responded to the background information questionnaire provided to them.
- To maintain confidentiality, no one was allowed to copy the test papers, take pictures of the test papers or keep the test papers in the school.
- After the test was over, test booklets were collected at the EDCU by a consulting firm. Each school submitted their monitoring report, test administrator's report and a list of participating and non-participating students.

Analysis methods

The data analysis methodology consists of two parts. The first part is item analysis, and the second part is data analysis and interpretation. In the first part, MS Excel, SPSS 23, R and different packages (TAM, psych, mirt, plyr, miceadds, CTT, janitor, Wright Map) of R (Statistical computation software) were used to code, recode and clean the database. During the data cleaning, duplicate cases, outliers and unidentified cases were cleaned. All the background variables were recorded to make them readable for R software. Also, dummy variables were prepared for conditioning the run in R.

R software was used to analyse the items to generate item parameters for each item of all five sets. Later, a joint file was prepared by combining all three sets of a subject, and this was useful to generate item-level parameters viz. difficulty, discrimination, item fit, distractor analysis, item characteristic curve (ICC) plots and test information function (TIF) plots. From the joint run, item parameters in the form of logits were generated, and those parameters were fixed for case analysis. After the case analysis, a '.wle' file was generated for case estimation, and this was used for conditioning the run.

In all subjects, the following major data analysis process was adopted.

- Data was cleaned and recoded.
- School codes were edited and finalised.

- The final input file for the TAM package for R software was prepared.
- Set-wise analysis of student responses was carried out using TAM.
- A single database of all five sets of data in each subject was prepared as the input file for final IRT analysis.
- Dummy variables were prepared for relevant variables of student background information and province codes.
- With the combined input of dummy background variables and student score, the final IRT analysis was carried out with NASA 2020 data, and the results were reanalysed by fixing the parameters of NASA 2017 items that were used as anchor items for NASA 2020. After fixing the parameters, the final IRT analysis was done by using the TAM package for R.
- After final analysis of R, data was converted to a Statistical Package for Social Science (SPSS) data file, and a sample file with calculated student weighting was merged to prepare the final analysis file with ten plausible values generated by TAM of R. This SPSS file was used to generate the outputs of the data as results.
- Results are now weighted mean that are calculated by using the weight cases method.

Tools development, their reliability and validity

Assessment framework

Curriculum-based test items were developed based on the assessment framework. The assessment framework is a plan of content, item type, content domain and proportion of test items to be included. It is a blueprint of the whole standardised NASA assessment. The assessment framework was developed before designing the test and developing the test items. The assessment framework was developed to:

provide a clear guideline for a sound assessment approach to inform policy makers and the other concerned stakeholders on quality of education. It includes domains to be assessed, the statement of criteria together with standards, specification of items, framework for contextual variables to be considered while conducting an assessment and brief guidelines for assessment design (ERO, 2017).

The assessment framework has identified and described the domains and constructs to be assessed in mathematics, Nepali, English and science. It has also proposed a framework for designing background questionnaires for students, teachers and head teachers. In addition, it has presented brief guidelines on the overall methodological approach to be adopted for the assessment (www.ero.gov.np – Assessment framework of grade 8).

Item selection for mathematics

Table 8 presents the content domain, criteria, weightage percentage, number and types of items, allocation of marks and distribution items in each of the six standards.

Table 8 Specification table for item selection

Content domain	Criteria nos.	Weightage	No. of objective items	Marks	No. of subjective items	Marks	Weightage for items of various standards
Geometry	1–13	40%	16	16	8	24	The weightage of items in each set should be around: Level 1: 10%, Levels 2, 3, 4 and 5: 20% each Level 6: 10%
Arithmetic	14–21	30%	12	12	6	18	
Data and Sets	22, 23	10%	4	4	2	6	
Algebra	24–32	20%	8	8	4	12	
Total		100%	40	40	20	60	

If content areas with a small number of items (weightage) have difficulty in covering the six levels of standards in one set of test booklets, these content areas may be covered by three sets of questions, which are administered at one time.

While developing and selecting items, various levels of cognitive domains should be taken into consideration. Items should be developed according to the six standards defined above; however, we should check and ensure that the various cognitive domains are adequately represented. Therefore, within the six levels of standards, various levels of cognitive domain should be included.

Item selection for Nepali language

Table 9 presents the item selection criteria for Nepali language.

Table 9 Item selection for Nepali language

Content domain	Criteria number	Weightage	Type of items	Marks	Weightage for items of various standards
Reading (vocabulary)		40%	SR and CR	28	Level 1: 10% Level 2, 3, 4 and 5: 20% each Level 6: 10%
Writing (vocabulary, functional grammar and spelling/ *barnabinyas*)		60%	SR and CR	42	
Total		100%		70	

What needs to be considered here is that the weights of different levels given in the table are only preliminary. The actual weight should be calculated by determining the minimum marks of each level using one of the methods related to level-wise minimum score (cut-score) based on the scores achieved by students. The classification and level suggested above helps in selecting questions for each level. (See detail in Framework of grade 8 assessment published in 2017 from www.ero.gov.np.)

Item selection for science

The following specification table presents content domain, criteria, weightage percentage, number and types of items, allocation of marks and distribution items in each of the six standards.

Table 10 Item selection for science

<table>
<tr><th>Content domain</th><th>Criteria nos.</th><th>Weightage</th><th>No. of objective items</th><th>Marks</th><th>No. of subjective items</th><th>Marks</th><th>Weightage of items of various standards</th></tr>
<tr><td>Physics</td><td>1–10</td><td>26%</td><td>12</td><td>12</td><td>5</td><td>15</td><td rowspan="6">The weightage of items in each set should be around: Level 1: 10%, Levels 2, 3, 4 and 5: 20% each
Level 6: 10%</td></tr>
<tr><td>Chemistry</td><td>11–15</td><td>22%</td><td>9</td><td>9</td><td>4</td><td>12</td></tr>
<tr><td>Biology</td><td>16–18</td><td>20%</td><td>8</td><td>8</td><td>4</td><td>12</td></tr>
<tr><td>Geology and astronomy</td><td>19–21</td><td>12%</td><td>3</td><td>3</td><td>3</td><td>9</td></tr>
<tr><td>Environment Education</td><td>22–24</td><td>20%</td><td>8</td><td>8</td><td>4</td><td>12</td></tr>
<tr><td></td><td></td><td></td><td>40</td><td>40</td><td>20</td><td>60</td></tr>
</table>

If content areas with a small number of items (weightage) have difficulty in covering the six levels of standards in one set of test booklets, these content areas may be covered by three sets of questions, which are administered at one time.

While developing and selecting items, various levels of cognitive domains should be taken into consideration. Items should be developed according to the six standards defined above; however, we should check and ensure that the various cognitive domains are adequately represented. Therefore, within the six levels of standards, various levels of cognitive domain should be included.

Item selection for English language

Table 11 presents the content domain, criteria, weightage percentage, number and types of items (selected response –SR, and constructed response – CR), allocation of marks and distribution items in each of the six standards.

Table 11 Table of specification for item selection

Content domain	Weightage (%)	Marks	Weightage for items of various standards
Reading	60%	48	The weightage of items in each set should be around: Level 1: 10% Levels 2, 3, 4 and 5: 20% each Level 6: 10%
Writing	40%	32	
Total	100%	80	

Note:

1. The total number of SR items (multiple-choice questions) should be between 18 and 24. The number of CR items carrying 1 mark each (very short answer questions) should be between 6 and 12, so that the total number of questions carrying 1 mark each will be 28–32. The number of CR items carrying 2, 3 or 4 marks each should be 16–24, depending on how many marks each question carries, ensuring that the marks for the test total 80.
2. When selecting the items for each content domain, it is necessary to select a reasonable ratio of both SR and CR items.

Item development and selection

Item development workshop

The item development process began with a one-day orientation for the trained item writers on test item development, followed by a workshop to write draft items by school and university teachers. After the developed items had been set on computer, a workshop with experts was organised. Experts at the workshop reviewed the items to ensure their alignment with the curriculum framework and also checked their level and appropriateness.

Once the items had been reviewed by the experts, the subject committee meeting finalised the test item booklets. After this, final language editing and layout design took place before printing in a secured press.

Pre-test of test items

To generate parameters for all items, they were targeted for pre-test on 300 students, but there was a small variation in the number of students in pre-test (of not more than ten students). Altogether six versions of the subject-wise booklets were pre-tested in the sample districts and schools. The pre-test was administered as in Table 12.

Table 12 Number of schools and students participating in pre-testing

S. no.	Subject	No. of sets piloted	No. schools piloted	No. of students participating
1	Mathematics	6	80	1800
2	Nepali	6	80	1800
3	Science	6	80	1800
4	English	6	80	1800

After the pre-test, the items were analysed to produce item parameters. Those parameters were

- difficulty level
- item–rest correlation
- internal consistency
- distractor analysis.

The proportion of items in the final booklets for the four subjects was thus maintained.

Item booklet, scoring key and optical mark recognition design

Selected items in each subject were arranged into three booklets, with some linking items between the booklets. Scoring keys for SR items and scoring schemes for CR items were prepared for each booklet. Based on the booklets and scoring schemes, optical mark recognition (OMR) sheets were designed for data generation and entry.

Preparation of the scoring scheme and guidelines

A group of teachers and experts in the respective subjects worked on the compilation, review and finalisation of the scoring schemes for each subject. For multiple choice and other SR items, answer keys were reviewed and reconfirmed. For CR items, the possible answers as well as marks to be provided in each step were reviewed and confirmed. For dichotomous items, conditions for 0 and 1 credit were clearly specified. For CR items with partial credit conditions, each of the credits 0, 1, 2 and so on were clearly stated. Along with scoring schemes for each subject, some guidelines for scoring were also prepared. Rubrics were developed, including score distribution, in various skills of writing and levels of proficiency.

Review of test booklets and scoring schemes

At the final stage of item selection and item booklet preparation, the committee for each subject reviewed the items and item booklets, editing them, confirming the data and formatting the items. The subject committees prepared the final test booklets, which were then set as printing-ready copy (PRC). The subject committees also reviewed the scoring schemes.

While selecting the items and preparing the test booklets for the final test, the following criteria were considered:

- curriculum-based
- coverage of all content areas
- proper representation of various cognitive domains
- assessment of the various proficiency levels
- items having a range of difficulty from p-value 0.15 to 0.90
- proper discrimination power of the items, item rest correlation r>0.02
- comparability with previous NASA and TIMSS.

Preparation of the item register

Working with subject experts, ERO prepared an item register in each subject on an Excel spread sheet. The unique item identification (ID), item descriptor for each item, scoring keys for selected response type of items or (SR) items and various credits, as well as a description of each constructed response items for CR items, were included in the item register. Two examples are given in Tables 13 and 14.

Table 13 Example of an item register for science

SN	Item_ ID	set_1	set_2	set_3	set_4	set_5	Key	Hierarchy	Item_ type	Content_area	Item_description	Learning objective	Old_ Link_2017	Link with TIMSS	Model used for IRT
1	31		1	12			B	Understanding	SR	Physics	To say relationship between atmospheric pressure and altitude				2PL
2	51			1	9		B	Remembering	SR	Physics	Unit of physical quantity				2PL
4	56			6	14		C	Understanding	SR	Chemistry	Example of element				2PL
6	32		2	13			B	Understanding	SR	Chemistry	Utility of chromatography				2PL
7	57			7	15		A	Remembering	SR	Biology	Classify plants into monocot and dicot				2PL
9	33		3	14			D	Applying	SR	Geo and Astro	Members of solar system				2PL
10	98					14	C	Applying	SR	Biology	Importance of plumule and radical				2PL
87	54			4			D	Applying	CR	Chemistry	Position of atoms when car runs over a can			TIMSS	GPCM
88	89				30		2	Applying	CR	Physics	Calculate the volume of a rectangle from given length, width and height				GPCM
89	99					20	2	Applying	CR	Physics	Calculate final velocity when car starts from rest				2PL

90	1	1					A	Remembering	SR	Physics	Concept of acceleration		Old_2017		2PL
91	2	2					C	Remembering	SR	Biology	Sub-divisions of fungi		Old_2017		2PL
92	3	3					B	Understanding	SR	Geo and Astro	Relation of climate and weather		Old_2017		2PL
93	4	4					A	Understanding	SR	Chemistry	Name of substance having pH value 7		Old_2017		2PL
94	5	5					B	Remembering	SR	Physics	What happens to gas molecules by increment of temperature		Old_2017	TIMSS	2PL
95	6	6					D	Remembering	SR	Biology	Energy observed in photosynthesis			TIMSS	2PL
96	7	7					A	Remembering	SR	Environment	Know producers can use sun's energy to make food			TIMSS	2PL
97	8	8					A	Applying	SR	Biology	Know the concept WBC can destroy bacteria			TIMSS	2PL
98	9	9					A	Understanding	SR	Geo and Astro	Know the concept of the origin of living beings in the water first time			TIMSS	2PL
99	10	10					B	Applying	SR	Biology	Respiratory organ of frog on land			TIMSS	2PL
100	16	16					1	Applying	CR	Physics	Concept of flotation			TIMSS	2PL
101	17	17					1	Understanding	CR	Chemistry	Atomic number			TIMSS	2PL
102	18	18					2	Reasoning	CR	Environment	Suggest ways to minimise amount of carbon dioxide in the environment			TIMSS	GPCM

Table 14 Example of an item register for Nepali language

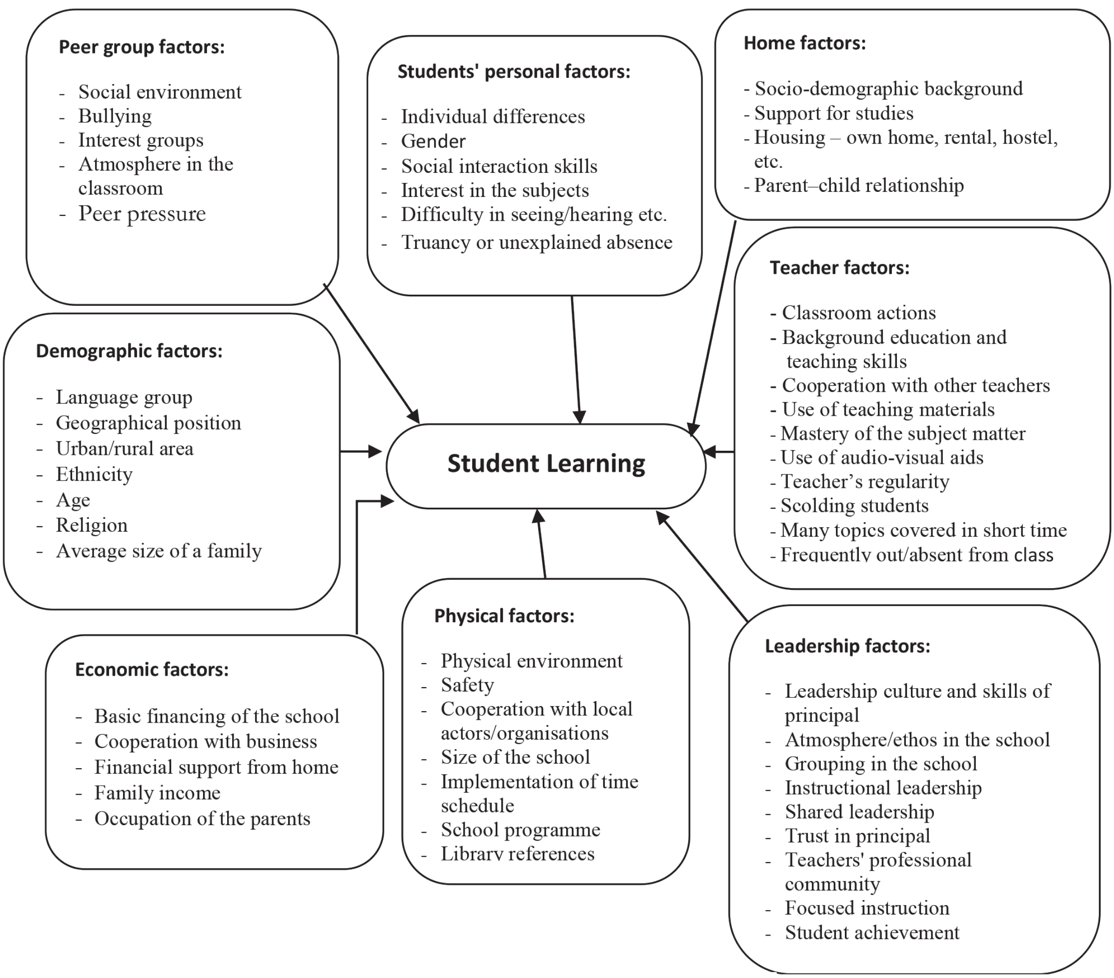
Item_ Serial	Item_ID	Set1	Set2	Set3	Set4	Set5	Content	Item Type	Key	Full_Score	Description	Linked-2017	Model
1	G8N20q1	1				1	Reading	SR	B				2pl
2	G8N20q2	1				1	Reading	SR	A				2pl
3	G8N20q3	1				1	Reading	SR	C				2pl
4	G8N20q4	1				1	Reading	SR	D				2pl
5	G8N20q5	1				1	Reading	ShortAnswer-CR		1			2pl
6	G8N20q6	1				1	Reading	ShortAnswer-CR		2			gpcm
7	G8N20q7	1				1	Reading	ShortAnswer-CR		1			2pl
8	G8N20q8	1				1	Writing	Summary-CR		2			gpcm
9	G8N20q9	1				1	Writing	Argument-CR		3			gpcm
10	G8N20q10	1	1				Reading	SR	B				2pl
11	G8N20q11	1	1				Reading	SR	D				2pl
12	G8N20q12	1	1				Reading	SR	C				2pl
13	G8N20q13	1	1				Reading	SR	A				2pl
14	G8N20q14	1	1				Reading	ShortAnswer-CR		1			2pl
15	G8N20q15	1	1				Reading	ShortAnswer-CR		2			gpcm
16	G8N20q16	1	1				Reading	ShortAnswer-CR		1			2pl
17	G8N20q17	1	1				Writing	Dialogue-CR		3			gpcm
18	G8N20q18		1	1			Reading	ShortAnswer-CR		1		YES	2pl
19	G8N20q19		1	1			Reading	ShortAnswer-CR		1		YES	2pl
20	G8N20q20		1	1			Reading	ShortAnswer-CR		1		YES	2pl
21	G8N20q21		1	1			Reading	ShortAnswer-CR		2		YES	gpcm

Note: Tables 13 and 14 are shown for two subjects as an example of how item registers are organised. Space constraints mean that a lot of information and some columns are omitted.

Background variables

The ERO has developed a framework for collecting background information by means of questionnaires after studying students, teachers and school survey instruments used in various international assessments such as the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) and the Pan-Canadian Assessment Program (PCAP) together with the tools used in previous NASA tests conducted by ERO and discussions with academics, practitioners, parents, teachers and the students. The student attitude scale used in the previous NASA was revised and used. Figure 5 shows the overall framework adopted for the background information questionnaires used in the study.

Figure 5 Concept of contextual variables. (Source: ERO, 2018)



The student background variables shown in Table 15 were included in the assessment.

Table 15 Student background variables/variable blocks

1. School ID 2. Location of school 3. Student's gender 4. Student's age 5. Language spoken at home 6. Caste/ethnicity 7. Identity with geography 8. Time spent beyond school time 9. Support for study at home 10. Availability of textbooks 11. Time to reach school 12. School opening and attendance days in last month 13. Homework and feedback 14. Student's future aim 15. Attitude of student towards subject	16. Student's subject-related activities in classroom 17. Mother's education 18. Mother's occupation 19. Father's education 20. Father's occupation 21. Number of family members 22. Home possession and accessories 23. Activities in leisure time at school 24. Frequency of extra activities at school 25. Frequency of participation in extra activities 26. Attitude towards teacher 27. Attitude towards school 28. Bullying at school 29. Teacher's time on task Other relevant variables from background information questionnaire administered.

Teacher questionnaire

The teacher questionnaire was administered to collect the following information:

- gender, age, first language
- teaching conditions, including class size, access to resources, percentage of students having textbooks, access to substitute teachers in case of absence
- educational experience, teacher qualifications and teaching experience
- teaching–learning practice and conditions at school
- professional engagement with learning, such as access to and interest in professional development, interest in teaching and time spent on preparation for classes
- availability of instructional support, such as classroom visits and feedback by head teacher, school supervisor
- teaching methodology, such as medium of instruction, use of assessment and style of teaching
- satisfaction with working conditions, such as tenure, pay rate and level of supervision
- relationship between the school and community, such as interactions with parents, involvement in school committees

- attitude of cooperation from students.

Head teacher questionnaire

A questionnaire for head teachers was administered to collect the following information:

- gender and age
- educational and management experience and qualifications
- school environment, including the quality of buildings and facilities, as well as availability of resources
- school records, such as fluctuations in student numbers, student and teacher absenteeism
- professional engagement of school leadership, such as access to and interest in professional development and interest in education
- leadership style and use of time
- assessment of teachers' work
- satisfaction with working conditions
- relationship with the community.

Students' attitude survey

In order to find the relationship between the attitude of students towards the subject and their achievement, the attitude survey questionnaire was administered. The questionnaire was adapted from a shortened version of FSMAS – Fennema Sherman Mathematics Attitude Scales – (Fennama & Sherman, 1976). The attitude survey was included in the students' background information questionnaire. The following are the statements used to identify the attitude of students towards the subject:

Self-confidence

1. Studying mathematics makes me feel nervous.
2. I am always under a terrible strain in a mathematics class.
3. I am able to solve mathematical problems without much difficulty.

Value

4. Mathematics is important in everyday life.
5. Mathematics is one of the most important subjects for people to study.
6. High-school mathematics courses will be very helpful to me, no matter what I decide to study.

Enjoyment

7. I usually enjoy studying mathematics in school.
8. Mathematics is dull and boring.
9. I am happier in a mathematics class than in any other classes.

Motivation

10. I prefer not to continue mathematics in college.
11. I am willing to take more than the required classes of mathematics.
12. I plan to take as many courses of mathematics as I can during my education.

Socio-economic status survey

The survey to assess the socio-economic status (SES) of the family was included in the students' background questionnaire. The aggregate of the student's responses to the questions on the following seven factors indicates the SES of the student's family:

- two variables related to parental education, including mother's and father's education
- two variables related to parental occupation, including mother's and father's occupation
- availability of various home accessories,
- availability of possessions at home
- type of school (public or private) attended by student.

Test administration

Preparation for test administration begins with the printing, packing and delivery of the test items and background questionnaires. ERO conducted a one-day orientation on test administration and the test booklet collection process with the head teacher of each sample school in 75 districts. With the help of two teachers, the head teacher of each sample school administered the test. The subject teacher and head teacher of the sample school completed the teacher's and head teacher's questionnaires respectively. Then students' answer sheets together with the teacher's and head teacher's responses were collected at the scoring centre in Kathmandu. The test administration process is described in this section.

The following activities were completed before administering the test to maintain peace and security in a low-stake environment.

- Delivery of head teacher guidelines for test administration. These guidelines gave every step of test administration on what to do and what not to do.
- Delivery of test booklets to the sample schools.

- Orientation for district-level head teachers and monitors.
- Arrangements for test administration by allocating monitoring teams to the centres, scheduling test administration.
- Random selection of participating students, where the number of students was more than the number of students required for the sample class.

Test administration had three parts. The first part was the background questionnaire for students, the head teacher and the corresponding subject teacher. After completion of the students' background information questionnaire, a ten-minute break was scheduled. After the break, the two-hour test was completed.

Mathematics/Science Group: Time allocation = 30 minutes to fill in background information questionnaire + 1 hour mathematics + 1 hour science

Language test group: Time allocation = 30 minutes to fill in background information questionnaire + 1 hour Nepali + 1 hour English

To ensure proper administration of the test, monitoring and sample school visits were conducted by different agencies. The Educational Development Coordination Unit (EDCU – the then District Education Office) not only managed the whole process of test administration but also monitored the administration process at school level. The ERO also sent at least one person to each district to facilitate and monitor the administration of the test. Alongside this, the consulting firm also monitored the test process. After the test, the consulting company collected the booklets and delivered them to the Kathmandu centre. After collecting the answer sheets in the Kathmandu centre, data preparation was completed by adopting the following steps:

- OMR sheet development and printing
- answer sheet coding, marking and scrutiny
- OMR score input and data cleaning
- submission of clean data and marked answer sheets to the ERO.

Item parameter estimation, item review and calibration

Item parameter estimation of each item was carried out and the items were reviewed accordingly. During the analysis, decisions were made on whether or not to use any particular items in the analysis. Classical as well as IRT parameters were estimated to review the items. Item parameters in IRT were used not only for item selection but also to estimate students' latent ability. Based on the item parameters of linking items, three versions of the tests were calibrated, and these three sets were integrated into a single set for analysis. Item parameter estimation, item review and calibration of the test were some of the key processes of IRT analysis from which students' ability was estimated and data was further analysed.

Reliability and validity

The validity of the test items was assured by using the assessment framework. The item-level parameters and set-level reliability of mathematics, Science, English and Nepali are given below.

Reliability

Reliability refers to the consistency of a measure. Reliability is a very important piece of validity evidence (Cohen et al., 2007). It also refers to the quality of items and consistency of the results. Although there are various ways of measuring reliability, internal consistency is considered one of the most used reliability measures. In Table 16, the Expected A Posteriori (EAP) reliability of every five booklets for all four subjects is presented.

Table 16 Reliability of item booklets in mathematics, science, Nepali and English

	Mathematics	Science	Nepali	English
EAP Reliability	0.875	0.806	0.772	0.804

The overall EAP reliability for all subjects was in a good and acceptable range.

Validity

Since all test items were developed and standardised according to the National Curriculum and Assessment Framework, the NASA 2020 test sets were considered to be valid. In the test, although there were more items, a few items were discarded due to misfit of the items in the analysis: a total of 92 items in mathematics, 70 in science, 37 in Nepali and 24 in English were used in the analysis.

Item parameters

In Table 17, IRT parameters are presented.

Table 17 Example of IRT-based item parameters (mathematics – NASA 2020)

SN	Item	Alpha	Beta	tau.Cat1	tau.Cat2	tau.Cat3
1	G8M17A01	1.233	-1.432	NA	NA	NA
2	G8M17A2C2	1.211	-0.83	NA	NA	NA
3	G8M17A3C	1.022	-0.63	NA	NA	NA
4	G8M17A4B3	1.526	-0.265	NA	NA	NA
39	G8M17A39	0.954	1.233	0.741	-0.162	-0.579
40	G8M17A40	1.15	1.043	1.254	-1.254	NA
41	G8M17A41	1.157	1.714	1.002	-1.002	NA
42	G8M17A42	0.937	1.619	1.165	-0.386	-0.779
43	G8M17A43	1.113	2.146	0.404	0.055	-0.46

This table is a part of the item parameter table. Similar parameter outputs are generated before analysing person-ability (see Appendices 1–4) for all subjects.

Item analysis was carried out by using R software that generated various item-level statistics and curves as well. Some examples of the item analysis output are given below.

Figure 6 Sample of item characteristics curve in mathematics

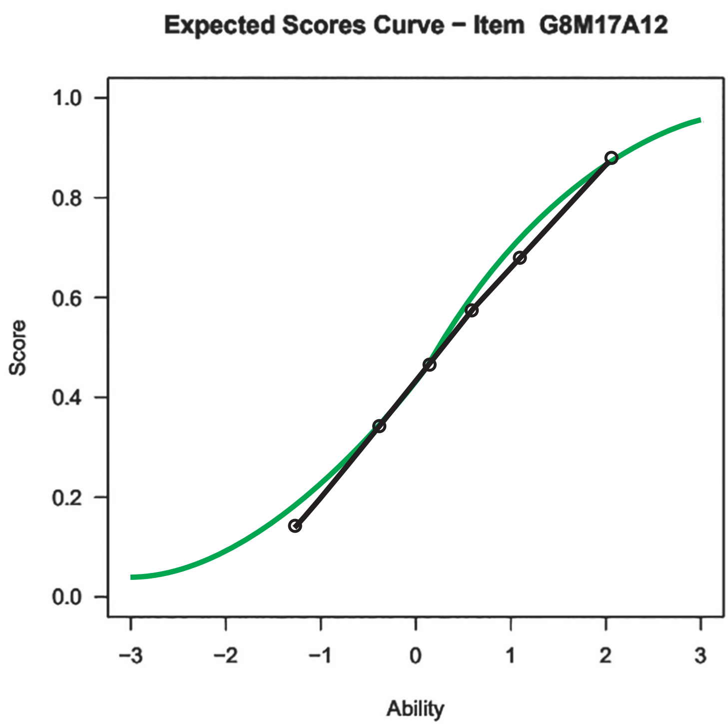


Figure 7 Sample of item characteristics curve in science

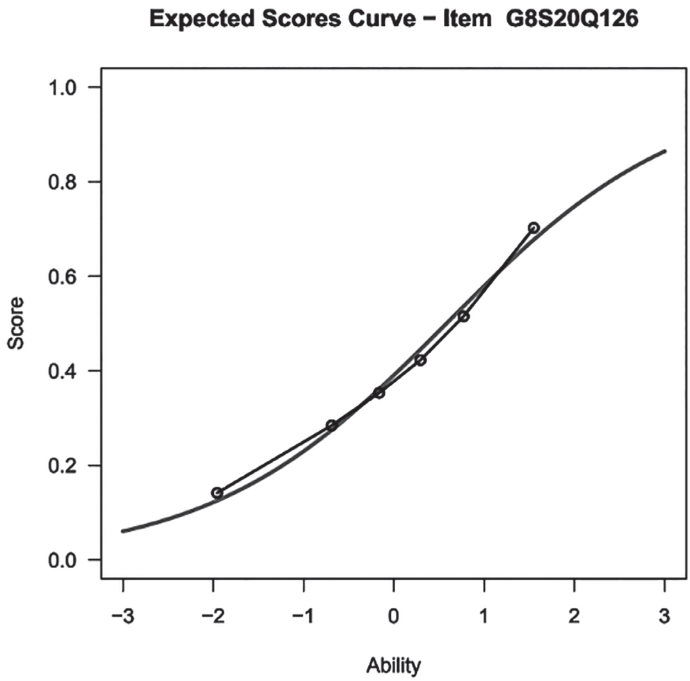


Figure 8 Sample of item characteristics curve in Nepali

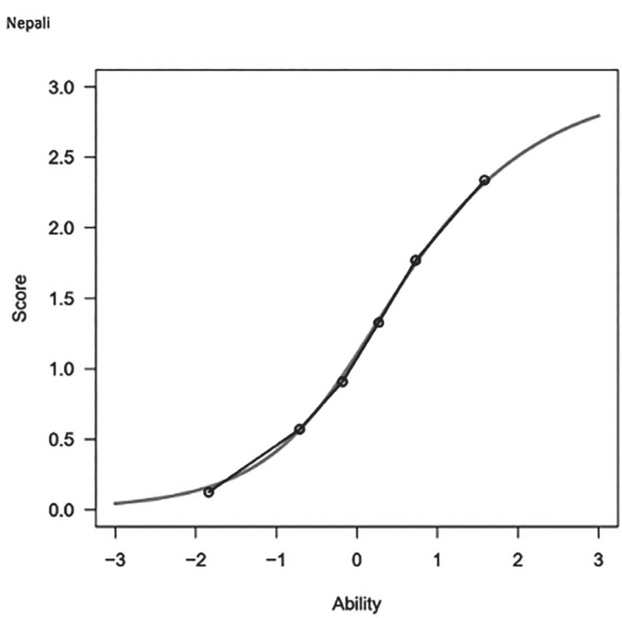
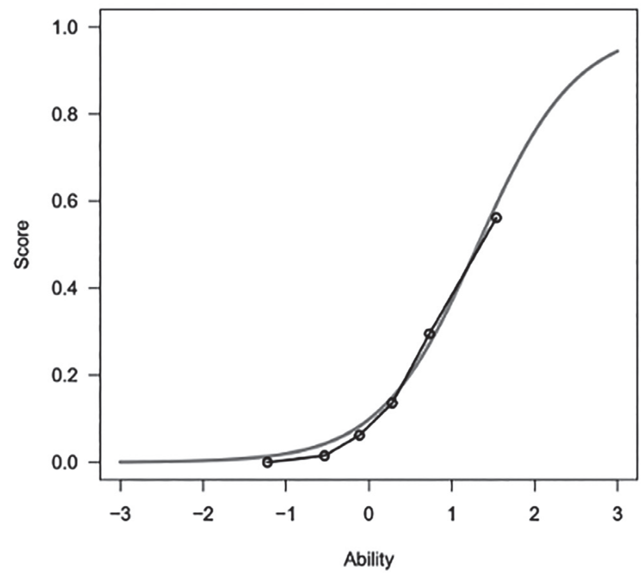


Figure 9 Sample of item characteristics curve in English



Plausible values and results

Plausible values (PVs) improve the precision of prediction ability for the population estimates. The PVs are calculated with conditioning background variables and some school-related indices. Conditioning provides unbiased estimates for modelled parameters. In this assessment, ten plausible values (PV1–PV10) were used to estimate population ability.

In this context, Yamamoto & Kulick (2000) state that the PVs approach 'uses students' responses to the items together with all background data in order to estimate directly the characteristics of student populations and sub-populations' (cited in Laukaitytė, 2016, p. 9). However, PVs are not individual test scores; they are the measures of the performance of the population. It produces an unbiased estimate of population parameters if assumption of scaling is reasonable, but it is not fair to use it for level of student ability.

It is useful to understand how plausible values are generated and how they are presented in the report. Some of the test items used in the 2017 assessment were also used as anchor items in the 2020 assessment. All the 2020 items have been calibrated by fixing the 2017 parameters for those anchor items. The 2020 results obtained in this way are comparable to the 2017 results. Also, the 2020 results will either be greater than, equal to or less than the 2017 results. In 2017, students' average score was 500, but in 2020, it may be different. This has all been analysed using IRT, with 2PL used to analyse items of 1 point, and GPCM for questions with 2 or more points. TAM was used to analyse this.

The latent ability (theta) of each student is obtained from the analysis stated above. Ten plausible values were calculated to find out the ability of the population on the basis of the latent ability, and the dummy variables made on the basis of the variables asked in the stu dents' background questionnaire. After transforming the average of those plausible values to zero (0), the following formula was used to present the results:

Student Achievement Score = 500 + Average PV × 50

Table 18 shows the 14 variable inputs while generating the ten PVs.

Table 18 Variables that were used to prepare dummies

index	Variable	Variable label
1	bq1Palika	Local level
2	bq3sex	Sex of the students
3	bq5LangRecoded	Language spoken at home
4	bq8ethnicgroup	Ethnic group
5	bq12support	Who helps you to read/write at home ?
6	bq13aim	Your future aim
7	bq16mother_education	Mother's education level
8	bq17mother_occupaton	Mother's occupation
9	bq18father_education	Father's education level
10	bq19father_occupation	Father's occupation
11	bq22mobile	Do you have a mobile ?
12	bqN1textbook	Do you have a Nepali textbook?
13	bqN2homework	Does your teacher give homework ?
14	bqN3feedback	Does your teacher give feedback ?

Table 19 is an example of plausible values in Nepali, drawn by conditioning run.

Table 19 Ten plausible values of all individual students calculated (Example from grade 8 mathematics, NASA 2020)

index	UID	theta	PV1	PV2	PV3	PV4	PV5	PV6	PV7	PV8	PV9	PV10	PV1G8M	PV2G8M	PV3G8M	PV4G8M	PV5G8M	PV6G8M	PV7G8M	PV8G8M	PV9G8M	PV10G8M
1	1001	-1.03325	-0.39	-1.09	-0.90	-1.19	-0.60	-1.42	-0.99	-1.20	-1.25	-1.08	464.50	430.08	439.39	425.21	453.98	413.86	435.20	424.44	422.18	430.46
2	1002	1.629559	1.46	1.73	1.20	1.86	1.43	1.40	1.72	1.93	1.11	1.56	555.09	568.47	542.62	574.67	553.57	552.11	567.90	578.37	537.97	559.97
3	1003	-0.34039	-0.43	0.72	-0.21	-0.67	-0.49	-0.70	-0.48	-1.38	-0.13	-0.72	462.28	518.85	473.25	450.49	459.43	449.35	459.77	415.70	477.15	448.04
4	1004	1.971653	1.93	1.85	1.87	1.97	2.04	1.68	1.68	1.74	1.65	1.89	578.16	574.13	575.17	580.06	583.50	566.20	566.08	568.72	564.76	576.32
5	1005	-0.26027	-0.28	-0.53	-0.73	0.71	-0.33	-0.70	-0.11	-0.34	-0.58	-0.63	470.00	457.48	447.86	518.52	467.39	449.14	478.32	467.07	454.97	452.43
6	1006	-0.94169	-0.92	-0.78	-0.82	0.14	-0.95	-0.59	-0.45	-0.76	-0.79	-0.71	438.40	445.28	443.17	490.19	437.15	454.49	461.49	446.16	444.98	448.49
7	1007	-0.31311	-0.47	-0.14	-0.57	-0.35	-0.89	-1.45	-0.59	-0.90	-0.89	-0.18	460.57	476.78	455.60	466.42	439.64	412.44	454.40	439.16	440.09	474.85
8	1008	-1.34683	-1.79	-1.40	-0.73	-1.46	-1.56	-1.76	-1.95	-0.85	-0.92	-1.42	395.52	414.77	447.51	411.88	407.12	397.35	387.78	441.92	438.27	413.66
9	1009	0.018818	-0.02	0.47	0.00	-0.59	0.22	0.21	-0.41	0.12	0.24	-0.36	482.39	506.66	483.62	454.52	494.32	493.95	463.67	489.41	495.31	465.81
10	1010	-0.99557	-0.63	-0.84	-1.85	-0.89	-0.61	-1.48	-0.50	-1.55	-1.51	-1.09	452.57	442.19	392.88	439.76	453.87	410.86	459.25	407.28	409.48	430.10
11	1011	0.484374	0.68	-0.35	0.23	0.45	0.61	0.84	0.42	0.53	0.24	0.13	516.86	466.38	494.87	505.63	513.64	524.94	504.33	509.62	495.35	489.84
12	1012	0.627769	0.65	0.09	0.71	0.63	0.03	0.43	1.06	0.11	-0.01	0.24	515.30	487.88	518.31	514.70	485.13	504.88	535.66	489.17	483.24	495.32
13	1013	-1.40716	-1.61	-2.11	-1.52	-1.24	-1.70	-1.62	-1.63	-0.89	-1.03	-0.39	404.63	379.79	409.05	422.90	399.90	404.19	403.73	440.12	433.22	464.44
14	1014	-1.90985	-0.90	-1.28	-1.18	-1.85	-2.00	-0.75	-1.19	-1.29	-1.92	-1.57	439.58	420.98	425.88	392.69	385.39	446.69	425.21	420.35	389.51	406.37
15	1015	-0.81038	-0.97	-0.33	-1.11	-0.26	-0.85	-1.02	-1.14	-1.44	-0.74	0.08	436.20	467.17	428.91	471.01	441.83	433.35	427.65	412.89	447.03	487.38

Provincial results

Provincial results are presented in the report because Province is the lowest reporting unit of the explicit strata. Provincial results provide the opportunity of comparing the results of major variables. Those results are weighted means over population based on the sample.

CHAPTER 3

Assessment Results In Mathematics Science, Nepali And English

Results of Mathematics

Introduction

In this section, the results of the responses of 21,814 students from 75 districts and 900 schools, who participated in NASA 2020 in mathematics, are analysed. The results are presented in the form of proficiency levels, their description and comparison. Population estimates presented in this section are based on the ten plausible values drawn from WLE. The comparisons are made on the basis of groups formed from background information variables such as students' family background, socio-economic status, ethnicity, gender, home language, school type, home environment, province, etc.

In this assessment, some of the test items used in the 2017 assessment were also used as anchor items. All the 2020 items have been calibrated by fixing the 2017 parameters of those anchor items. Therefore, the 2020 results obtained are comparable to the 2017 results with average achievement score of 500. IRT has been used to analyse this, with 2PL used to analyse items of 1 point and GPCM used for questions with 2 or more points. TAM package in R Statistical Software was used to analyse this.

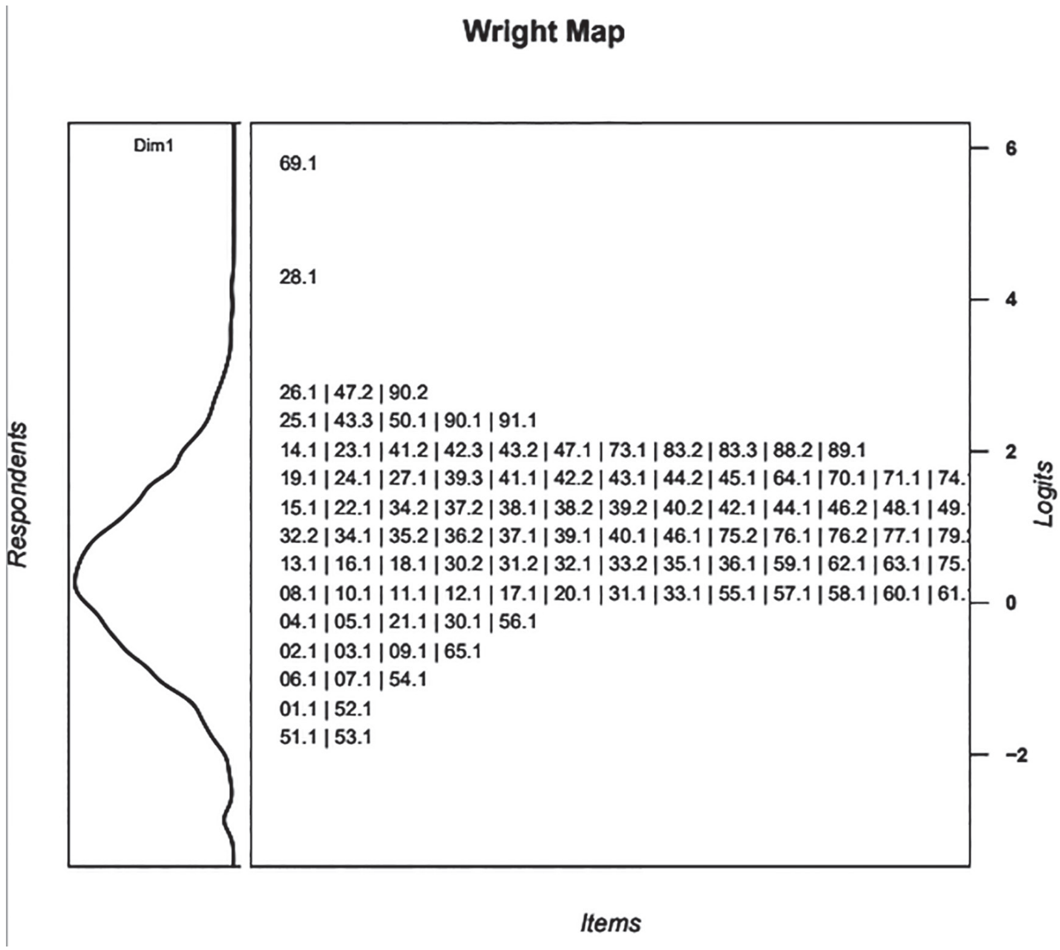
The latent ability (theta) of each student is obtained from the analysis stated above. Ten plausible values were calculated to find out the ability of the population on the basis of the latent ability, and the dummy variables made on the basis of the variables asked in the students' background questionnaire. After transforming the average of those plausible values to zero (0), the following formula was used to present the results:

Student Achievement Score = 500 + PV × 50

The Wright map for mathematics

The Wright map is organised as a frequency curve and vertical histogram. The frequency curve on the left side shows the respondents, and the histogram on the right side shows the test items. The left side of the map shows the distribution of the measured ability of the respondents from most able at the top to least able at the bottom. The items on the right side of the map are distributed from the most difficult at the top to the least difficult at the bottom. In Figure 10, student ability (θ) on the left and NASA 2020 items to the right are plotted on the same scale. When a person and an item lie at the same level, the probability of that response by the particular respondent is 50%. Figure 10 presents the NASA 2020 Wright map for mathematics.

Figure 10 Wright map showing respondent and item on the same scale



The curve on the left side represents the histogram of distribution of students; their latent ability is displayed in the logit scale ranging from -2 to +6 The distribution of students against the items administered (item numbers are shown to the right side) reveals that most of the items were difficult for the students. Although items were pre-tested and based on the grade 8 curriculum, most of the students were below the average latent ability '0'. This indicated that test items were difficult for the participating students. This further depicts that the performance level of grade 8 students in mathematics was not as expected by the curriculum.

National mean in mathematics

This comparison is made based on the linking items administered in both years. While comparing the national mean, NASA 2017 linking item parameters were fixed. Based on those fixed item parameters, NASA 2020 items were calibrated. The national mean score in NASA 2020 is 483, in comparison with 500 in NASA 2017.

Figure 11 Comparison of national means of NASA 2020 and NASA 2017

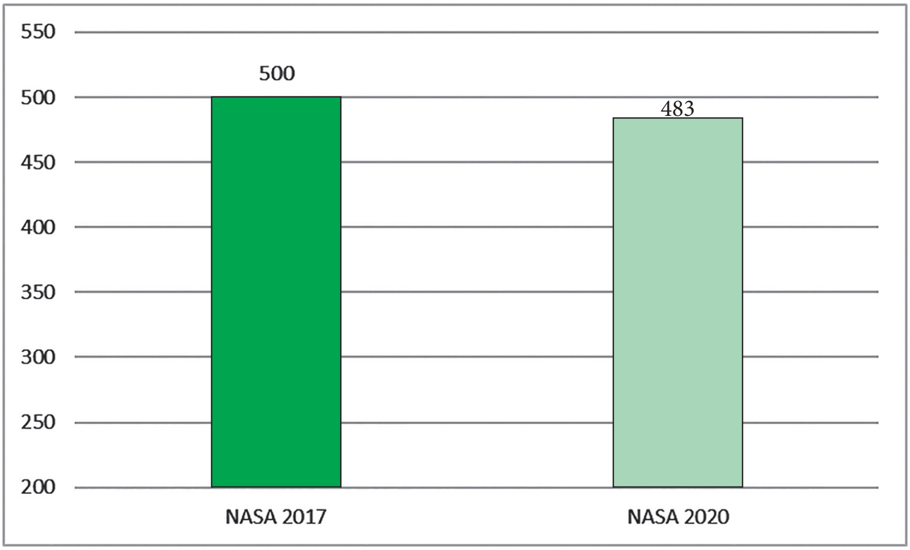


Figure 11 shows the comparison of the mean in mathematics between NASA 2017 and NASA 2020. The mean score of 2017 was 500, whereas in 2020 it was 483. The trend reveals a decrease during these three years, with the mean score falling by 17 scale score. This shows that learning in the classroom setting has deteriorated during those years. As a result, the mean score has noticeably decreased.

The standard error and confidence intervals were obtained by the Bootstrap method (Table 20).

Table 20 The standard error and confidence intervals obtained by the Bootstrap method

		Statistic	Std. Error	Bootstrap*		BCa 95% Confidence Interval	
				Bias	Std. Error	Lower	Upper
Mean_10PV	N	21814		0	0	.	.
	Mean	483.4616	0.31221	-0.0112	0.3025	482.8128	483.9995
	N	21814		0	0	.	.

Note: * indicates, unless otherwise noted, that bootstrap results are based on 450 bootstrap samples.

Proficiency level defined by cut scores

The NASA 2020 Framework has defined six levels of proficiency. According to the framework, the lowest level was level 1 (Below basic level) and the highest level was Advanced level. The scale score ranges set in NASA 2017 were used in NASA 2020 to make the levels comparable.

Table 21 Cut scores for the six proficiency levels in mathematics

Proficiency level	Scale score range
Level 6: Advanced level	606 and above
Level 5: Proficient 3 level	553–606
Level 4: Proficient 2 level	501–553
Level 3: Proficient 1 level	448–501
Level 2: Basic level	395–448
Level 1: Below basic level	395 and below

Table 21 shows that Basic level covers 395–448 scale score and Proficient level 1 represents 448–501, whereas Proficient level 3 denotes 553–606 scale. These benchmarks were set in 2017 and the same benchmarks are used to for consistency in comparison.

Proficiency level-wise distribution of the students

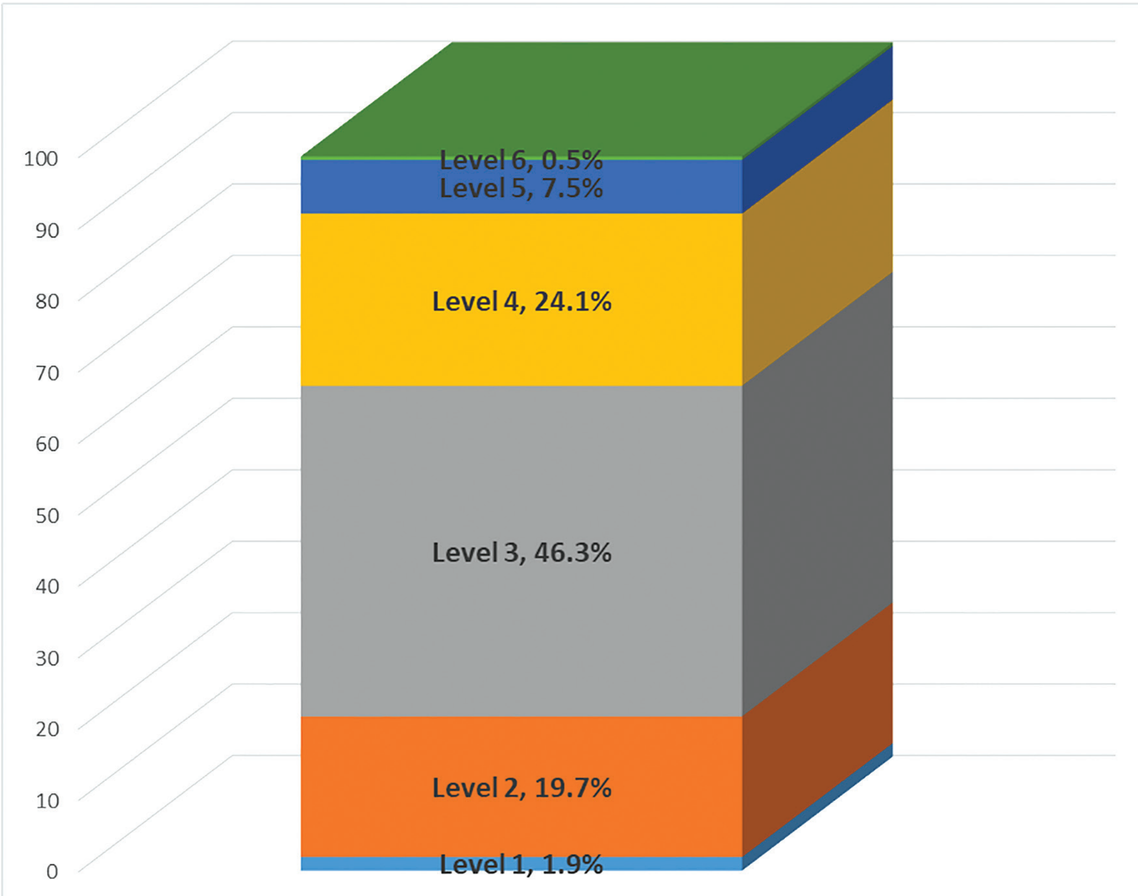
The test scores were also analysed in terms of the six proficiency levels of students' achievement. The levels included are 1, 2, 3, 4, 5, 6, from lowest (Below basic) to highest (Advanced) level. Scores below 395 were considered level 1, while scores within the range of 395–448 were categorised as level 2. Similarly, scores within the ranges of 448–501 and 501–553 were labelled level 3 and level 4 respectively. Scores within the range of 553–606 and above 606 were categorised as level 5 and level 6 respectively.

Table 22 Distribution of students in the six levels of proficiency

Proficiency level	Score range	N	% of students
Level 6: Advanced level	606 and above	2371	0.5
Level 5: Proficient 3 level	553–606	34853	0.5
Level 4: Proficient 2 level	501–553	111460	24.1
Level 3: Proficient 1 level	448–501	214427	46.3
Level 2: Basic level	395–448	90943	46.3
Level 1: Below basic level	395 and below	8614	1.9
Total		462668	100

Table 22 reveals that 1.9% were below the basic level, whereas only 0.5% were at an advanced level. Most students fall under Proficient level 1 and Proficient level 2, which comprised 46.3% 24.1% respectively, and only 7.5% of respondents were categorised as Proficient level 3. Figure 12 shows the percentage of students across the six proficiency levels.

Figure 12 Percentage of students at the six levels of proficiency



The data in Figure 12 shows that if students in level 4 or above are taken as achieving the minimum level of proficiency, 32.1% students achieved above the minimum level of proficiency where as 67.9% of students having shown their performance to be below minimum level of performance. The results indicate that learning opportunities in mathematics were not adequate to the majority of students.

Overall mean score by province in mathematics

The weighted mean score stratified by the seven provinces is presented in Figure 13. The blue horizontal line represents the national mean in 2020.

Figure 13 Overall mean score in mathematics by province

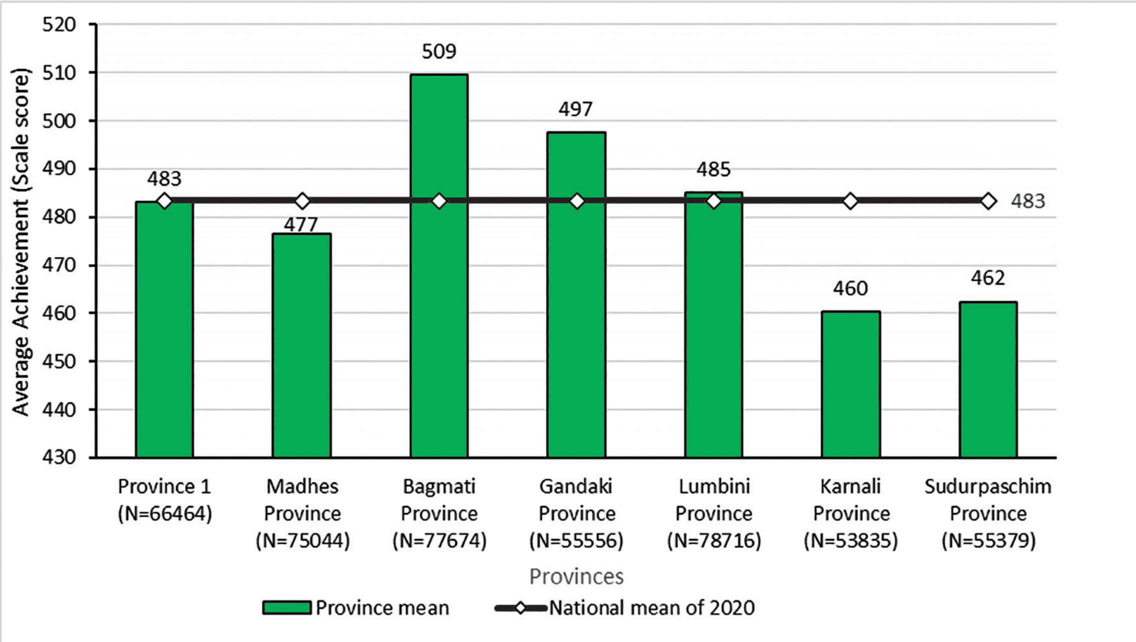


Figure 13 shows a comparison of the overall mean score in mathematics of the seven provinces. The blue horizontal line represents the national average (483) of all provinces. Karnali has the lowest mean (460), followed by Sudurpaschim (462), which are below the national average, and the highest means are Bagmati province (509) followed by Gandaki (497), which are above the national average. Province 1 and Lumbini maintained the national mean score, comprising 483 and 485 respectively.

Province-wise student performance level

The Federal Democratic Republic of Nepal is divided into seven provinces and 753 local government units. While picking up the schools as PSUs, provinces were regarded as strata. The average scores described in this section are the transformed/scale score of 483 national average. The national mean is taken as a reference to contrast with the provincial mean. Those provinces whose average score exceeds the mean score are acknowledged as better-performing provinces, whereas those with an average score below 469 are presumed to be of substandard performance. The descriptive measures of students province-wise are presented in Figure 14.

Figure 14 Province-wise student performance level in mathematics

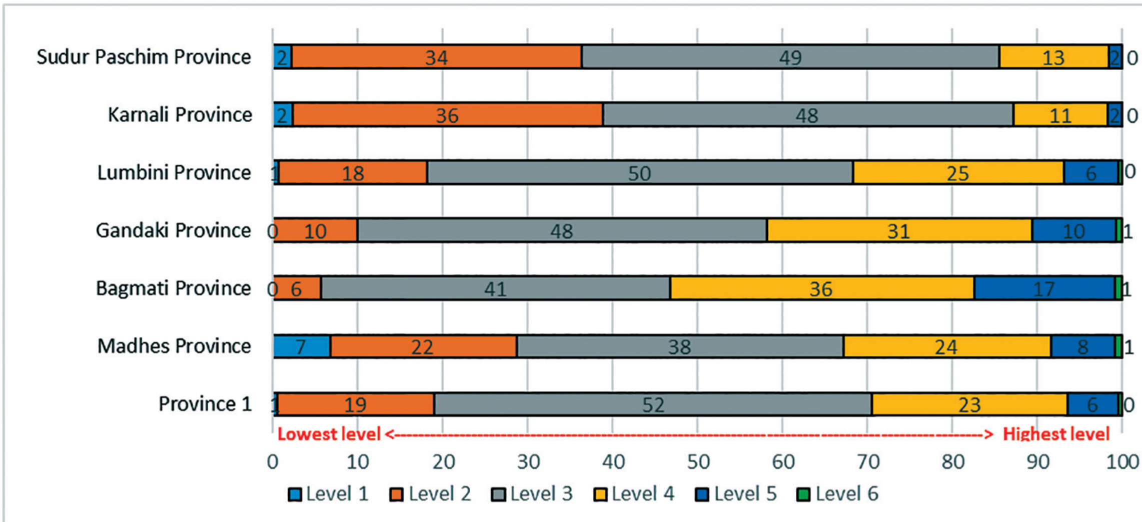


Figure 14 shows that Madhesh Province has the greatest number (7%) of Below basic-level students, whereas Gandaki and Bagmati provinces negligible number of Below basic-level students. Similarly, Province 1, Lumbini, Karnali and Sudurpaschim negligible number of Advanced-level students. The figure also reveals that there is a variation in the proportion of students who meet the minimum level of learning according to province. Both Karnali and Sudur Paschim provinces lie at the bottom in the proportion of students who meet the minimum level of performance, while Bagmati and Gandaki are at the top. The province-wise proportion of students who achieved above the minimum level of proficiency are: Province 1 (29%), Madhesh Province (33%), Bagmati Province (54%), Gandaki Province (42%), Lumbini Province (31%), Karnali Province (13%), and Sudur Paschim Province (15%). Those students who passed the Proficient 1 level are considered as achieving the minimum level of proficiency in mathematics (see definition of minimum level of proficiency in the NASA-2019 report published in 2020.

Results by gender, ethnicity and home language in mathematics

Achievement by gender

Based on the gender of the respondents, students were categorised into three sections – boys, girls and not specified – for uniform and proportionate results. Girls and boys should have equal opportunity and support in their studies; however, the results vary between them. To what extent this has been realised among mathematics students in grade 8 has been an area of investigation. With gender as an implicit stratum, a comparison was made between the number of students in the six defined proficiency levels, as presented. Figure 15 presents the achievement of girls and boys in Mathematics.

Figure 15 Achievement in mathematics by gender

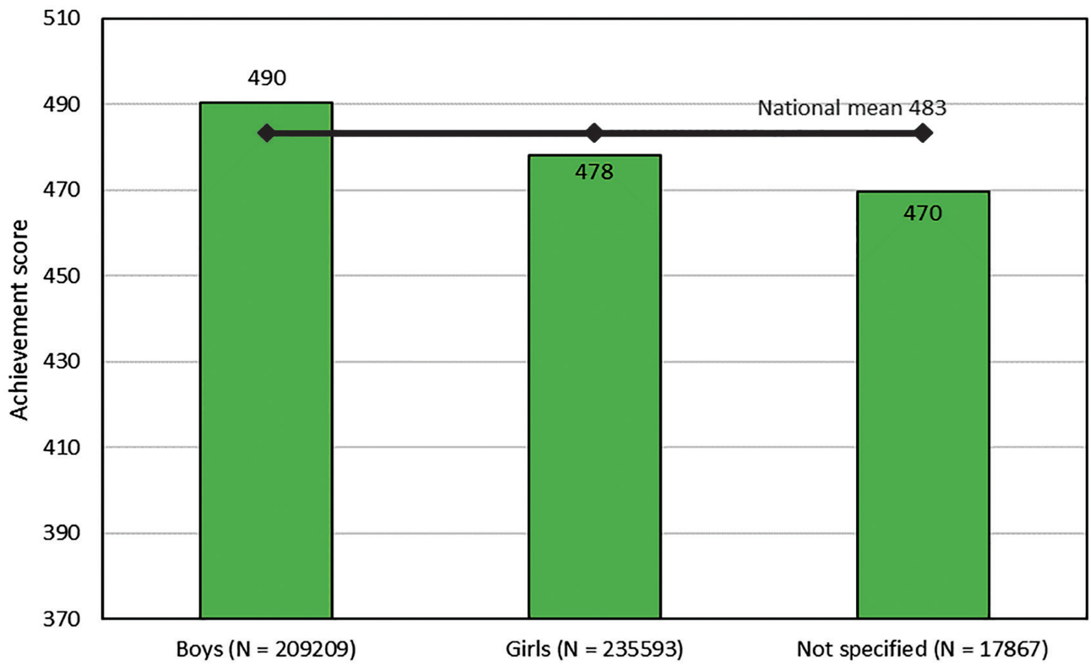


Figure 15 shows that students who did not specify their gender were the most disadvantaged (470), followed by girls (478), whereas boys scored 490, which is slightly above the national average. The difference in mean achievement score is significant at 95% confidence level ($p<0.05$) and the effect size is medium (Cohen's $d = 0.26$) in the population, because these students remained at Below basic level and could not learn any one of the content matters adequately. Figure 16 presents the distribution of students in six proficiency levels.

Figure 16 Distribution of students in different levels of performance by gender

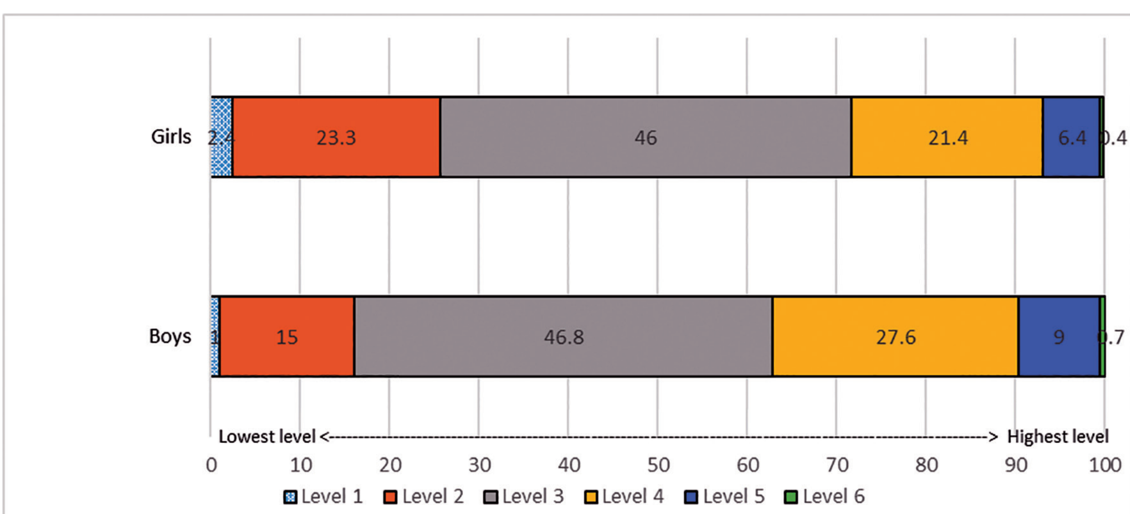


Figure 16 shows the variations in proficiency level based on gender. The achievement of students varies between girls and boys at different levels. As can be seen, only 15% of boys scored Level 2, whereas 23% of girls were placed in Level 2. Although there was similarity in the results between boys (46.8%) and girls (46%) at Level 3, a distinct variation was found at Level 4. Only 21.4 % of girls, compared to 27.6% of boys, were in the same level. Similarly, 6.4% of girls were in Level 5, whereas 9% of boys were in proficiency Level 5. Overall, the distribution of boys was slightly higher than girls in the higher proficiency levels.

Achievement by age group

In NASA 2020, the respondents were asked to provide their age as a background variable to analyse the relationship between age and educational achievement. The age groups of the students were categorised into five strata: 12 years or less, 13 years, 14 years, 15 years and 16 years or above. The details of mean score of their educational achievement is presented in Figure 17.

Figure 17 Achievement in mathematics by age group

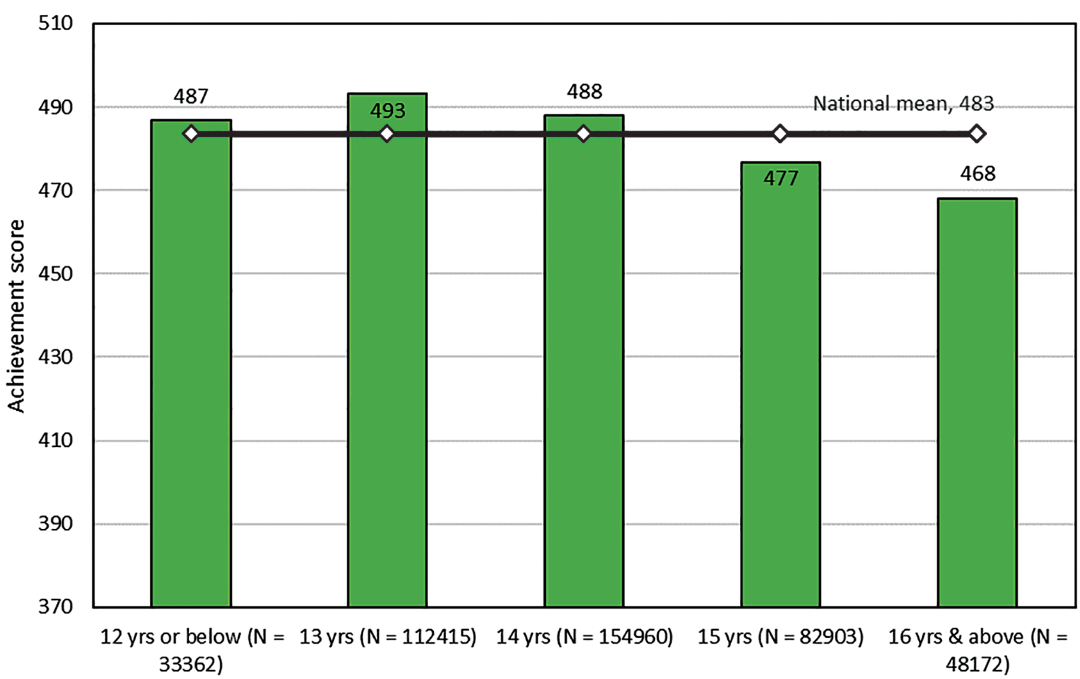


Figure 17 reveals that the 13 years age group achieved the highest mean score (493), and the lowest mean score (468) was for the age group 16 or above. The age groups 12 years or below, 13 years and 14 years scored higher than the national average mean of 483, whereas the age groups 15 years and 16 years or above scored below the national mean. Overall, the results indicate that 13 years has the highest achievement, and as the age increases, the educational achievement decreases.

Home language and achievement

NASA 2020 investigated the association of home language with students' educational achievement. In the study, 449,846 respondents stated that they spoke Nepali as their home language, whereas 12,822 stated that they spoke other languages at home. The achievement of students was as presented in Figure 18.

Figure 18 Achievement in mathematics by home language

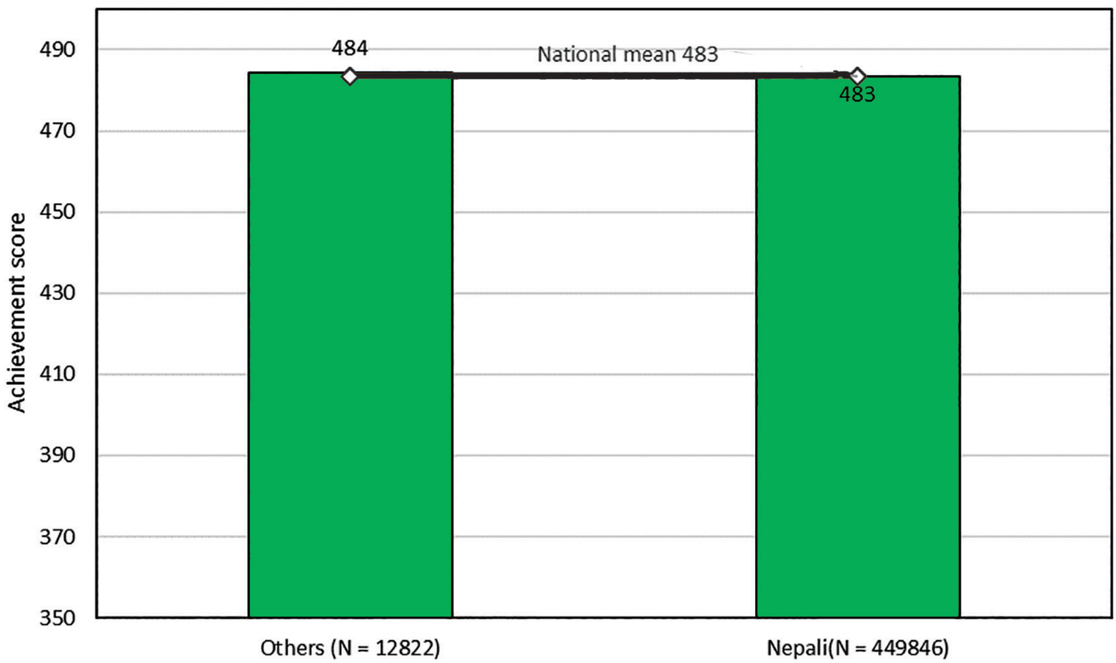


Figure 18 reveals that Nepali speakers at home scored 483, which is very slightly lower than the mean score of 484 of students who spoke other languages at home. Interestingly, this result contradicts previous NASA results (ERO, 2013; ERO, 2015; ERO, 2016; ERO, 2018; ERO, 2019). The results indicate that there is no significant difference in educational achievement in Mathematics between Nepali speakers or other language speakers at home.

Achievement by family size of students

In NASA 2020's background questionnaire, respondents were asked to state the size of their family. The sizes ranged from four to ten or more members. Table 23 illustrates the association of family size and students' educational achievements.

Table 23 Achievement in mathematics by family size

Number of family members	Mean	N	Std. deviation	Std. error of mean
4 or less	495.6971	124297	46.38461	0.13157
5	485.9804	110663	43.22271	0.12993
6	480.9481	82860	43.91496	0.15256
7	475.4769	48496	41.98469	0.19065
8	474.7918	26983	42.53167	0.25892
9	475.1799	14489	47.41655	0.39393
10 or more	476.3512	36457	45.73569	0.23953
Total	484.7918	444244	45.11227	0.06768

Table 23 depicts that the learning achievement of students was highest (495.69) when the family size was four members, and with an increased family size of eight members, the achievement in mathematics was 474.79. The results indicate that the higher the number in the family, the lower the learning achievement in mathematics.

School–home distance and achievement

In NASA 2020, one of the areas of questionnaire was the distance between school and home. To analyse this in association with achievement, the distances were categorised into five groups of walking times, comprising: up to 15 minutes, 16–30 minutes, 30 minutes to 1 hour, 1–2 hours, more than 2 hours (Figure 19).

Figure 19 Students' achievement by the distance between home and school

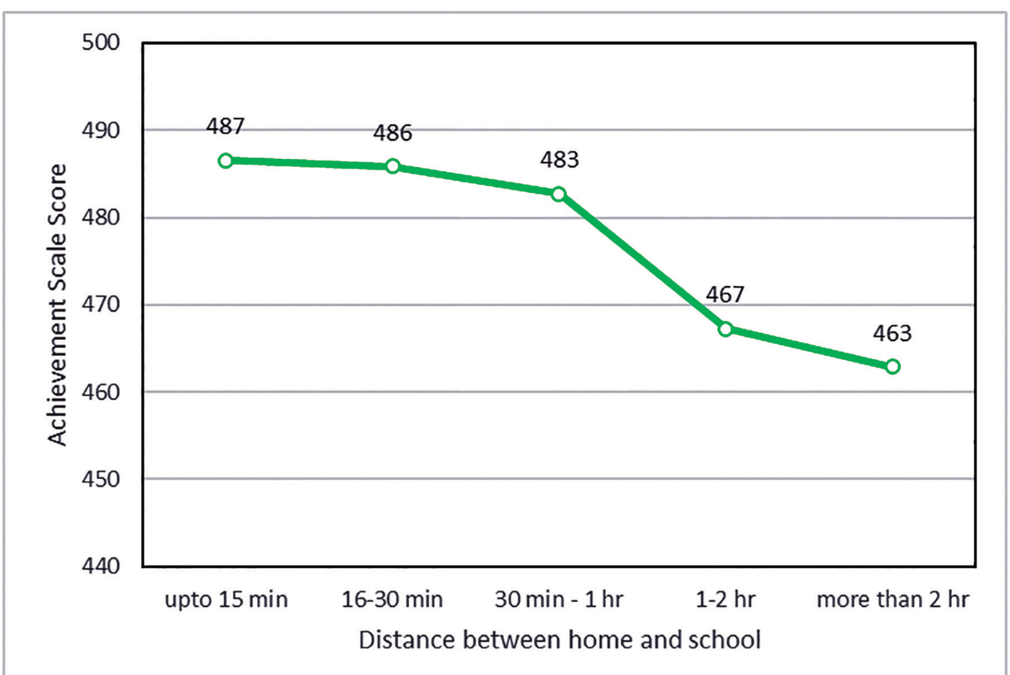


Figure 19 illustrates the results of the distance between school and home in relation to respondents' educational achievement in mathematics. The achievement of students whose homes were within 15 minutes, 16–30 minutes and 30 minutes to 1 hour walking distance of their school scored higher than the national mean, scoring 487, 486 and 483 respectively, whereas students whose homes were 1–2 hours or more than 2 hours walking distance from their school achieved below the national mean, with 467 and 463 respectively. The data revealed that up to a 15-minute walk between school and home is the ideal distance, as these students scored highest (487), whereas students with a walk of 2 hours or more achieved the lowest score (i.e. 463).

Medium of instruction and achievement

In NASA 2020, the medium of instruction was also investigated by asking respondents to state their medium of teaching for mathematics. The respondents' answers were categorised as Nepali, English and Other & not specified languages (Figure 20).

Figure 20 Achievement in mathematics by medium of instruction in the school

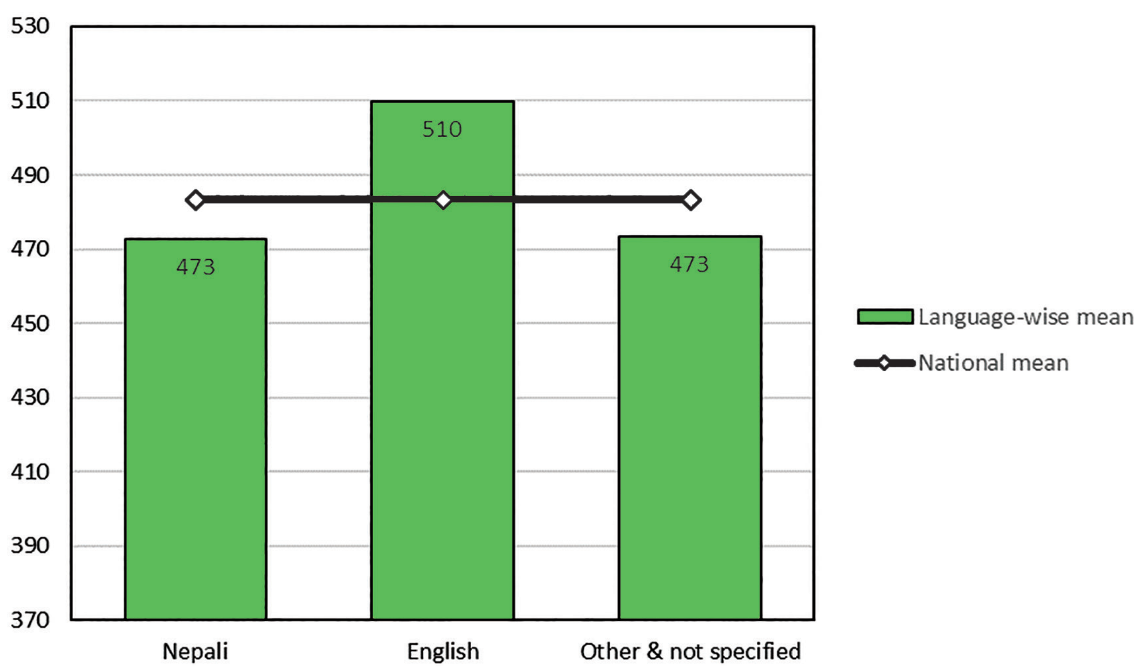


Figure 20 depicts the mean scores of students based on their medium of instruction. As can be seen, students whose medium of teaching was English scored highest (510), whereas students whose medium of instruction was Nepali and Other & not specified languages scored below the national mean score (473 each). An interesting fact revealed by NASA 2020 is that English is the best medium of instruction for achieving a higher score. However, it is possible that high ability students are disproportionately studying in English medium schools.

Achievement by time spent in different activities – before and after school hours

Students' engagement in different activities before and after school was put into seven categories – watching TV, playing with friends and chatting, household work, study/ homework, working for a wage, studying other books and supporting siblings – which were measured from no time given to more than four hours.

Table 24 Achievement by time spent in different activities – before and after school hours

Category	Time spent				
	No time given	Less than 1 hr	1–2 hr	2–3 hr	More than 4 hr
Watching TV	465.1	489.1	503.5	503.5	475.1
Playing with friends/chatting	475.4	489.5	488.5	481.0	475.9
Household work	470.3	493.4	489.4	481.8	472.4
Study/homework	458.7	472.0	489.5	496.0	494.0
Working for a wage	492.8	481.0	473.6	470.5	469.9
Studying other books	470.6	493.2	491.0	481.3	468.1
Supporting siblings	479.6	492.3	488.4	477.1	467.9

In Table 24, it can be seen that the highest achievement score is found among those students in the habit of watching TV for 1–2 hours and 2–3 hours (mean = 503.5), whereas the lowest scores are found among those not spending time watching TV (mean = 465.1) and not giving time for study/homework (mean = 458.7).

Achievement by provision of study support at home

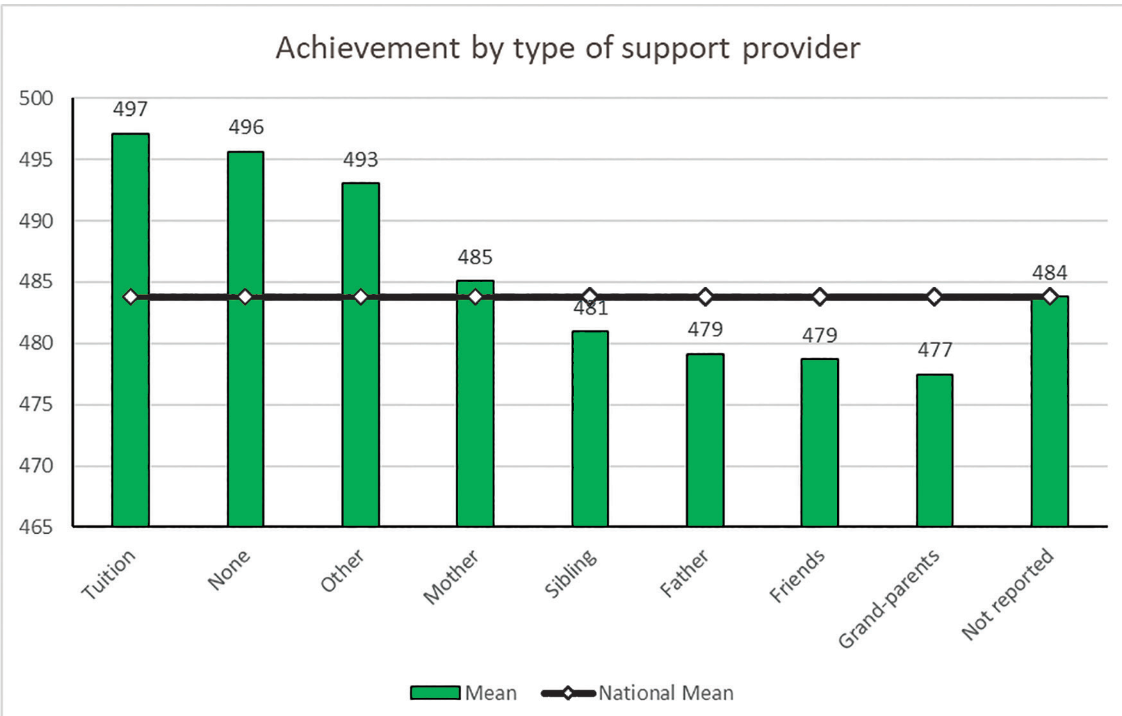
In NASA 2020, one of the questions students were asked was in relation to support in their studies at home. The responses were categorised into nine groups, including father, mother, sibling, friends, grandparents or tutors, and are presented in Table 25.

Table 25 Achievement by support provider for study at home

Support for study at home	Mean	N	Std. deviation	Std. error of mean
Tuition	497.1159	40985	42.06534	0.20778
None	495.652	18641	45.20722	0.33111
Other	493.0156	11312	45.83728	0.43096
Mother	485.0903	29897	46.27267	0.26762
Sibling	480.9569	211430	42.6263	0.0927
Father	479.1415	71490	46.18688	0.17274
Friends	478.7143	30400	41.35666	0.2372
Grandparents	477.4633	2242	46.31804	0.97817
Not reported	483.8009	46272	60.54001	0.28144
Total	483.382	462668	45.9388	0.06754

Table 25 shows that support provided by a tutor was most effective, with students achieving 497 scale score. Similarly, students who were supported by their mother or other achieved 485 and 493 respectively, which were above the national average. Students who were supported by grandparents and friends achieved 477 and 478 respectively, which were below the national average and remained less effective. The results can be presented in a bar diagram as in Figure 21.

Figure 21 Achievement by support provider for home study



The achievement of students who take tuition or are supported by their mother, other and none are significantly higher than remaining ($p=0.000<0.001$, $\eta2=0.015$). It means that students who are showing better results are receiving support from either their mother or someone from outside the home. On the other hand, the achievement scores of students getting support from siblings, their father, friends and grandparents are less than the national average score. The results also suggest that those students taking tuition classes, in the habit of self-learning (none) and supported by other have better performance in mathematics.

Achievement by career aspiration of students

In NASA 2020, students could indicate their future career goal, although most (highest frequency) did not report this. Their responses were categorised as teacher, government official, private sector employee, businessman, work abroad, farmer, and doctor and engineer as in Table 26.

Table 26 Achievement by career aspiration of students

Career aspiration	Mean	N	Std. deviation	Std. error of mean
Teacher	462.0688	116909	36.87014	0.10783
Govt. official	488.2592	69517	41.77858	0.15846
Private sector employee	480.1791	7321	41.26981	0.48232
Businessman	475.7919	20035	44.88478	0.31711
Work abroad	479.2385	14864	43.71691	0.35857
Farmer	440.2423	207	51.84924	3.60512
Doctor and engineer	462.5229	11867	42.58256	0.3909
Not specified	495.3046	221948	47.43944	0.1007
Total	483.382	462668	45.9388	0.06754

Table 26 shows the students' choices in their future career and their achievement level. The overall mean score was 483 based on their choice of future career goal in relation to their achievement, which can be displayed as in Figure 22.

Figure 22 Achievement by future career aspiration of students

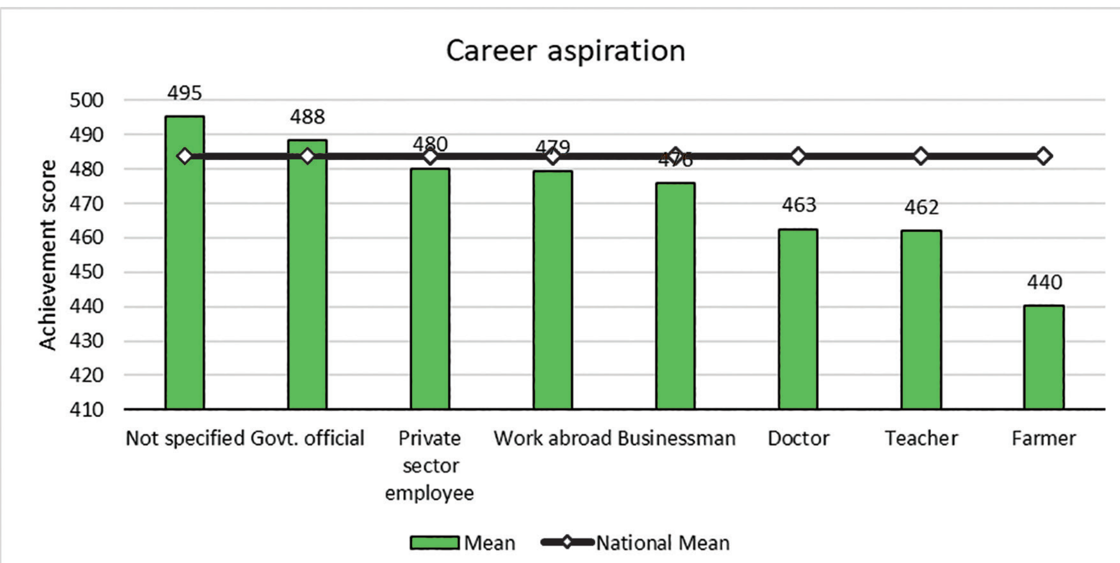


Figure 22 shows that students who did not specify their future aim from the provided list and those interested working for the government have higher achievement scores than their counterparts, and their scores are also found to be higher than the national average. Those students aiming to be a farmer in the future have a lower performance in mathematics (mean = 440). Additionally, it was interesting that those aiming to be a doctor or teacher also have lower performance in mathematics compared to all others except those wishing to be a farmer.

Achievement by type of activities in school leisure time

In NASA 2020, students were asked to state their leisure-time activities in school. Their engagement during leisure time was categorised as classwork, homework, play, return home. The details have been given in Table 27.

Table 27 Achievement by type of activities in school leisure time

Leisure-time activity	Mean	N	Std. deviation	Std. error of mean
Classwork	485.62	258404	43.32	0.08523
Homework	484.30	145905	45.37	0.11876
Play	483.78	30718	50.04	0.28551
Return home	452.84	11190	44.20	0.41787
Not specified	460.15	16452	65.98	0.51442
Total	483.38	462668	45.94	0.06754

The results indicate that those students who spent their leisure time in classwork (mean = 485.62, SD = 43.32) and homework (mean = 484.30, SD = 45.37) performed better in mathematics. However, those engaged in uncategorised activities (not specified) have the lowest performance (mean = 460.15, SD = 65.98).

Figure 23 Achievement by type of activities in school leisure time

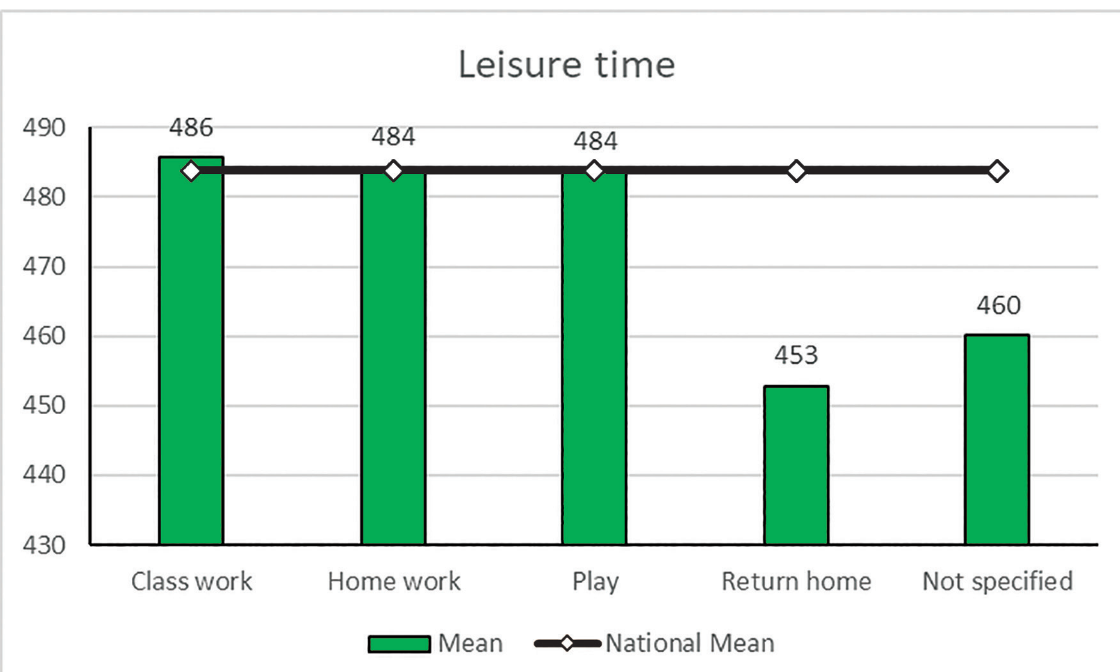


Figure 23 reveals that students using their school leisure time for classwork, homework and engaging in games and sports achieved 486, 484 and 484 respectively, which were above the national average, whereas those students who return home or do not have specific activities achieved below the national average.

Achievement by extra-curricular activities

Extra-curricular activities are considered an important part of the all-round development of students in schools. In NASA 2020, achievement varied on the basis of the students' involvement in extra-curricular activities as shown in Table 28.

Table 28 Achievement by extra-curricular activities

Extra-curricular activity	Mean	N	Std. deviation	Std. error of mean
Not specified	447.0424	17666	55.49242	0.41751
Regular	484.9594	161414	47.43089	0.11806
Sometimes	486.0003	261632	43.3014	0.08466
Never	469.8236	21957	42.01257	0.28353
Total	483.382	462668	45.9388	0.06754

Table 28 depicts that students who were involved in extra-curricular activities either regularly or sometimes achieved above the national average, whereas students who never engage in extra-curricular activities achieved below the national average. The data can be shown in diagram form as in Figure 24.

Figure 24 Achievement by extra-curricular activities

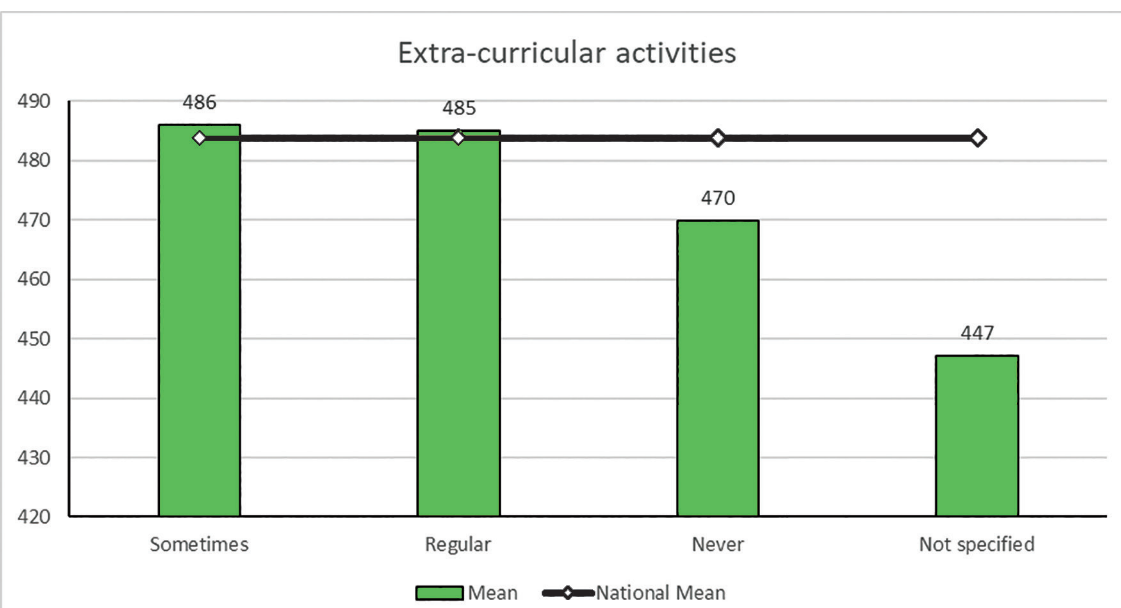


Figure 24 shows that the students in the habit of participating either regularly or sometimes in extra-curricular activities have better performance in mathematics than the national average. However, those never participating in or not specifying whether they participate in extra-curricular activities have poorer performance than others and the national average.

Achievement by mother's education

Mother's education is one of the indicators of a family's socio-economic status. In this analysis, the student achievement score is compared with the student’s mother's education status, which is presented in Table 29.

Table 29 Achievement in mathematics by mother's education

Mother's education	Mean	N	Std. deviation	Std. error of mean
Illiterate	470.7823	140040	39.99264	0.10687
Just literate	480.2205	116207	38.56269	0.11312
Grade 8 pass	483.2366	82736	42.51071	0.14779
Grade 10 pass	499.8578	61299	45.41682	0.18344
Grade 12 pass	515.1191	31363	47.34718	0.26735
Bachelor's	526.4532	12161	51.2975	0.46516
Master's or above	545.589	4781	47.27797	0.68372
Not specified	434.894	14080	56.45703	0.47579
Total	483.382	462668	45.9388	0.06754

The relationship between mother's education and the student's learning achievement score is also presented in the Figure 25.

Figure 25 Achievement by mother's education

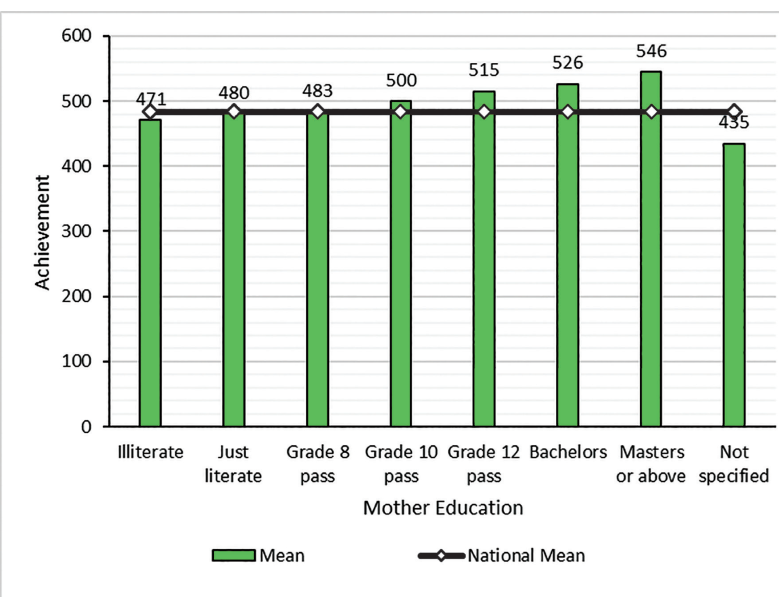


Figure 25 shows that the achievement score of students having a mother with a (Grade 12 pass (515), Bachelor's 526 and Master's or above 546) was found to be higher than the national average. That score is highest in the case of qualification to the level of Master's and above. Furthermore, the lowest achievement scores of 471 and 435 respectively were found to be in the case of those students having an illiterate mother and those not specifying their mother's education. The score achieved by students of an illiterate mother was 471, which was below the national average, whereas the score achieved by students whose mother had a Master's degree qualification was 546. The results indicate that there was a significant difference in the achievement of students whose mothers were educated compared to those whose mothers were uneducated.

Achievement by mother's profession

Similarly, mother's professional background has a positive effect on the achievement of students. There were nine categories of mother's profession in NASA 2020. The details of the mothers' professions as given by the students have been presented in Table 30.

Table 30 Achievement in mathematics by mother's profession

Mother's profession	Mean	N	Std. deviation	Std. error of mean
Agriculture and household work	475.1077	269843	40.16332	0.07732
Household work only	497.5889	99321	46.53034	0.14764
Work in other's house	472.4378	4107	44.78384	0.6988
Work for wage	481.8771	6105	43.23027	0.55328
Working abroad	488.4612	8471	41.2679	0.44838
Teaching	514.9342	11485	49.97339	0.46632
Business	498.898	30694	45.56693	0.26009
Government job	512.7193	9493	48.53096	0.49809
Other	514.9803	8562	47.17105	0.50978
Not specified	445.3424	14586	65.35993	0.54117
Total	483.382	462668	45.9388	0.06754

Table 30 illustrates that there is a direct correlation between the mother's profession and the achievement of students. Students whose mother is engaged in teaching or a government job achieved 515 and 513 respectively, whereas students whose mothers were involved in working in others' houses or in agriculture achieved 472 and 475 respectively, which are fairly below the national average. This can be presented in diagram form as in Figure 26.

Figure 26 Achievement in mathematics by mother's profession

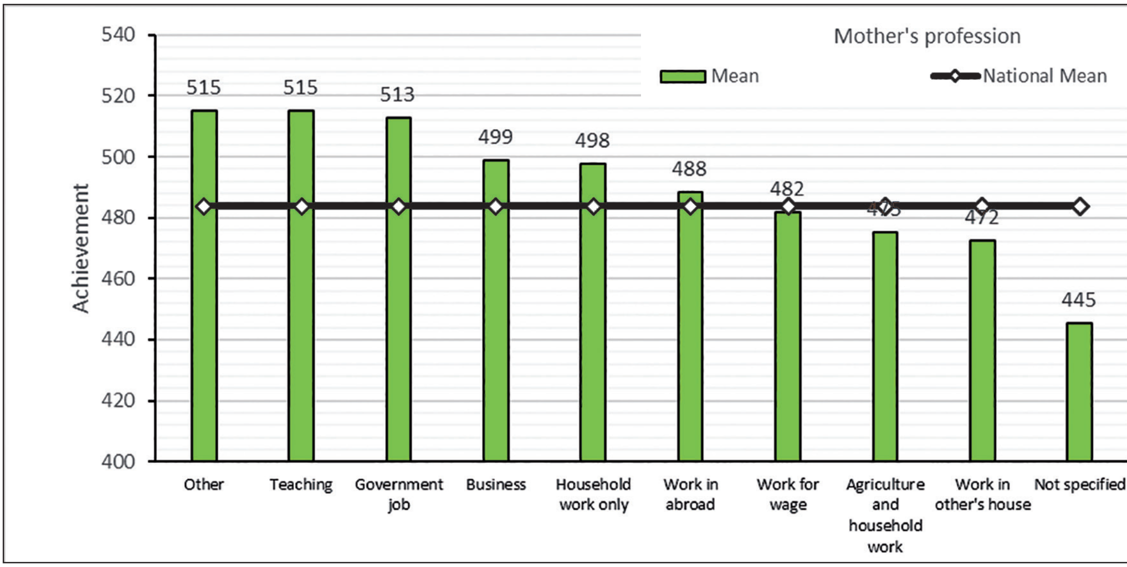


Figure 26 shows that the achievement score of students having mothers whose occupation is teaching, a government job, business, household work only, working abroad and other have better performance in mathematics when compared to the national average and the remaining categories. However, the result is poorest among those students who did not specify their mother's job.

Achievement by father's education

Similar to mother's education level, the study asked students about their father's education level. Their responses were categorised into seven headings, such as Illiterate, Just literate, Grade 8 pass, Grade 10 pass, Grade 12 pass, Bachelor's, and Master's or above. The responses from students and their achievements are presented in Table 31.

Table 31 Achievement in mathematics by father's education status

Father's education	Mean	N	Std. deviation	Std. error of mean
Illiterate	465.0383	61461	38.7951	0.15649
Just literate	476.5968	90163	37.06439	0.12344
Grade 8 pass	477.3406	106494	39.32138	0.12049
Grade 10 pass	487.3889	101394	43.84367	0.13769
Grade 12 pass	503.3059	54699	46.42912	0.19852
Bachelor's	519.9101	21200	49.18056	0.33777
Master's or above	538.4194	12597	48.82017	0.43498
Not specified	443.7324	14660	63.82716	0.52716
Total	483.382	462668	45.9388	0.06754

Table 31 depicts that those students whose parents were illiterate or below grade 8 level achieved below the national average, whereas those students whose father passed grade 10, grade 12, or had bachelor's- or master's-level education, have achieved fairly above the national average. This is shown as a bar diagram in Figure 27.

Figure 27 Achievement by father's education status

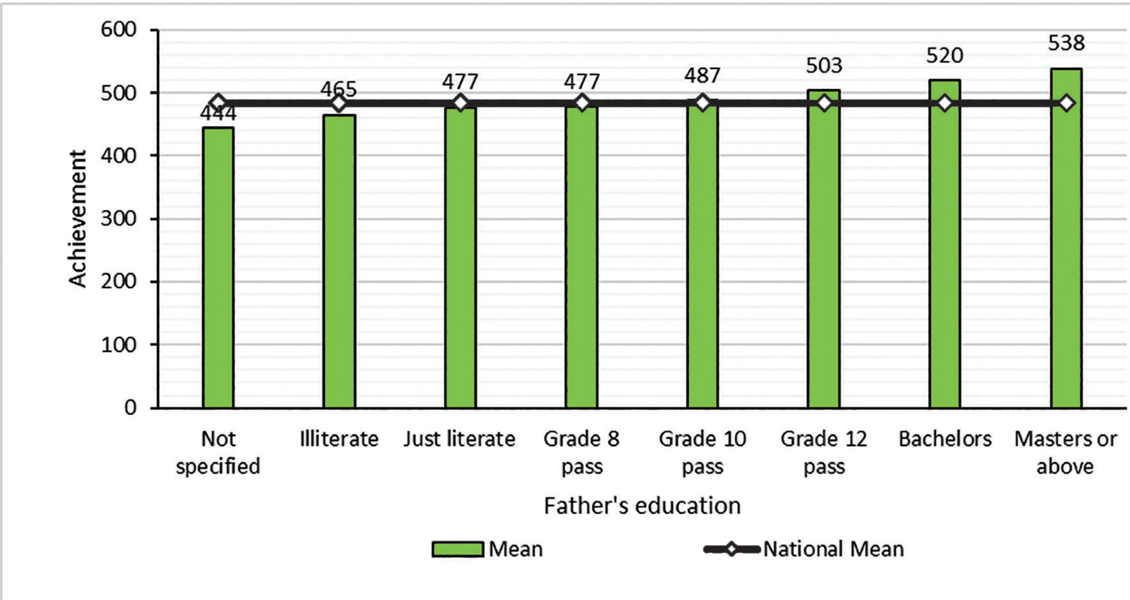


Figure 27 shows that the achievement score of students was found to be higher than the national average among those with a father having qualifications above school education (Grade 10). However, the result is highest for Master's and above (mean = 538) and lowest for those not specifying their father's educational qualification (mean = 444).

Achievement by father's occupation

In NASA 2020, the father's occupation has also shown a direct correlation with the achievement of students. The different job categories for NASA 2020 are shown in Table 32.

Table 32 Achievement in mathematics by father's occupation

Father's profession	Mean	N	Std. deviation	Std. error of mean
Agriculture and household work	469.9945	134483	39.57804	0.10792
Household work only	460.9116	9948	41.11586	0.41224
Work in other's house	464.4285	8491	37.53697	0.40736
Work for wage	482.5459	34734	37.0112	0.19859
Working abroad	482.6825	107541	42.38185	0.12924
Teaching	506.8175	14212	48.59307	0.40761
Business	499.4774	67429	46.6308	0.17958
Government job	501.1342	32725	46.48856	0.25699
Other	509.4864	32456	44.73225	0.2483
Not specified	456.3889	20651	60.42054	0.42045
Total	483.382	462668	45.9388	0.06754

Table 32 shows that students whose fathers are involved in teaching, have a government job or who work in business have achieved higher scores than the national average, whereas students whose fathers have been engaged in agriculture or who work in others' houses achieved below the national average. This data is displayed in Figure 28 as a bar diagram.

Figure 28 Achievement in mathematics by father's occupation

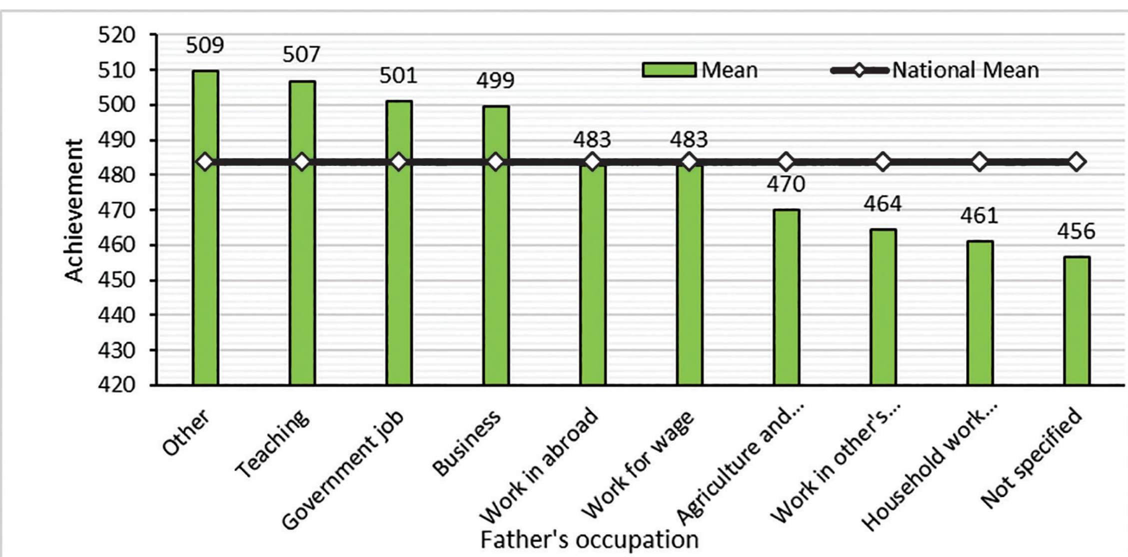


Figure 28 shows that those students having fathers engaged in teaching, a government job or other have performance that is higher than the national average. However, the results were found to be poor for students whose father works in the household or who did not specify their father's job. Additionally, students whose fathers are engaged in teaching and other professional areas have the best results when compared to other listed categories.

Achievement by possessions at home

Based on the information given by students, data was tabulated regarding the various possessions found in their homes, such as Table for study, Separate study room, Peaceful space to study, Computer for schoolwork, Children's magazine, story/poetry and pictures, Reference book for schoolwork support, Internet, and dictionary. The number of possessions are split into eight categories: none, one, two, three, four, five, six, and seven. Table 33 presents how many home possessions the students had or did not have and their achievement in mathematics.

Table 33 Achievement by number of home possessions

Number of home possessions	N	Mean	Std. deviation	Std. error of mean
None	11603	451.3601	54.72738	0.50806
One	208656	473.0199	39.76623	0.08706
Two	51370	478.4841	42.02983	0.18544
Three	66847	485.2635	42.22689	0.16332
Four	55933	493.9466	45.32514	0.19165
Five	34662	503.7484	47.12013	0.25309
Six	18937	517.7993	51.33242	0.37302
Seven	14660	531.8751	56.44323	0.46618
Total	462668	483.382	45.9388	0.06754

Table 33 depicts that when the number of home possessions increases, the mathematics achievement of students also increases. This fact is shown in Figure 29 as a line graph.

Figure 29 Achievement by number of home possessions

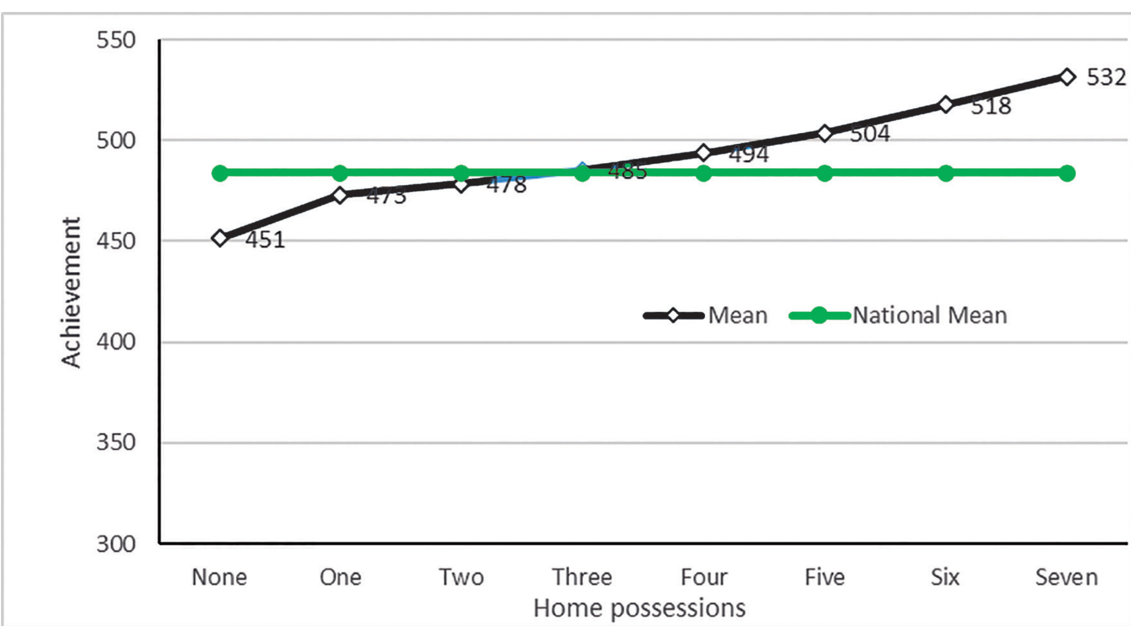


Figure 29 shows that the achievement score was found to be greater than the national average for students having more than three possessions, and lowest for those having only one possession or none.

Achievement by number of books available at home for study

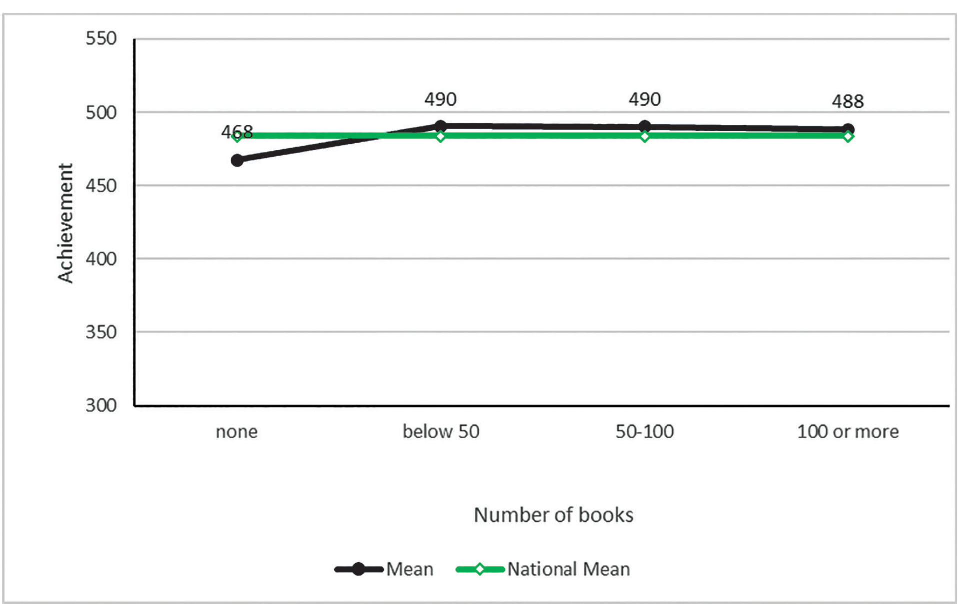
This study also inquired about the number of books available to students for studying at home. The responses are presented in Table 34.

Table 34 Achievement in mathematics by number of books available at home for study

No. of books other than textbook	Mean	N	Std. deviation	Std. error of mean
None	467.53	140133	42.40378	0.11328
Less than 50	490.3813	281014	44.89151	0.08468
50–100	490.0283	27740	49.70578	0.29844
100 or more	488.4684	13781	53.21638	0.45332
Total	483.382	462668	45.9388	0.06754

Table 34 shows the achievement score of students based on the number of books available at home. The results indicate that the achievement score of students was found to be poor among those not having books at home (mean = 467.53). For students having less than 50 (Mean = 490.38) or 50–100 (mean = 490.03) books at home, the scores were relatively good. In fact, the achievement score was found to be higher than the national average in all cases except where there were no books at home. This is also presented in Figure 30 as a line graph.

Figure 30 Achievement in mathematics by number of books available at home for study



Achievement by possession of a separate mobile phone

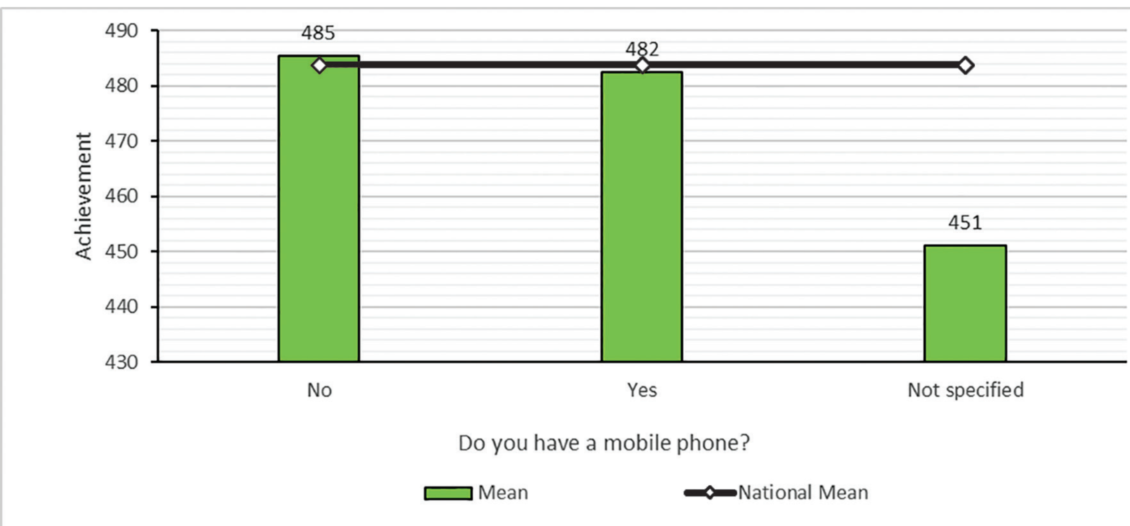
In NASA 2020, one of the questions asked was whether students possessed a separate mobile phone. Their responses are presented in Table 35.

Table 35 Achievement by students' possession of a separate mobile phone

Separate mobile	Mean	N	Std. deviation	Std. error of mean
No	485.3693	349248	44.00649	0.07446
Yes	482.46	94646	49.25974	0.16012
Not specified	451.0607	18774	51.25984	0.37411
Total	483.382	462668	45.9388	0.06754

Table 35 shows that most of the students don't have a separate mobile phone. The data can be presented in bar diagram form as in Figure 31.

Figure 31 Achievement by students' possession of a separate mobile phone



From Figure 31 those students not having a personal mobile phone have better results in mathematics than others and the national average. However, the achievement score was found to be poor for those students who did not specify whether they have a mobile phone. The data reveals that students who have a separate mobile are performing worse than those who do not have one. Thus, it appears a mobile phone does not support studying.

Achievement by access to social media

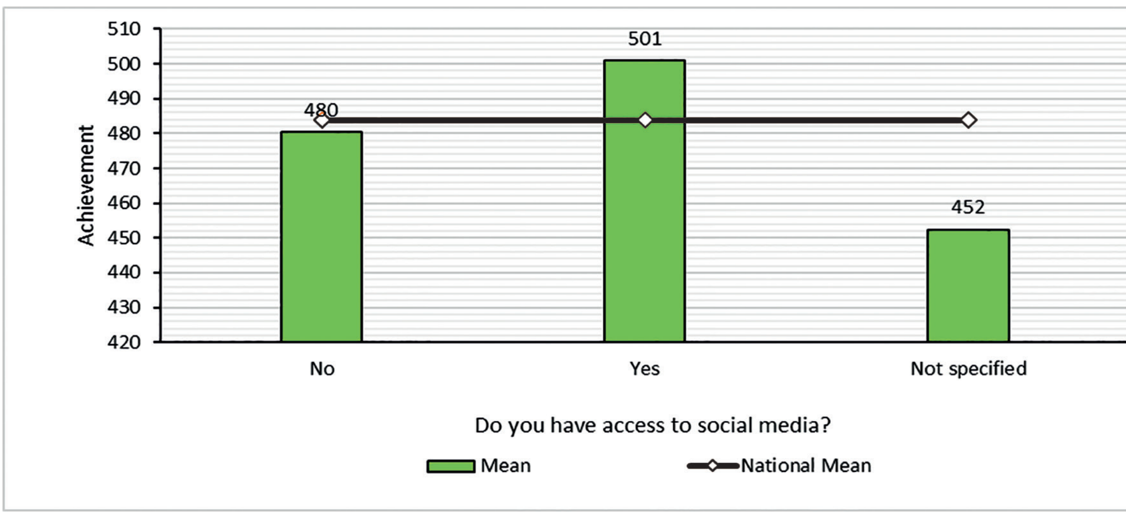
The study also asked a question regarding students' access to social media. The responses of the students are presented in Table 36.

Table 36 Achievement in mathematics by students' access to social media

Social media access	Mean	N	Std. deviation	Std. error of mean
No	480.4009	336003	42.40746	0.07316
Yes	500.8542	101635	49.67369	0.15581
Not specified	452.4525	25029	50.32966	0.31813
Total	483.382	462668	45.9388	0.06754

Table 36 shows that students who use social media perform better in mathematics than those who did not specify or who had no use of social media. The result is interesting, and the reason behind this may be that the social media users can solve their personal problems through virtual collaboration and communication. The data has been presented in Figure 32 as a bar diagram.

Figure 32 Achievement by access to social media



The data in Figure 32 reveals that those who operate social media are better performers although the achievement of students who own a mobile phone is weaker than those without (see figure 31). It is possible that those who use social media might have used learning materials like YouTube videos. However, the data does not specifywhat type of social media they have used.

Achievement by home Accessories

Students were asked to respond as to the accessories their family had access to at home - Car, TV, Computer, Motorcycle, and Permanent house. Initially students responded by number of facilities: 0 for none, one, two, three, etc. Later, the response was recoded into Yes/No whether they have an item or not, without considering how many they own. Thus, the categories were changed into 1 or 0 for yes or no. Later, all the categories were summed. The final summed variable of home facilities was used to compare the mean. Achievement based on home facilities is presented in Table 37.

Table 37 Achievement in mathematics by home facilities

Home accessories recorded	Mean	N	Std. deviation	Std. error of mean
None	461.0375	102642	40.99285	0.12795
Only one	474.9847	126931	39.17451	0.10996
Any two	488.2251	117474	42.46878	0.12391
Any three	499.8596	70670	46.19935	0.17379
Any four	519.6893	36275	46.938	0.24645
All five	518.9899	8675	54.26393	0.58259
Total	483.382	462668	45.9388	0.06754

The data in Table 37 reveals that the higher the number of home facilities, the better the achievement of students. Students who did not have any home facilities achieved 461, whereas students who possess four or five accessories achieved round 520 , which demonstrated a direct correlation between home facilities and achievement. This has been presented in Figure 33.

Figure 33 Achievement in mathematics by home facilities

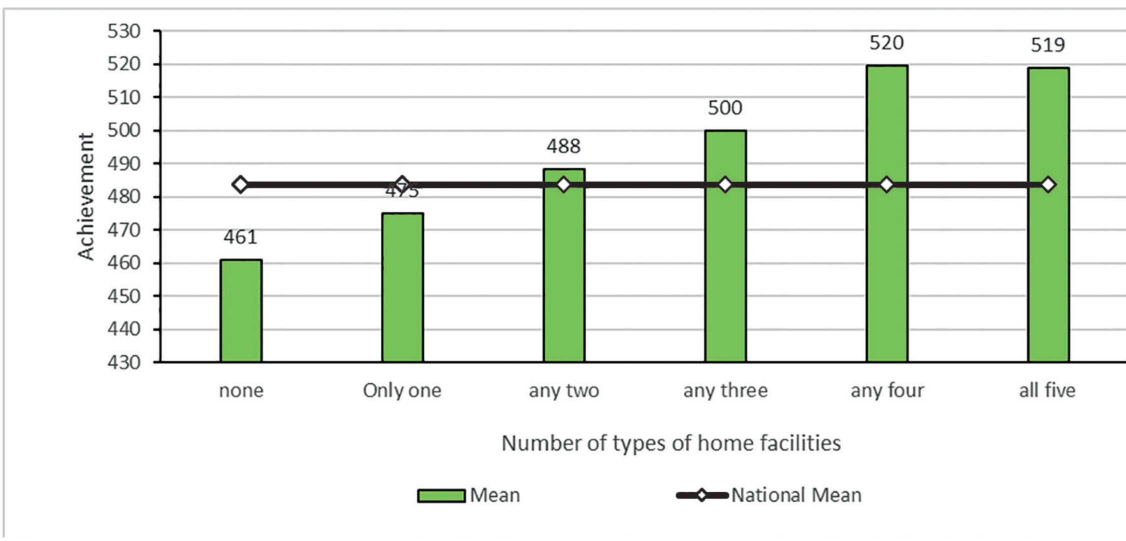


Figure 33 shows that those students having more than two home facilities have better performance in mathematics than the national average. The results further show that those having any four facilities have the highest achievement, and those not having any facilities have poor performance in mathematics. Furthermore, the result is significant ($p = 0.000 < 0.01$) with respect to number of types of home facilities with poor effect size (partial eta squared = 0.144).

Student's perception of teacher's behaviour

The attitude of students towards the teacher's behaviour also plays a vital role in their performance in mathematics. Details of the teacher's behaviour towards students and students' attitudes towards the teacher have been presented in Table 38.

Table 38 Attitudes of students towards teacher's behaviour

Statements	Totally agree	Somewhat agree	Somewhat disagree	Totally disagree	Not specified
Teacher teaches with love and care	80.9	12	1.5	1.2	4.4
Most of the teachers really listen to me	48	32.2	5.8	3.7	10.3
Teachers do not give corporal punishment	35.5	25.5	12.3	14.1	12.5
Teachers treat us equally and care about us	78.2	9	2.7	2.5	7.6
Teachers answer when we ask when we don't understand	85.5	5.5	1.1	1.5	6.5
Teachers give us homework	82.6	7	1.3	1.7	7.4
Teachers give and check homework	77.5	10.8	2	1.7	7.9
Teachers teach full time of the period	63.8	18.6	4.2	2.7	10.6

Table 38 shows the complexities of totally agree and somewhat agree. After merging the four categories – the first two into AGREE and the second two into DISAGREE – the data is shown in Table 39.

Table 39 Students' attitude on the teacher's behaviour

Statements	Agree	Disagree	Not specified
Teacher teaches with love and care	93	3	4
Most of the teachers really listen to me	80	10	10
Teachers do not give corporal punishment	**61**	**26**	**13**
Teachers treat us equally and care about us	87	5	8
Teachers answer when we ask when we don't understand	91	3	7
Teachers give us homework	90	3	7
Teachers give and check homework	88	4	8
Teachers teach full time of the period	82	7	11

Table 39 reveals that many students (more than 26%) believe that teachers give corporal punishment to students. However, 93% of students showed their positive attitude towards their teachers as they stated that teachers show love and care for them. Similarly, 90% of students agreed that their teachers give and check homework. Out of the total students. 87% agreed that they were treated equally by their teachers, and 91% of students agreed that their teacher answers their questions when they do not understand.

Students' attitudes towards school behaviour

Students' performances may also be influenced by their attitudes towards school behaviour. The data from respondents was categorised as the students' like or dislike of coming to school, peers' treatment of them, facilities available to them and their participation in club activities. The data is presented in Table 42.

Table 42 Students' attitudes towards school behaviour

Statements	Totally agree	Somewhat agree	Somewhat disagree	Totally disagree	Not specified
School: I like to come to school.	90.1	4.4	0.7	1.3	3.5
School: Peers treat me fairly as far as possible.	71.7	18.3	1.9	1.7	6.4
School: School facilities, like playing, drinking water and toilets, are good.	81.2	8.9	2	1.9	6
School: I participate in child club and children's programmes.	54.2	26.6	4.1	6.4	8.7

Table 42 shows the attitudes of students towards school. As shown in Table 42, 90% of students agreed that they like to come to school, whereas only 54.7 % students agreed that they participate in club activities. Almost 72% of students agreed that peers treat them fairly as far as possible. Among the students, 81% agreed that they have good facilities for playing, drinking water and good toilets.

Status of bullying in schools

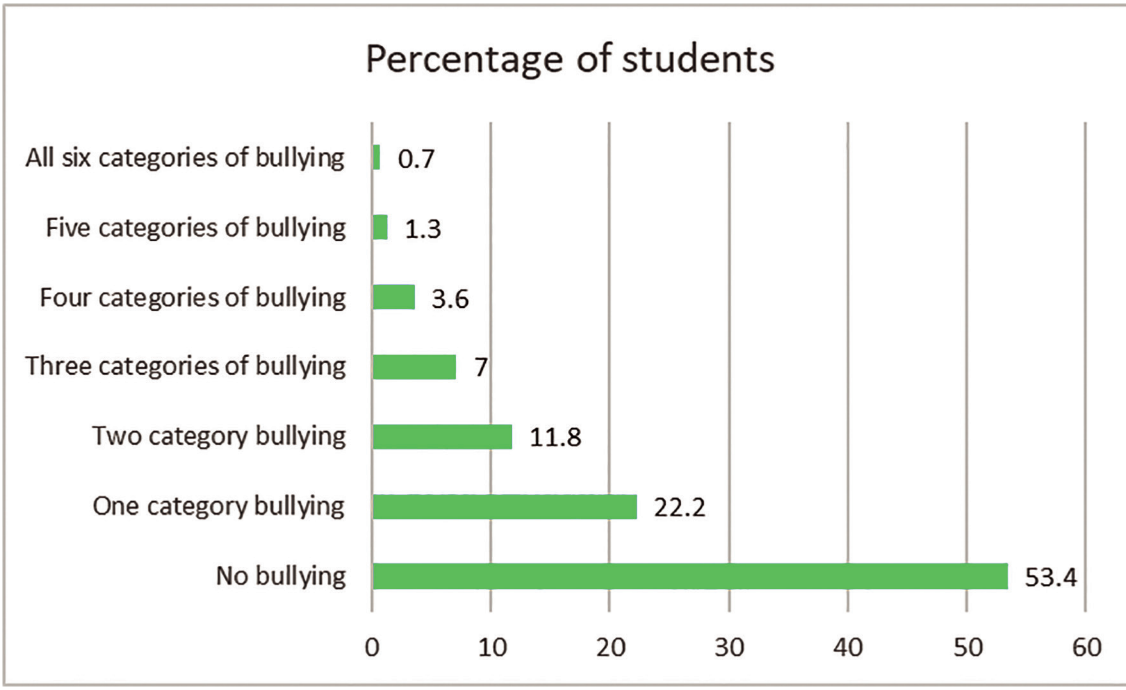
There were six categories of bullying that students were asked whether they had faced in the last month. The responses of students are summarised in Table 43.

Table 43 Existence of bullying behaviour in the school

Type of bullying	Number of students (%)	
	No	Yes
1. Some of my belongings were stolen	81.6	18.4
2. I was beaten by friends	87.8	12.2
3. Friends have misbehaved towards me	89.3	10.7
4. Friends have teased me	82.1	17.9
5. I was isolated	90.7	9.3
6. I was forced to do unwanted things	76.5	23.5

Table 43 shows that after analysing the data by summing up all six categories of bullying into a variable, the data obtained is as in the graph in Figure 34.

Figure 34 Presentation of existing bullying activities in school (%)



The chart in Figure 34 reveals that about 47% of students have experienced at least one type of bullying, showing that schools are not yet operating good discipline. However, the correlation between the sum of bullying and the student score, r = 0.04, revealed that there is almost zero correlation between student score and bullying.

Status of availability of scholarship in schools

In NASA 2020, students were asked 'Did you get a scholarship?' The number of students who did and did not receive a scholarship is presented in Table 44.

Table 44 Status of availability of scholarships in schools

Type of school	Scholarship	N	% of students
Community	Scholarship not received	268683	73.3
	Scholarship received	51847	14.1
	Not stated	46072	12.6
	Total	366602	100
Institutional	Scholarship not received	41180	49.7
	Scholarship received	11304	13.6
	Not stated	30428	36.7
	Total	82911	100

Table 44 shows that only 14.1% of students from community schools and 13.6% from institutional schools are getting scholarships from the institutions. Hence, relevant stakeholders should manage additional scholarship programmes in the institutions.

A comparison of achievement by scholarship received vs not received is presented figure 35.

Figure 35 Achievement in mathematics by scholarship received in community and institutional schools

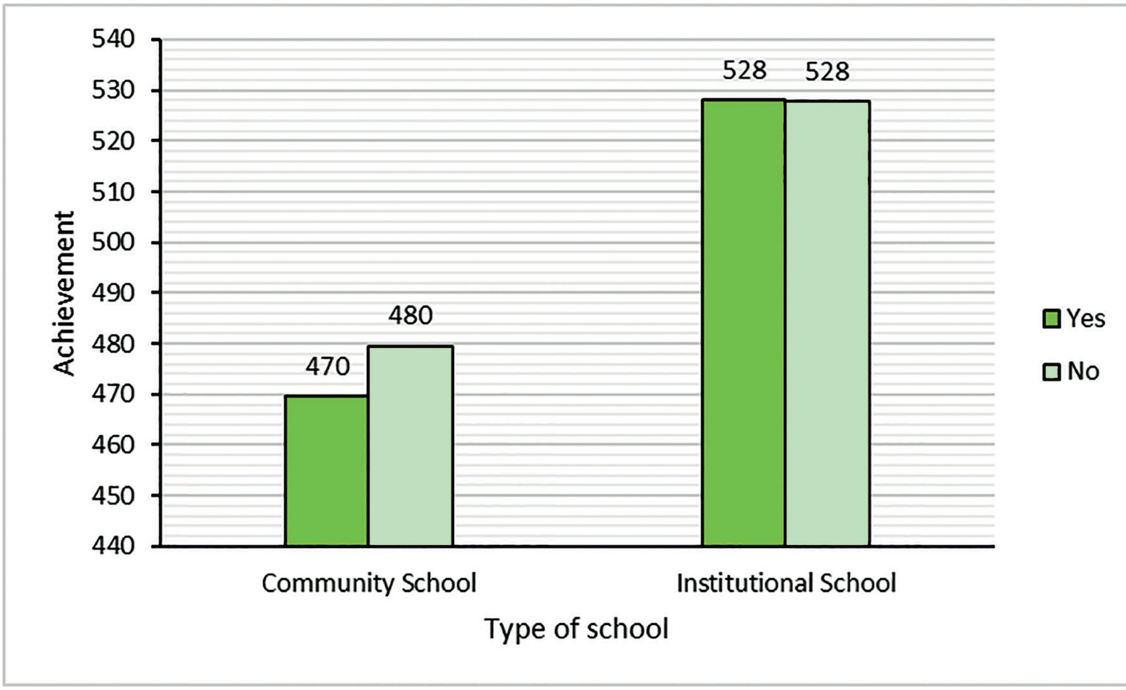


Figure 35 depicts that students in institutional schools who responded that they have received scholarships achieved 528 scale score, whereas community school students who responded that they did not get a scholarship achieved a score of 480. Those who received scholarship is community school have performed lower than who did not receive.

Sufficiency of scholarship

One of the questions asked to students was 'Is the scholarship sufficient for exercise book and pen purchase?' The students' responses are presented in Table 45.

Table 45 Response of students on sufficiency of scholarship

School type	Scholarship	N	%
Community	Sufficient	268683	73.3
	Not sufficient	51847	14.1
	Not stated	46072	12.6
	Total	366602	100
Institutional	Sufficient	41180	49.7
	Not sufficient	11304	13.6
	Not stated	30428	36.7
	Total	82911	100

Table 45 shows that around three-quarters (73.3%) of students from community schools responded that the amount of scholarship is sufficient for purchasing copy and pens, whereas only around half (49.7%) of students from institutional schools have the same response. Furthermore, 14.1% from community schools and 13.6% from institutional schools reported that the scholarship amount is insufficient for purchasing copy and pens.

Frequency of use of teaching materials

Students were asked whether their teacher uses any teaching materials on a regular basis or not. Their responses are presented in Table 46.

Table 46 Students' attitudes towards use of teaching materials

School Type		Number of students	Per cent
Community	Always uses	141335	38.6
	Uses 3–4 times a week	78635	21.4
	Uses 1–2 times a week	70641	19.3
	Never uses	35767	9.8
	Not stated	40225	11
	Total	366602	100
Institutional	Always uses	30429	36.7
	Uses 3–4 times a week	17623	21.3
	Uses 1–2 times a week	20525	24.8
	Never uses	9576	11.5
	Not stated	4758	5.7
	Total	82911	100

Table 46 shows that 38.6% of teachers from community schools and 36.7% from institutional schools are always using instructional materials in their teaching. However, teachers from 9.8% of community schools and 11.5% of institutional schools are in the habit of never using such materials in their classroom teaching.

Frequency of use of laboratory

One of the questions students were asked was 'How often does your teacher use the mathematics laboratory?' The responses of the students are presented in Table 47.

Table 47 Students' attitudes towards frequency of use of laboratory

Type of school		Frequency	Per cent
Community	Always uses	106332	29
	Uses 3–4 times a week	66937	18.3
	Uses 1–2 times a week	68330	18.6
	Never uses	73355	20
	Not stated	51648	14.1
	Total	366602	100
Institutional	Always uses	14943	18
	Uses 3–4 times a week	13509	16.3
	Uses 1–2 times a week	19501	23.5
	Never uses	27326	33
	Not stated	7632	9.2
	Total	82911	100

In Table 47, it can be seen that teachers in around one-third (29%) of community schools and less than one-fifth (18%) of institutional schools always use the mathematics laboratory. However, one-fifth (20%) of community school teachers and one-third (33%) of institutional school teachers never use the laboratory in the school.

Achievement by laboratory use

Students' mean achievement based on the use of the laboratory was analysed. The details of the results are presented in Table 48.

Table 48 Achievement by laboratory use

Type of school	How often does your teacher use the laboratory ?	Mean	N	Std. deviation	Std. error of mean
Community	Always uses	465.7865	106332	37.24817	0.11423
	Uses 3–4 times a week	474.6181	66937	37.98307	0.14681
	Uses 1–2 times a week	484.0856	68330	39.14708	0.14976
	Never uses	486.8037	73355	42.32075	0.15626
	Not stated	457.2476	51648	42.85198	0.18856
	Total	473.8122	366602	40.98102	0.06768

Institutional	Always uses	513.7444	14943	39.41032	0.32239
	Uses 3–4 times a week	522.6211	13509	39.06063	0.33606
	Uses 1–2 times a week	533.5839	19501	39.24155	0.28101
	Never uses	538.1553	27326	38.2634	0.23147
	Not stated	511.8046	7632	40.32006	0.46154
	Total	527.7238	82911	40.36773	0.14019

Table 48 shows that those teachers who never use a mathematics laboratory have the highest learning achievement in both community (mean = 486.09) and institutional (mean = 538.16) schools. However, the score was found to be poorest with respect to those teachers always using a mathematics laboratory in both community (mean = 465.79) and institutional (mean = 513.74) schools. The data shows use of the laboratory in Mathematics do not have a positive influence on learning.

Achievement by availability of textbook

In NASA 2020, one of the questions students were asked was 'Do you have a mathematics textbook?' Out of all the students, 7.2% reported that they did not have a textbook for mathematics. The responses have been presented in Table 49.

Table 49 Achievement by availability of textbook

Textbook	Mean	N	Std. Deviation	Std. Error of Mean
No	457.6821	33354	37.48862	0.20527
Yes	486.3512	419545	45.07028	0.06958
Missing	443.6113	9769	60.80368	0.61518
Total	483.382	462668	45.9388	0.06754

Table 49 shows that those students not having a textbook (mean = 457.68) perform worse than those with a textbook (mean = 486.35), who have good performance in mathematics. Hence, government and other stakeholders should ensure these resources are available to the students.

Achievement by homework

This study analysed whether homework plays an important role in learning achievement. The responses of the students were categorised and are presented in Table 50.

Table 50 Achievement by homework provided or not

bqN2homework	Mean	Std. deviation	Std. error of mean	N	% of total N
Always	486.256	44.84796	0.07156	392805	84.90%
Sometimes	474.2013	42.73806	0.18086	55840	12.10%
Never	450.7477	53.86993	0.82828	4230	0.90%
Not stated 99	434.5489	60.85374	0.61492	9794	2.10%
Total	483.382	45.9388	0.06754	462668	100.00%

Table 50 shows that students who always get homework have the highest average achievement in mathematics (mean = 486.26, SD = 44.85), whereas the result was found to be poor for those never getting homework (mean = 450.75, SD = 53.87) from their teachers. The results also suggest that regularity in homework is necessary for improving mathematics achievement.

Achievement by feedback on homework

An important question students were asked as part of their feedback on homework was 'Does your teacher give feedback on homework?' Students' responses are presented in Table 51.

Table 51 Achievement by feedback on homework

bqN3feedback	Mean	Std. deviation	Std. error of mean	N	% of total N
Always	484.8572	44.58223	0.07668	338002	73.10%
Sometimes	485.3213	45.00751	0.1381	106209	23.00%
Never	457.553	52.22912	0.66465	6175	1.30%
Not stated	438.9989	59.2852	0.53495	12282	2.70%
Total	483.382	45.9388	0.06754	462668	100.00%

Table 51 shows that the average achievement score in mathematics was found to be high among those students getting feedback either some of the time (mean = 485.32) or always (mean = 484.86). However, the result is poor for those who never get feedback (mean = 457.55). Hence, every teacher should give regular feedback on their students' homework.

Achievement by Teachers' regularity in class

Teachers' regularity was also questioned and analysed in NASA 2020. The details are shown in Table 52.

Table 52 Achievement by teachers' regularity in class

Teacher's regularity in class	Mean	Std. deviation	Std. error of mean	N	% of total N
Spends whole time	486.1381	44.71362	0.07077	399166	86.30%
Comes late and goes quickly	480.2301	47.49449	0.2898	26858	5.80%
Generally, does not come to the class	464.7573	40.7844	0.26929	22937	5.00%
Not stated	440.4605	56.08271	0.47904	13706	3.00%
Total	483.382	45.9388	0.06754	462668	100.00%

Table 52 shows that students' achievement scores in mathematics are found to be significantly higher where teachers spend the whole time in the classroom (mean = 486.14, SD = 44.71), and the results are significantly poorer where teachers do not come to class (mean = 464.76, SD = 40.78). Hence, teachers reliably spending the whole time in the classroom has a positive impact in mathematics achievement.

Students' self-concept of mathematics

To measure the self-confidence of students, they were asked a question about the usefulness of mathematics, and their responses were transformed from negative to positive by recoding: 1 = 4, 2 = 3, 3 = 2, 4 = 1. The frequency of the transformed responses is presented in Table 53.

Table 53 Proportion of responses on self-concept of usefulness of mathematics

Statements	Not stated	Totally disagree	Partially disagree	Partially agree	Totally agree
1. Mathematics learning helps me to do household work.	4.4	1.5	1.5	12.4	80.2
2. Mathematics helps me to learn other subjects.	5.9	2.5	3.3	24	64.4
3. I like to practise mathematics.	5.4	1.4	2	14.3	76.8
4. I have to do well in mathematics to learn other subjects.	5.6	1.7	1.8	11.7	79.3

Table 53 shows that just over four-fifths of students totally agree that learning mathematics helps them to do household work (80.2%), and it helps for learning other subjects (64.4%). Furthermore, more than three-quarters of students like to practise mathematics (76.8%) and believe they have to do well in mathematics to learn other subjects (79.3%), indicating that a high level of competency in mathematics supports and enhances the learning of other subjects.

Self-confidence in mathematics

The students' responses are transformed from negative to positive by recoding: 1 = 4, 2 = 3, 3 = 2, 4 = 1. The frequency of the transformed responses is presented in Table 54.

Table 54 Self-confidence in mathematics (%)

Self-confidence	Not stated	Totally disagree	Partially disagree	Partially agree	Totally agree
Usually I do well in mathematics.	5.6	1.5	3.3	30.3	59.3
I want to learn mathematics more.	5.4	1.2	1.5	11.1	80.8
I enjoy learning mathematics.	6.5	1.5	2.2	16.7	73
I can learn mathematics fast.	7.9	3.8	6.3	40.5	41.5
I find mathematics difficult.	8.5	19.4	12.6	31.9	27.6

Table 54 shows that just over four-fifths of students are interested to learn mathematics (80.8%), around two-thirds (59.3%) of students are doing well in mathematics and around three-quarters (73%) of students enjoy learning mathematics. However, only two-fifths (41.5%) of students pick up mathematics quickly, and around two-thirds (59.5%) of students still find mathematics a difficult subject. Hence, it should be improved through the use of learner-centred instruction and by use of sufficient learning resources and technologies.

Attitudes towards teaching–learning activities

Students were asked to express their attitudes towards the teaching–learning activities. Their responses were categorised as presented in Table 55.

Table 55 Attitudes towards teaching–learning activities

Statements	In most of the lessons	In some lessons	Never	Not stated
Teaching–learning: We do exercises from the textbook.	84.6	10.2	0.6	4.7
Teaching–learning: We work in groups with peers.	66.6	25.9	2.2	5.3
Teaching–learning: We solve mathematics problems ourselves.	70.7	21.2	1.7	6.4
Teaching–learning: We use geometry box materials in mathematics.	65.6	23.2	4	7.2
We start to do homework in the classroom.	31.4	27.5	26.4	14.7

Table 55 shows that most students do practice in mathematics from their textbook (84.6%), which may mean that the learners do not have access to or the idea of using other resources. Two-thirds (66.66%) of students work with their peers, which is a good concept for collaborative learning. Hence, this should be encouraged. The majority of students are in the habit of solving problems themselves (70.7%), which promotes students' problem-solving skills. Additionally, some students (4%) are not using geometry box materials in the lessons. Hence, relevant stakeholders should focus on these arguments, and if learners do not have access to such resources, this should be managed by special support. Around a quarter (26.4%) of students never start their homework in the classroom. Hence, they may need some special counselling because collaborating with their peers in the classroom can help if they face challenges in mathematics.

Achievement by type of schools

Students' reactions regarding the school type and their achievement were categorised and are presented in Figure 36.

Figure 36 Achievement by type of school

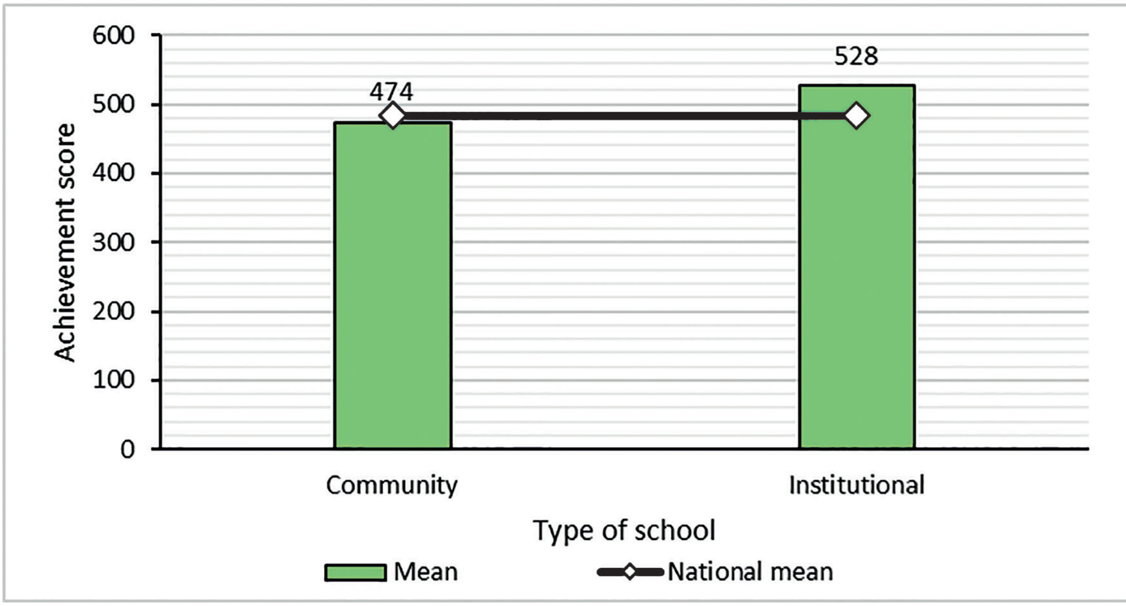


Figure 36 shows that the difference is stasistically significant; at 99% confidence level, the $p < 0.0001$. The partial correlation coefficient ($r = 0.455$) reveals that the effect of the type of school is high.

Science Results

Introduction

In this section, the results of the responses of 21,146 students from 75 districts and 900 schools, who participated in NASA 2020 in science, are analysed. The results are presented in the form of proficiency levels, their description and comparison. Population estimates presented in this chapter are based on the ten plausible values drawn from WLE. The comparisons are made on the basis of groups formed from background information variables such as students' family background, socio-economic status, ethnicity, gender, home language, school types, home environment, province, etc.

In this assessment, some of the test items used in the 2017 assessment were also used as anchor items. All the 2020 items have been calibrated by fixing the 2017 parameters of those anchor items. The 2020 results obtained in this way are comparable to the 2017 results. IRT has been used to analyse this, with 2PL used to analyse items of 1 point and GPCM used for questions with 2 or more points. TAM was used to analyse this result.

The latent ability (theta) of each student is obtained from the analysis stated above.

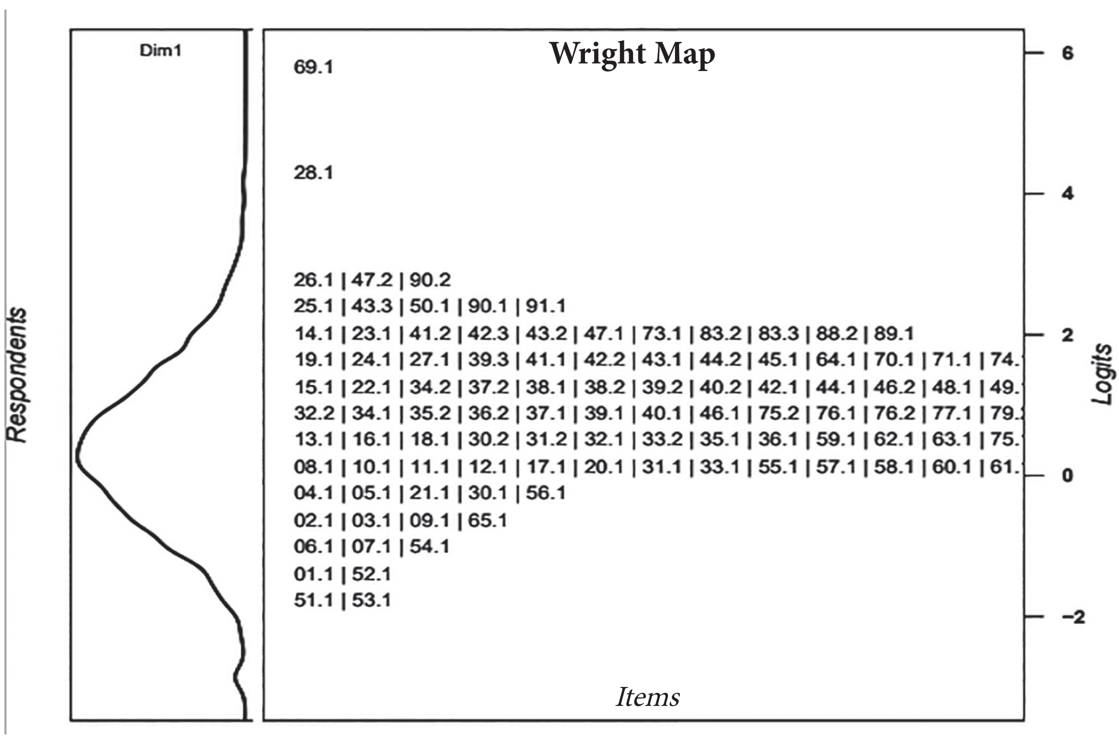
Ten plausible values were calculated to find out the ability of the population on the basis of the latent ability, and the dummy variables made on the basis of the variables asked in the students' background questionnaire. After transforming the average of those plausible values to zero (0), the following formula was used to present the results:

Student Achievement Score = 500 + PV × 50

The Wright map for science

The Wright map is organised as a frequency curve and vertical histogram. The frequency curve on the left side shows the respondents and the histogram on the right side shows the test items. The left side of the map shows the distribution of the measured ability of the respondents from most able at the top to least able at the bottom. The items on the right side of the map are distributed from the most difficult at the top to the least difficult at the bottom. In Figure 37, student ability (θ) in the left and NASA 2020 items to the right are plotted on the same scale. When a person and an item lie at the same level, the probability of that response by the particular respondent is 50%. Figure 37 presents the NASA 2020 Wright map for science.

Figure 37 Wright map showing respondent and item on the same scale



The curve on the left side represents the number of students; their latent ability is displayed in the logit scale ranging from -4 to +6. The distribution of students against the items administered (item numbers are shown to the right side) reveals that most of the items were difficult for the students. Although items were pre-tested and based on the grade 8 curriculum, most of the students were below the average latent ability '0'. This indicated that test items were difficult for the participating students. This further depicts that the performance level of grade 8 students in science was not as expected by the curriculum.

National mean in science

This comparison is made based on the linking items administered in both years. While comparing the national mean, NASA 2017 linking item parameters were fixed. Based on those fixed item parameters, NASA 2020 items were calibrated. The national mean score in NASA 2020 is 470, in comparison with 500 in NASA 2017.

Figure 38 Comparison of national means of NASA 2020 and NASA 2017

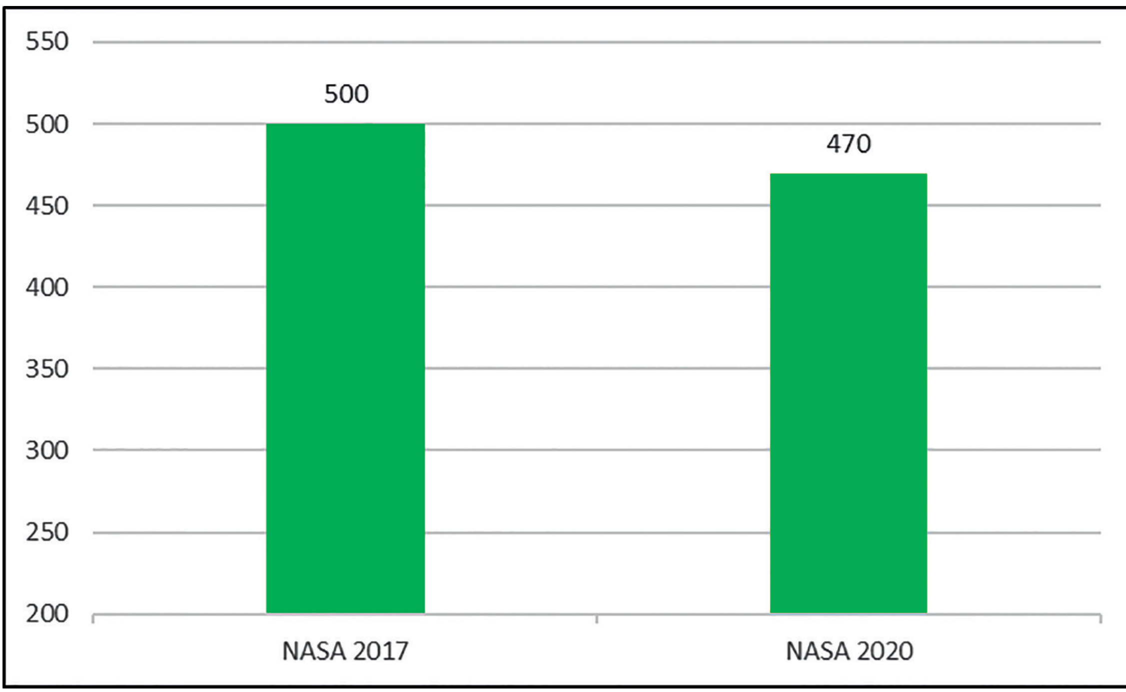


Figure 38 shows the comparison of the mean between NASA 2017 and NASA 2020. The mean score of 2017 was 500, whereas in 2020 it was 470. The trend reveals a decrease during these three years, with the mean score falling by 30 scale score. This shows that learning in the classroom setting in Science has deteriorated during those years. As a result, the mean score has noticeably decreased.

Proficiency level in science defined by cut scores

The NASA 2020 Framework has defined six levels of proficiency. According to the framework, the lowest level was level 1 (Below basic level) and the highest level was Advanced level. The scale score ranges set in NASA 2017 were used in NASA 2020 to make the levels comparable. The proficiency level and scale score range are shown in Table 56.

Table 56 Proficiency levels and scale score range

Proficiency level	Score range
Level 6: Advanced level	575 and above
Level 5: Proficient 3 level	529–575
Level 4: Proficient 2 level	482–529
Level 3: Proficient 1 level	438–482
Level 2: Basic level	390–436
Level 1: Below basic level	below 390

Table 56 shows that Basic level covers 390–436 scale score and Proficient level 1 represents 438–482, while Proficient level 2 denotes 482–529 scale scores as fixed parameters in the scale score. Similarly, Proficient level 3 refers to 529–575, while Advanced level covers scale scores above 575. These benchmarks were set in 2017, and the same benchmarks are used for consistency in comparison.

Proficiency level-wise distribution of the students

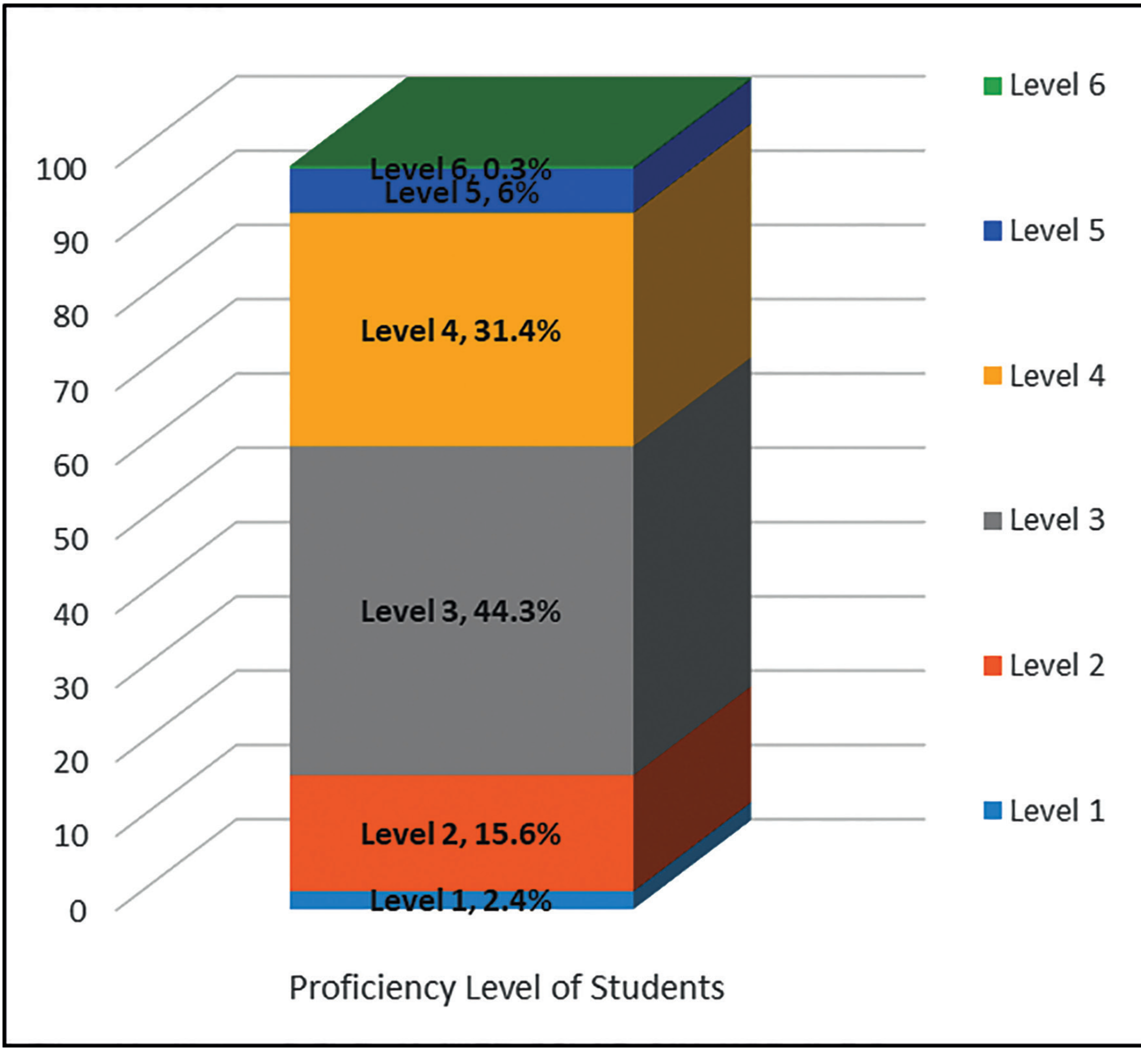
The test scores were also analysed in terms of the six proficiency levels of students' achievement. The benchmark consists of six levels in low (Below basic) to high (Advanced) order. Scores below 390 were considered Below basic level, while scores within the range of 390–436 were categorised as Basic level. Similarly, scores within the ranges of 438–482and 482–529 were labelled Proficient level 1 and Proficient level 2 respectively. Scores within the range of 529–575 and above 575 were categorised as Proficient level 5 and Advanced level 6 respectively.

Table 57 Distribution of students in different levels of proficiency

Proficiency level	Score range	N	Percentage of students
Level 6: Advanced level	575 and above	1124	0.3
Level 5: Proficient 3 level	529–575	26783	6
Level 4: Proficient 2 level	482–529	140883	31.4
Level 3: Proficient 1 level	438–482	198954	44.3
Level 2: Basic level	390–436	70239	15.6
Level 1: Below basic level	390 and below	10860	2.4
Total		448844	100

Table 57 also shows that the total number of students participating in the test was 448,844. The data reveals that 10,860 students (i.e. 2.4%) were below the basic level, whereas only 1,124 (0.3%) were at an advanced level. Most students fall under Proficient level 1 and Proficient level 2, which comprised 198,954 (44.3%) and 140,883 (31.4%) respectively, and only 26,783 (6%) of respondents were categorised as Proficient level 3.The distribution of students across the six proficiency levels is represented in Figure 39.

Figure 39 Distribution of students across the six levels of proficiency



The data in Figure 39 reveals that if students in level 4 or above are taken as achieving the minimum level of proficiency, the greatest proportion of students (62.3%) fall below this, with only 37.7% of students having shown their performance to be above the minimum level of performance. Altogether, 18% students are at below the basic level and only 6.3% student's fall on or above proficient level 3. The result shows that approximately two-thirds of grade 8 students were deprived of learning opportunities in science.

Overall mean score by province

The weighted mean score stratified by the seven provinces is presented in Figure 40. The orange horizontal line represents the national mean in 2020.

Figure 40 Overall mean score in science by province

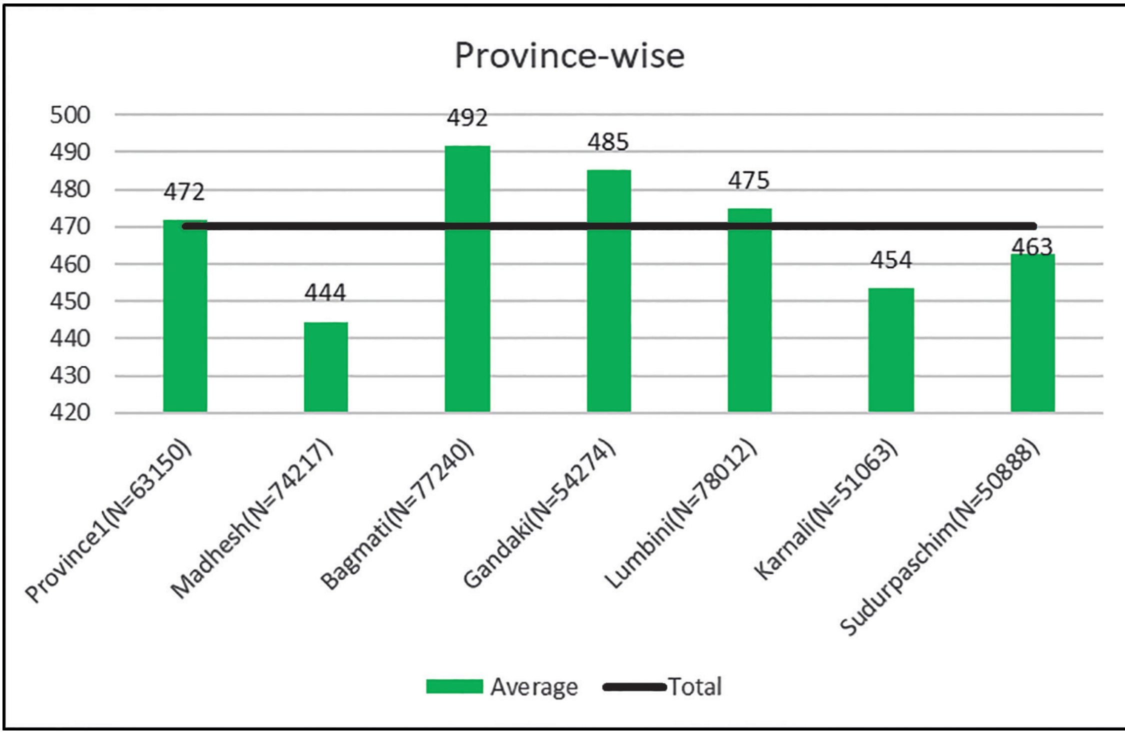


Figure 40 shows a comparison of the overall mean scores for the seven provinces. The orange horizontal line represents the national average (470) of all provinces. Madhesh province has the lowest achievement (444), which is below the national average, and Bagmati province has the highest mean at 492 scale score. Karnali and Sudurpaschim provinces scored slightly above the mean score of Madhesh, at 454 and 463 respectively. Lumbini and Gandaki provinces also scored slightly above the national mean, at 475 and 485 respectively. Province 1 maintained the national average with a score of 472. Overall results show that Province 1 and Lumbini maintained the national average, Bagmati and Gandaki scored quite a bit above the national average, while students in Madhesh, Karnali and Sudurpaschim scored below the national average in science in NASA 2020.

Province-wise student performance level

The Federal Democratic Republic of Nepal is divided into seven provinces and 753 local government units. While picking up the schools as PSUs, provinces were regarded as strata. The average scores described in this section are the transformed/scale score of 470 national average. The national mean is taken as a reference to contrast with the provincial mean. Those provinces whose average score exceeds the mean score are acknowledged as better-performing provinces, whereas those with an average score below 469 are presumed to be of substandard performance. The details are presented in Figure 41.

Figure 41 Distribution of students in different levels of proficiency by province

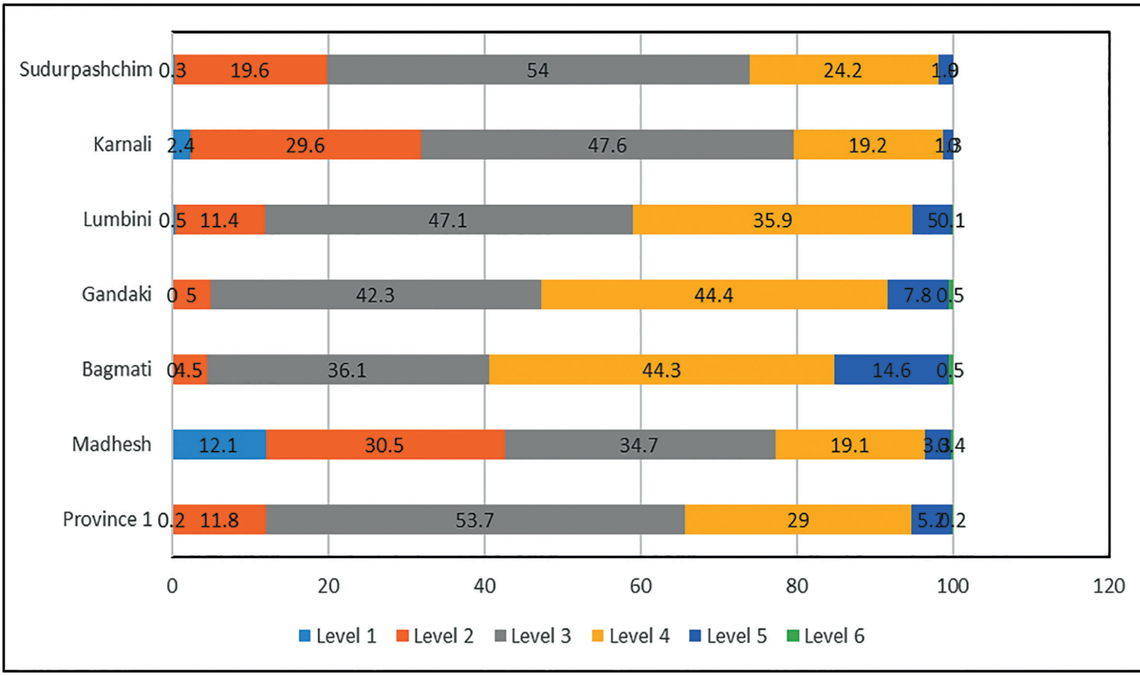


Figure 41 presents the distribution of students' performance levels in the seven provinces across levels 1–6. The data reveals that Province 1 and Sudurpaschim have the highest number of students in the lowest proficiency level among all provinces. This indicates that the grade 8 science curriculum was least effectively delivered or learned in these areas. Gandaki and Bagmati provinces have the lowest number of students at Below basic level and the highest number of students at Advanced proficiency level. This reveals that students in these two provinces have learned the content of the science curriculum the best. Overall, the results indicate that Bagmati students have the highest mean score (59.4%), followed by Gandaki (52.7%), while students in Karnali have the lowest mean score (20.5%), followed by Madhesh (23.3%). This means that these students have passed the minimum level of proficiency in NASA 2020.

Distribution of students by type of local level

Nepal's political sector is divided into local levels such as Metropolitan City, Sub-metropolitan city, municipality or rural municipality. In this study, when categorising school location into rural and urban areas, rural municipality can be considered as a rural location and others as urban locations. Based on these two categories, data showing the difference in students' distribution over the six proficiency levels in science is presented in Figure 42.

Figure 42 Distribution of students in different proficiency levels by local level (locations) for science

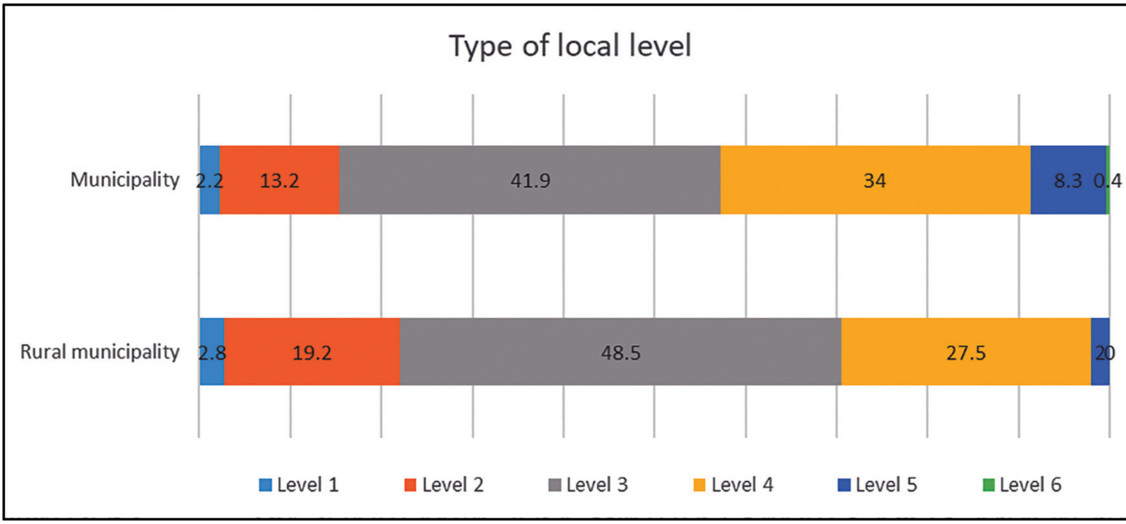


Figure 42 clearly presents evidence that the distribution of students varies between urban and rural areas in all proficiency levels. In rural areas, there were more students at Below basic level (19.2%) than in urban areas (13.2%). In level 3, there were also more rural students (48.5%) than urban students (41%), which clearly shows that rural students achieve less than urban students. Regarding the top three proficiency levels, more urban students (34%) have achieved higher proficiency levels than rural students (27.5%). Overall, the results indicate that 42.4% have passed the minimum level of proficiency in urban areas, whereas only 29.5% students have passed the minimum level of proficiency in rural areas. This means that science teaching has been less effective in rural areas, where only a third of students have passed the minimum level of proficiency.

Student distribution by gender in science

In the NASA 2020 questionnaire framework, students stated their gender. Based on the gender of the respondents, students were categorised into three section – boys, girls and not specified.Girls and boys should have equal opportunity and support in their studies; however, the results vary between them. Figure 43 shows mean score achieved by Science students in Grade 8 based on gender.

Figure 43 Achievement by gender in science

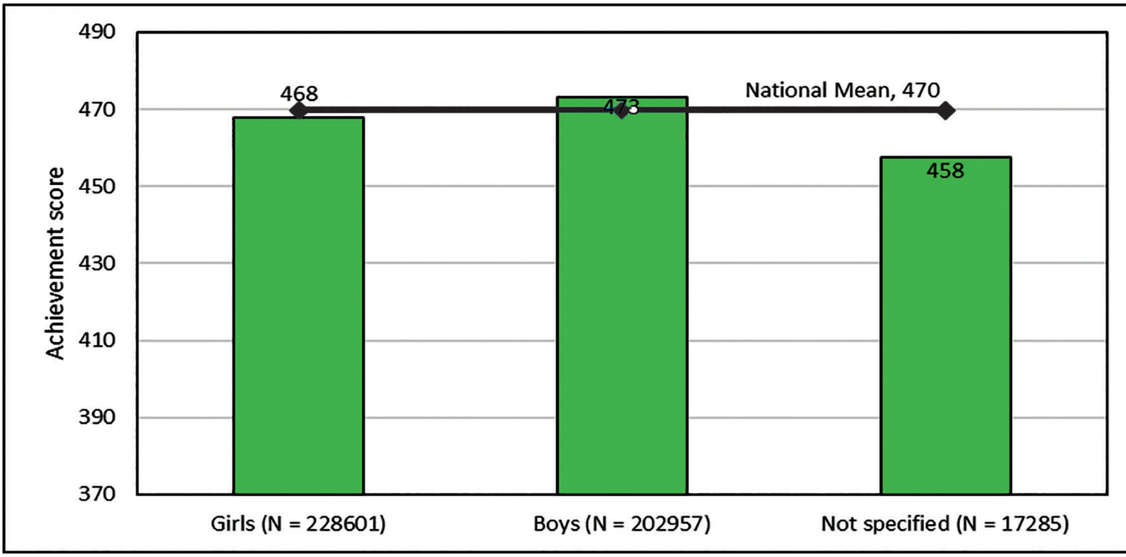


Figure 43 shows that of the total number of students, 228,601 were girls, whose mean score of 468 was below the national mean, and the total number of boys was 202,957, whose mean score was 473, slightly above the national average.

Girls and boys should have equal opportunity and support in their studies, but results vary as to the extent this has been realised among grade 8 science students. With gender as an implicit stratum, a comparison was made of the distribution of students across the six defined proficiency levels. The results are presented in Figure 44.

Figure 44 Distribution of students in science across different levels of proficiency by gender

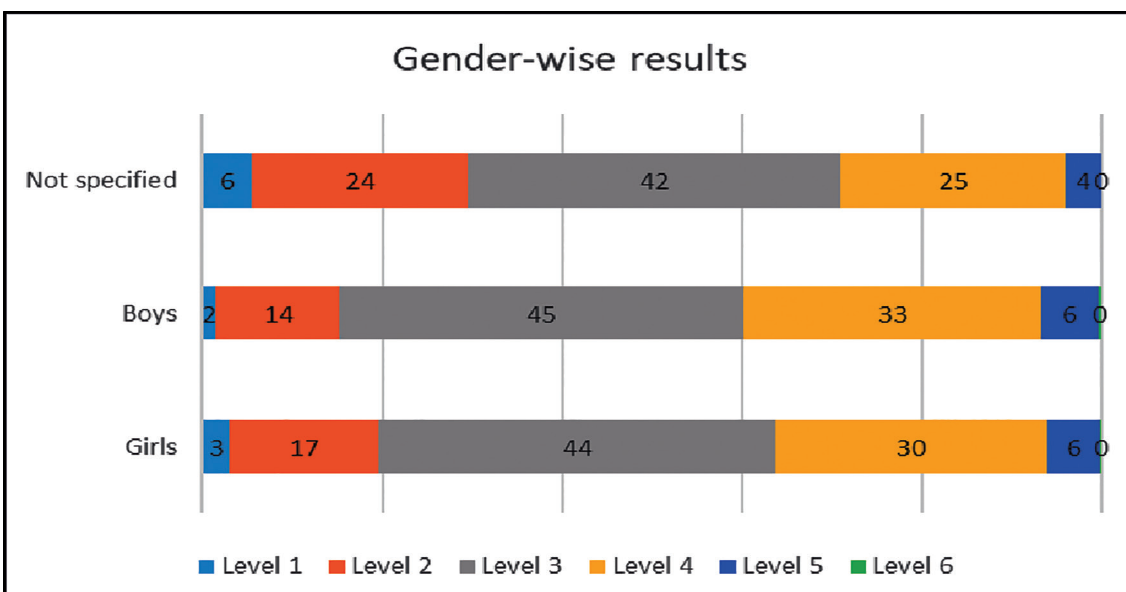


Figure 44 shows that the variation between boys and girls in their achievement. Approximately 39% of boys passed the minimum proficiency level, whereas 36% of girls passed the minimum level proficiency, so slightly less than boys. Of the students who did not state their gender, only 29% passed the minimum level of proficiency in NASA 2020. These results indicate a slightly bigger distribution of boys than girls in the higher proficiency levels in terms of achievement.

Achievement by home support for studying

Support at home for students' studies plays an important role in increasing learning achievement. The variables in NASA 2022 were categorised as father, mother, tuition, siblings, friends, grandparents, others and nobody. Table 58 provides the details.

Table 58 Achievement by home support for studying

Support beyond school time	Mean	N	Std. deviation	Std. error of mean
None	480.9026	18166	36.76442	0.27277
Tuition	478.3082	39826	36.15528	0.18117
Others	475.5917	11188	39.09831	0.36964
Mother	470.9765	28966	37.81444	0.22219
Siblings	469.8938	205286	35.55487	0.07847
Friends	466.3774	29758	36.67002	0.21257
Father	465.5138	69213	39.74046	0.15106
Not reported	464.0068	44199	57.30245	0.27256
Grandfather and grandmother	460.3448	2242	37.43056	0.79048
Total	469.7619	448844	39.52154	0.05899

Table 58 illustrates the number of students who were supported by different people at home beyond school time. The data shows that the largest number of students were supported by their siblings (205,286), father (69,213) and mother (28,966) at home. Their achievement levels scored 470, 465 and 470 respectively. The results indicate that the support from the father (465) is less effective than the support provided by the mother (471), and the least-effective support is grandparents (460), as the students achieved below the national mean. Learning supported by tuition and Other were better, as they scored 478 and 475 respectively. However, good students have no need of any support, as a score of 480 was achieved by students who received no support at home. The details are also displayed in Figure 45.

Figure 45 Achievement by home support for studying

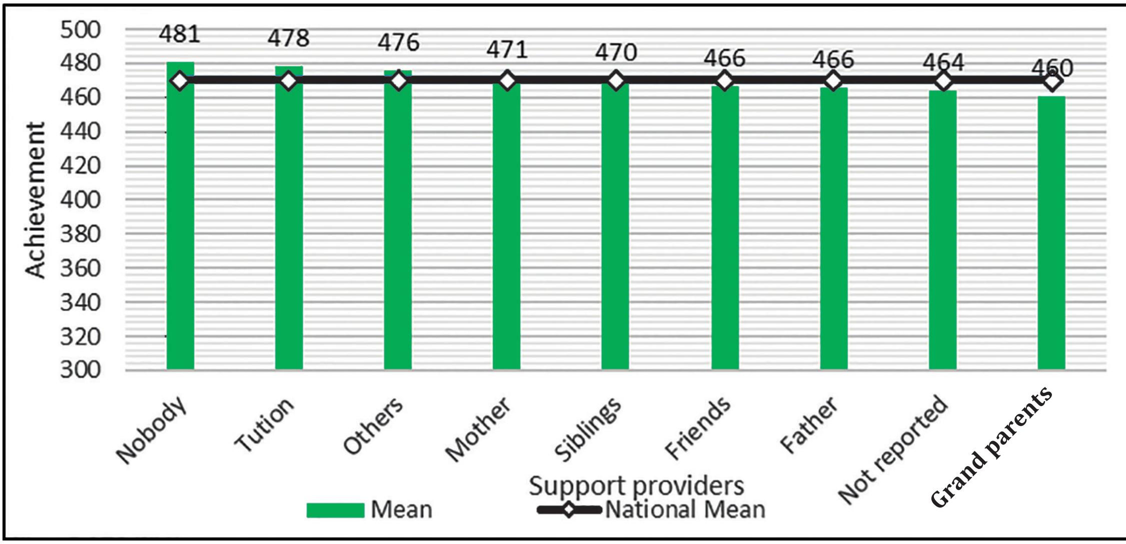


Figure 45 reveals that students who were not supported by anyone achieved the highest score (481), whereas students supported by grandparents scored the lowest (460). Support from siblings, mother, other and tuition are also noteworthy, as the students achieved higher than or equal to the national average. The mean score indicates that students who were supported by their mother achieved higher (471) – a bit above the national average – whereas students who were supported by their father achieved (465), below the national average.

Overall result show that after school teaching, home support plays a vital role in better learning achievement, but a few exceptionally capable students do not need any support or they can find the support themselves and get better scores.

Achievement by career aspirations

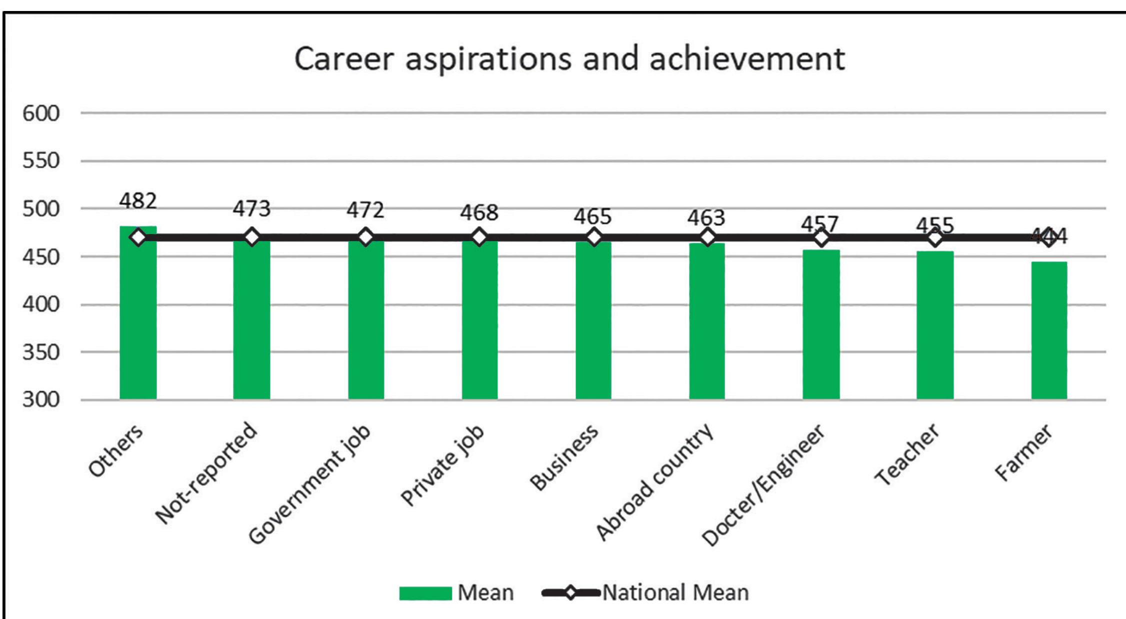
Career aspirations can be a great motivating force for students and can affect their achievement. Acknowledging this fact, students were asked about their aim for the future regarding different types of professions. The professions included employment in the private sector, farming, working abroad, businessman, doctor/engineer, civil servant and teacher, and all other professions were categorised into the heading 'Other'. Table 59 depicts the achievement of students in accordance with their career aspirations.

Table 59 Achievement by future career goal

Your future aim	Mean	N	Std. deviation	Std. error of mean
Other	481.6239	135719	37.34814	0.10138
Not-specified	473.2792	80081	45.46795	0.16067
Government job	471.9411	67035	35.99492	0.13902
Private job	468.1583	7009	34.50128	0.41209
Business	464.7095	19207	36.28735	0.26184
Work abroad	463.2143	14595	35.8114	0.29642
Doctor/Engineer	456.66	11515	36.81936	0.34312
Teacher	454.9782	113475	35.4454	0.10522
Farmer	443.8368	207	42.74015	2.97176
Total	469.7619	448844	39.52154	0.05899

Table 59 shows the number of students who considered their possible career aspirations in relation to their achievement in science. The data showed that out of 448,844 students, 113,475 of them hope to be a teacher, 11,515 a doctor or engineer, 67,035 hope to work in government sectors, 19,207 in business, 14,595 want to work abroad, 7,009 in the private sector, 80,081 did not specify, 135,719 were aiming for other professions and, last of all, 207 saw their future career in farming. Figure 46 shows their mean score in relation to their future goal.

Figure 46 Achievement by career aspirations



The data in Figure 46 illustrates that students who want to work in a government job or 'other' job category, or who did not specify, achieved above the national average, that is, 472, 473 and 482 respectively. The mean score for private sector jobs, business, working abroad, doctor/engineer, teacher and farmer was below the national average. Overall, results show that no particular job categories have especially positive effects on students' achievements, as those students who did not report a job category or specified 'other' have achieved higher mean scores than those wishing to work as a doctor/engineer, teacher, in business or working abroad. This difference in mean score by career goal is significant at 99% confidence level ($p<0.001$).

Achievement by use of leisure time in the classroom

Leisure time allows students to devote their time to their interests and relieves them from the constant pressure of study. They can utilise their time effectively. This leisure time can also be of utmost importance in planning their studies and achieving more highly. The students were asked how they used their leisure time, and they selected from options such as returning home, playing, doing homework, and classwork. Table 60 shows the achievement score of students involved in different activities during their leisure time in school.

Table 60 Achievement by use of leisure time in the classroom

Leisure-time activities	Mean	N	Std. deviation	Std. error of mean
Classwork	472.537	250020	36.83699	0.07367
Homework	471.4015	142284	36.7648	0.09747
Play	466.681	29932	43.47919	0.25131
Return home	430.3461	10942	41.73934	0.39902
Not reported	443.9973	15666	63.84493	0.51009
Total	469.7619	448844	39.52154	0.05899

Table 60 shows that approximately 56% of students engage in classwork, and 32% of students do their homework during their school leisure time, and they achieved scores of 472 and 471 respectively. These scores are slightly above the national average, whereas students who engage in play, return home or who did not specify achieved 466, 430 and 443 respectively as their mean score. These results clearly show that students who are involved in their classwork or homework during their leisure time score higher than those students who are engaged in games or return home. Figure 47 presents the details of the impact of students' use of leisure time on their achievement.

Figure 47 Achievement by use of leisure time in the classroom

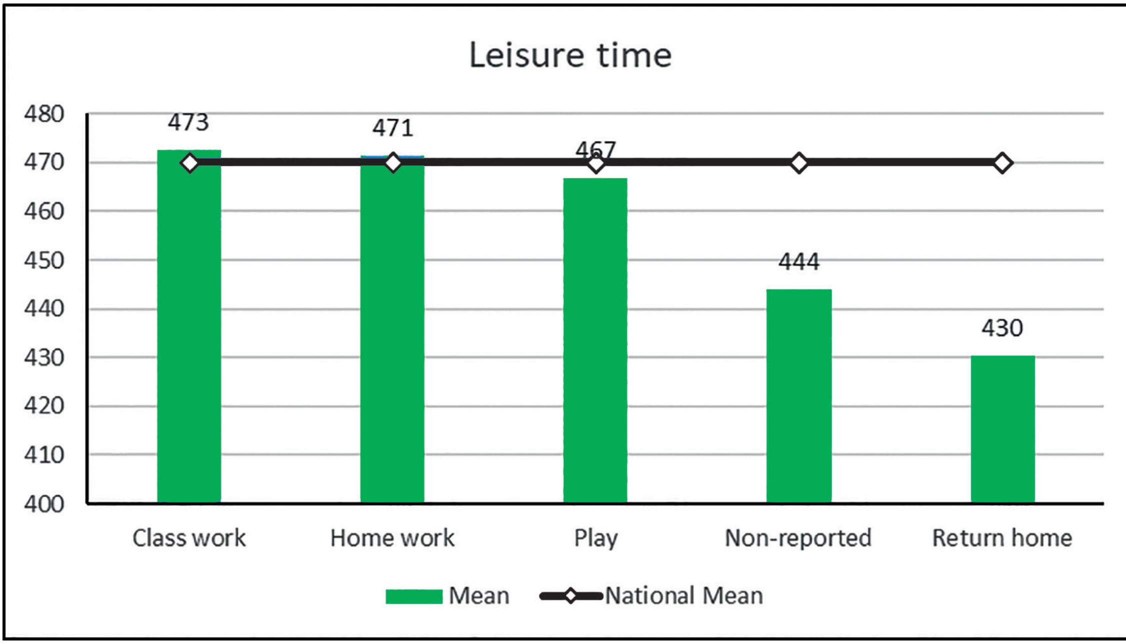


Figure 47 shows the achievement score of students involved in different activities during their leisure time at school. The students were using their leisure time in different activities such as returning home, playing, doing homework or doing classwork. The students engaged in classwork and homework achieved 473 and 471 respectively, which are above the national average. Figure 47 also reveals that the students who return home achieved 430, which is the lowest score. Similarly, students who did not respond and those who enjoy playing games and sports during their leisure time achieved below the national average. The results indicate that engaging students in classwork and homework during their leisure time has a positive impact in their learning achievement.

Achievement by frequency of extra-curricular activities

In NASA 2020, students were asked how often their schools offered extra-curricular activities, as learning achievement depends on students' physical and mental engagement in activities such as athletics, sports, voluntary work, photography, drama, music, etc. The involvement of students in various extra-curricular activities develops social and soft skills that promote their well-being. Studies also show that regular physical activities can facilitate stable, long-term, enhanced behavioural engagement with school in students (OECD, 2019), fostering social and emotional well-being (NCF, 2009). The frequency of extra-curricular activities in the school and students' involvement was analysed in comparison with their educational achievement in science. Table 61 shows the responses of the students on the frequency of extra-curricular activities in their school and their achievements.

Table 61 Achievement by frequency of extra-curricular activities

Extra-curricular activities	Mean	N	Std. deviation	Std. error of mean
Regular	470.3773	156548	40.84996	0.10324
Sometimes	472.9682	253957	35.8947	0.07123
Never	459.1344	21169	36.95867	0.25402
Not specified	429.8281	17169	54.65097	0.41708
Total	469.7619	448844	39.52154	0.05899

Table 61 depicts that 35% of students agreed that their school organised regular extra-curricular activities, 57% of them stated that their school engaged them in extra-curricular activities sometimes, and 5% of students responded that their school never involved them in extra-curricular activities. Details of their achievement in relation to the frequency of extra-curricular activities are also presented in Figure 48.

Figure 48 Achievement by frequency of extra-curricular activities

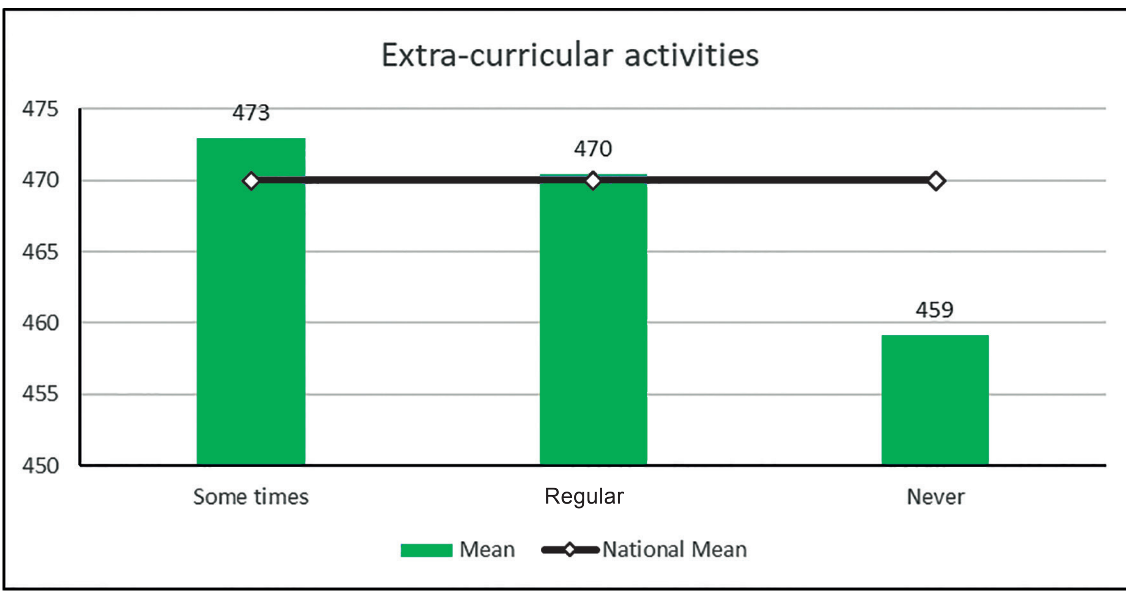


Figure 48 reveals that students who are sometimes engaged in extra-curricular activities achieved the highest score (473), which is above the national average, and students who were never involved in extra-curricular activities achieved the lowest score (459), which is below the national average. However, students who were engaged in regular extra-curricular activities scored 470, which is equal to the national average for science. The results indicate that less frequent extra-curricular activities are better than regular extra-curricular activities for students' achievement in science.

Achievement by mother's education

In NASA 2020, parents' educational background was also investigated. A mother's educational background can have a profound impact on the learning achievement of her children. Therefore, participants were asked in the questionnaire to state the academic qualification of their mother. The options were illiterate, literate, grade 8, grade 10, grade 12, Bachelor's, and Master's or above. Achievement in science based on mother's academic qualification is presented in Table 62.

Table 62 Achievement by mother's education

Mother's education level	Mean	N	Std. deviation	Std. error of mean
Master's or above	514.0954	4740	41.06691	0.59648
Bachelor's	501.4899	11934	43.66364	0.3997
Grade 12	494.1048	30722	38.28513	0.21843
Grade 10	480.9783	59619	38.71529	0.15856
Literate	470.6499	111883	32.89999	0.09836
Grade 8	468.5109	80399	36.67068	0.12933
Illiterate	459.9338	136067	35.13194	0.09524
Not reported	420.292	13480	58.03671	0.49987
Total	469.7619	448844	39.52154	0.05899

Table 62 shows that in NASA 2020, 30% of respondents stated that their mother was illiterate, and 25% stated that their mother was literate. The level of qualification of their mothers was 18% at Grade 8, 13% Grade 10, 7% Grade 12, 3% Bachelor's and 2% Master's or above. The students' achievement levels based on their mother's qualification are presented in Figure 49.

Figure 49 Achievement by mother's education

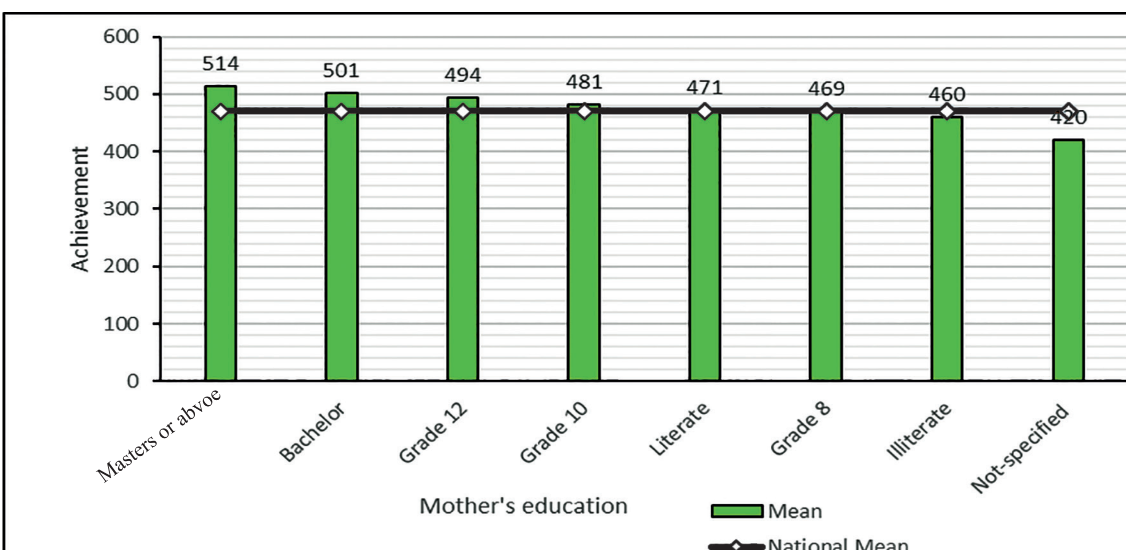


Figure 49 shows that there is a significant difference in the achievements of students whose mothers were literate and those whose mothers were illiterate. The achievement score for literate mothers' children was 471, whereas illiterate mothers' children achieved 460. The achievements of children whose mothers' educational levels were Grade 10, Grade 12, Bachelor's and Master's scored 481, 494, 501 and 514 respectively. The results indicate that the higher the mother's qualification level, the better their children's score.

Achievement by father's education

In NASA 2020, the father's education level also had a direct impact on the achievement of the students. The father's education was categorised in the same way as the mother's education level. The results are presented in Table 63.

Table 63 Achievement by father's education

Father's education status	Mean	N	Std. deviation	Std. error of mean
Master's or above	508.5525	12431	41.93918	0.37615
Bachelor's	494.8101	20600	40.31474	0.28088
Grade 12	483.6926	53520	39.69142	0.17157
Grade 10	471.995	98575	37.68256	0.12002
Literate	468.3904	87309	32.12619	0.10873
Grade 8	465.2977	103060	34.89851	0.10871
Illiterate	455.9847	59226	34.57738	0.14208
Not stated	429.5322	14122	63.07624	0.53078
Total	469.7619	448844	39.52154	0.05899

Table 63 shows that students whose father was illiterate achieved a score of 456, whereas students whose father was literate achieved 468. There was a significant difference in the achievement of students whose fathers were literate compared to those whose fathers were illiterate. Similarly, students whose fathers held a Master's degree or above achieved 509 scale score. The data can be represented as in Figure 50.

Figure 50 Achievement by father's education

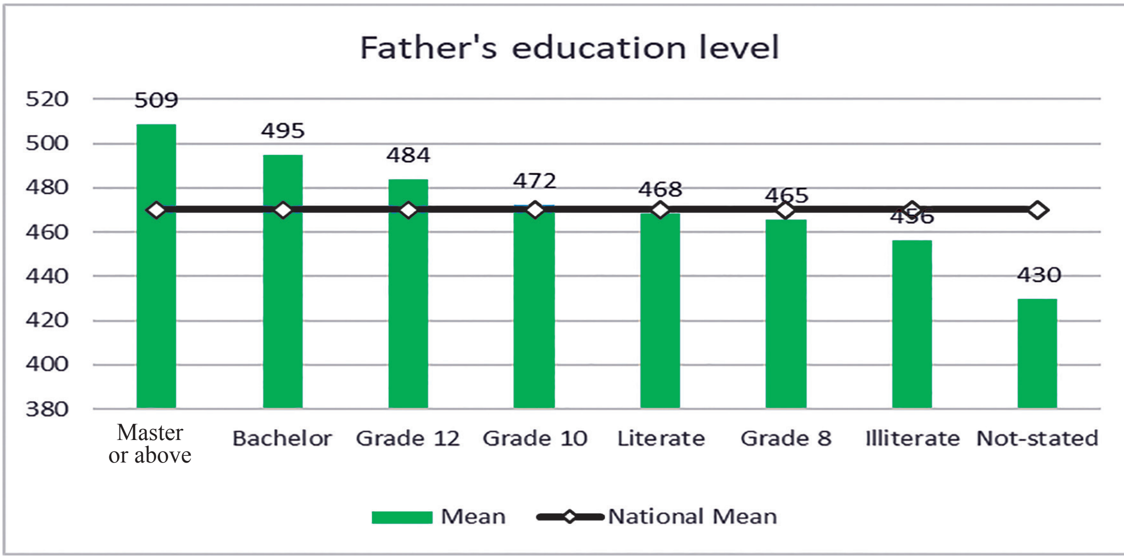


Figure 50 reveals that levels of achievement vary with the education level of the students' fathers. Students whose fathers were illiterate achieved a scale score of 456, whereas students whose fathers were literate achieved 468. There was a significant difference in the achievement of students whose fathers were illiterate compared to those whose fathers were literate. Data shows that the higher the father's level of education, the better the student's score, as students whose father passed grade 10 achieved 572, Grade 12 was 484, Bachelor's 495, and Master's or above 509. The results indicate that the higher the father's qualification, the better the student's achievement.

Achievement by mother's occupation

In NASA 2020, mother's occupation has been an important consideration in relation to students' learning achievement in science. For this purpose, nine different categories were established for analysis of mother's occupation. According to the responses given by sample students, the number of students having mothers in the various categories are indicated in Table 64.

Table 64 Achievement by mother's occupation

Mother's occupation	Mean	N	Std. deviation	Std. error of mean
Other	497.2213	8355	39.17896	0.42862
Teaching	494.8927	11195	39.67816	0.37501
Government job	488.2379	9225	40.85867	0.42541
Business	483.7024	29722	36.94617	0.2143
Foreign	478.3237	8223	34.41701	0.37955

Household	476.3401	97046	39.53007	0.12689
Labour	473.7384	5939	36.33478	0.47146
Agriculture + household	465.2568	261153	35.15791	0.0688
Work in other home	460.6791	3983	37.56759	0.59526
Not reported	425.8219	14003	63.88938	0.53992
Total	469.7619	448844	39.52154	0.05899

Table 64 presents the nine occupation categories carried out by the mothers of students. They were Agriculture (261,153), Household work (97,046), Work in other's house (3,983), Labour (5,939), Foreign employment (8,223), Teaching (11,195), Business (29,722), Government service (9,225) and Other (8,355). Details of the achievements of students in relation to their mothers' occupation are shown in Figure 51.

Figure 51 Achievement by mother's occupation

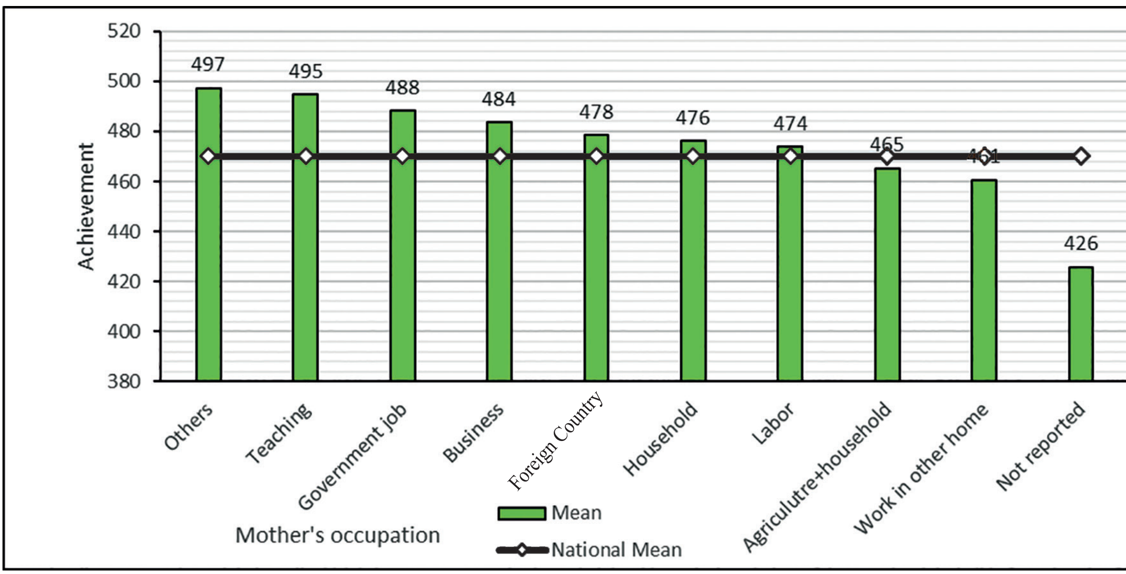


Figure 51 shows that the mother's occupation has a direct impact on the student's achievement. Children whose mothers were engaged in other profession than those mentioned in the questionnaire achieved the highest score (497), whereas children whose mothers worked in others' homes achieved the lowest score (461). Similarly, students whose mothers worked as labourers, in the household, abroad, in business, in a government job or in teaching scored 474, 476, 478, 484, 488 and 495 respectively, which are above the national average, whereas children whose mothers worked in agriculture plus household, or who work in others' homes have scores which are below the national average. There is a significant difference between high-achieving and low-achieving children at $p<.0.05$ based on their mother's professional background. Overall results show that there is a positive correlation between mother's occupation and children's learning achievement, as regular income makes the family financially sound.

Achievement by father's occupation

In NASA 2020, the professional background of the father also has a direct impact on the learning achievement of students. The father is taken to be the family breadwinner in Nepalese society and plays a dominant role in maintaining the financial status, which is related to providing better educational opportunities. The fathers' occupations were grouped into agriculture & household, only household, labour, working abroad, teaching, business, and government job. The details are presented in Table 65.

Table 65 Achievement by father's occupation

Father's occupation	Mean	N	Std. deviation	Std. error of mean
Others	497.2213	8355	39.17896	0.42862
Teaching	494.8927	11195	39.67816	0.37501
Government job	488.2379	9225	40.85867	0.42541
Business	483.7024	29722	36.94617	0.2143
Foreign (work abroad)	478.3237	8223	34.41701	0.37955
Household	476.3401	97046	39.53007	0.12689
Labour	473.7384	5939	36.33478	0.47146
Agriculture + household	465.2568	261153	35.15791	0.0688
Work in other home	460.6791	3983	37.56759	0.59526
Not reported	425.8219	14003	63.88938	0.53992
Total	469.7619	448844	39.52154	0.05899

Table 65 shows that students whose fathers' professions involved teaching, government jobs and others achieved 493, 488 and 497 respectively. Students whose fathers worked in agriculture and household work, or who worked in others' homes, achieved 465 and 461 respectively. This data can is presented as a bar diagram in Figure 52.

Figure 52 Achievement by father's occupation

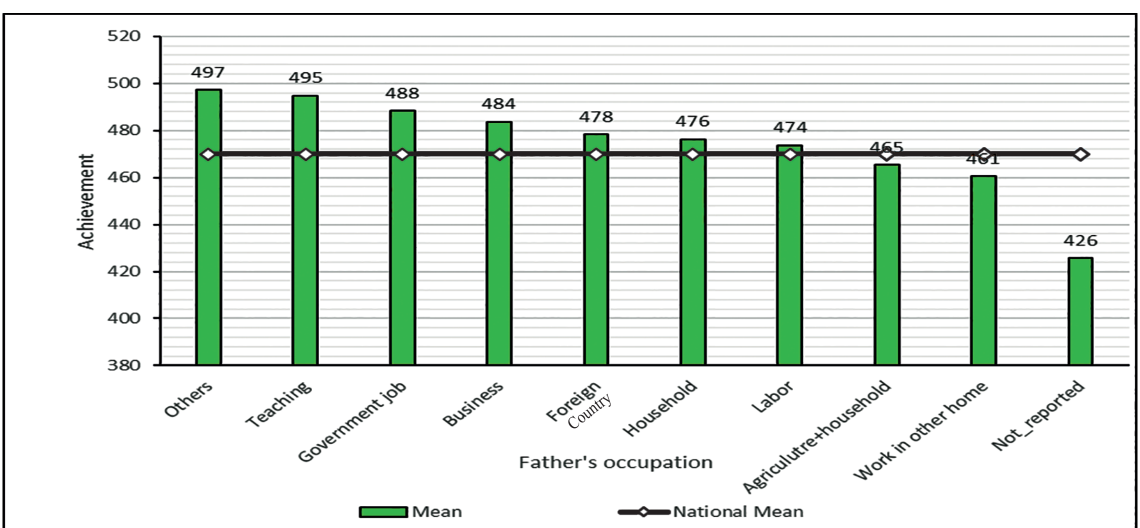


Figure 52 shows that students whose fathers worked in agriculture and household work or who worked in others' homes achieved 465 and 461 respectively, which are below the national average. Students whose fathers engaged in other categories of jobs achieved above the national average.

Achievement by availability of possessions at home

In NASA 2020, respondents were asked if their family possesses a separate study room, study table, computer, children's magazine, reference book, internet, and dictionary. Research shows that various facilities that support study play a vital role in learning achievement. In this study, students' home possessions have been categorised by number (one to seven items). Table 66 shows the achievement of students based on the availability of these home possessions.

Table 66 Achievement by availability of home possessions to students

Home possession	Mean	N	Std. deviation	Std. error of mean
None	440.1248	11210	53.36069	0.50398
One	461.9996	201869	35.43591	0.07887
Two	467.1732	49586	37.56554	0.1687
Three	471.6061	65255	35.23697	0.13794
Four	477.8703	54341	38.62457	0.16569
Five	484.5127	33752	40.20126	0.21882
Six	494.1349	18461	41.83113	0.30787
Seven or more	505.8645	14370	48.34225	0.40327
Total	469.7619	448844	39.52154	0.05899

Table 66 reveals that those students who do not have any home possessions scored 440, whereas those who responded to the seven or more possessions category scored 505 scale score. The results show that when the number of home possessions increases, the achievement of students increases significantly. The overall results indicate that the higher the number of home possessions, the better the mean scores which are above the national average. This data can be presented in line graph as in Figure 53.

Figure 53 Achievement by availability of home possessions to students

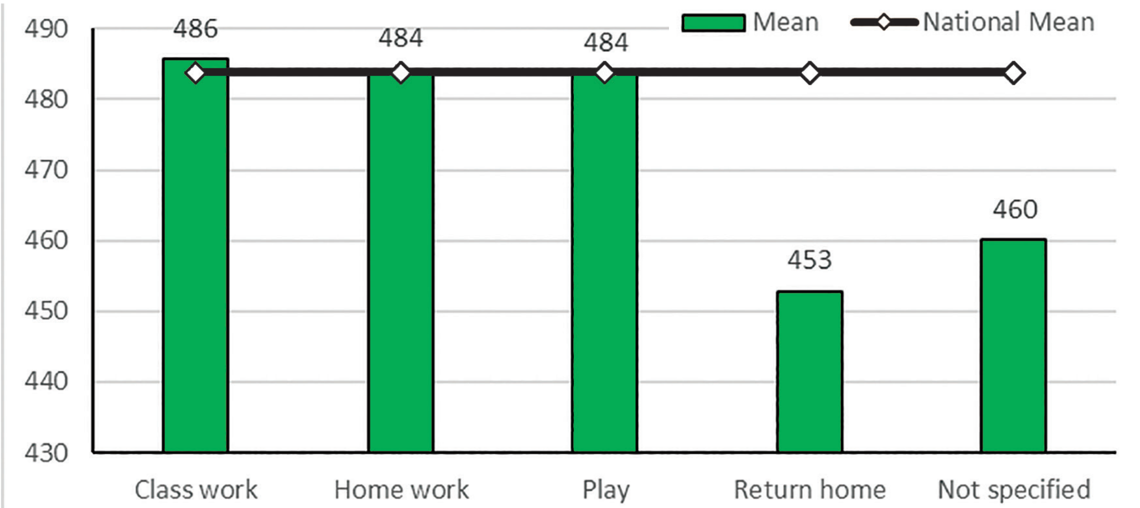


Figure 53 reveals that students who have access to a lower number of home possessions have achieved below the national average, whereas those students who have access to a larger number of home possessions have achieved higher scores which are above the national average. The blue line clearly shows that students who have access to no, one or two home possessions scored below the national average, i.e. 440, 462 and 467 respectively, whereas those who had access to three, four, five, six and seven home possessions scored 472, 478, 485, 494 and 506 scale score respectively. Overall, the results indicate that the number of home possessions has a positive correlation with the achievement of students.

Achievement by number of books available

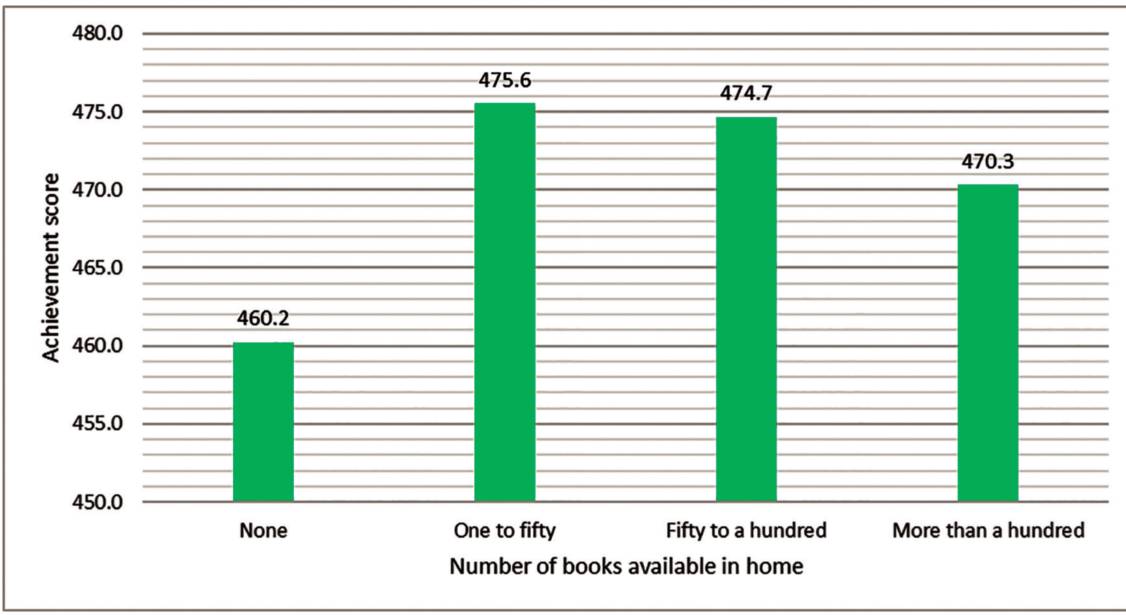
The NASA 2020 questionnaire includes a variable about the availability of books other than textbooks at home. The availability of books was categorized as one to fifty, fifty to a hundred, and more than a hundred. Table 67 provides the responses from the participants.

Table 67 Achievement by number of books available

Number of books other than textbooks	Mean	N	Std. deviation	Std. error of mean
None	460.2213	111270	36.36121	0.10901
One to fifty	475.5515	272775	37.27454	0.07137
Fifty to a hundred	474.6564	27202	42.87424	0.25995
More than a hundred	470.2968	13429	45.60377	0.39352
Not reported	442.5345	24166	49.49374	0.31838
Total	469.7619	448844	39.52154	0.05899

The number of students reported in Table 67 shows that 25% of respondents did not have any books, 61% had between one and fifty books, 6% had between fifty and a hundred books, and 3% of students had more than a hundred books available at home. The data reveals that students who have between one and fifty books at home achieved the highest scores (476), above the national average, while students who possess more than a hundred books achieve 470 scale score. The data provides evidence that one to fifty books best suits the students' level and helps to achieve higher scores. Figure 54 shows the differences in achievement based on the books available to students.

Figure 54 Achievement by number of books available in home



Overall, there is a significant difference between the scores of students who own books and those who do not have any books except textbooks at home. The results indicate that there is a positive relation between having additional books and the achievement of students.

Achievement by distance between home and school

One of the NASA 2020 background questions was about the walking distance between home and school. Students' responses in relation to their achievement are presented in Figure 55.

Figure 55 Achievement by distance between home and school

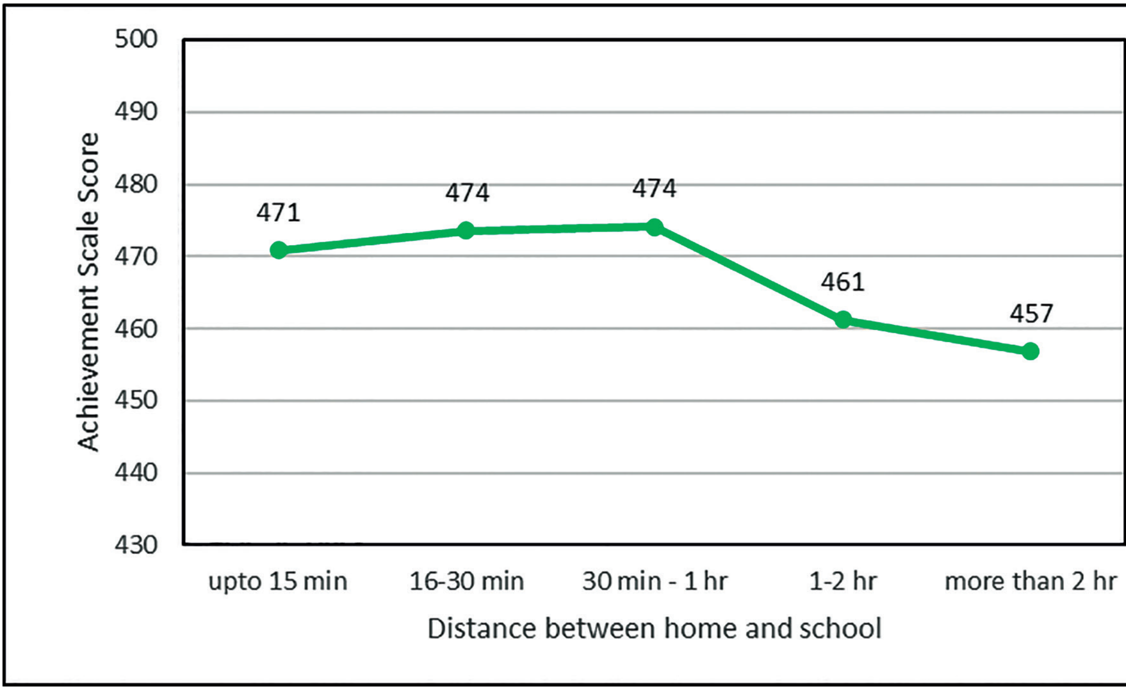


Figure 55 reveals that students whose walking distance from home to school was up to 15 minutes, 15 to 30 minutes, and between 30 minutes and 1 hour achieved 471, 474 and 474 respectively, whereas the achievement for those students whose walking distance from home to school was between 1 and 2 hours or more than 2 hours was distinctly lower, with scale scores of 461 and 457 respectively. The results display that there is a positive correlation between the home–school distance and the student's achievement.

Achievement by possession of a separate mobile phone

In NASA 2020, one of the questions asked was whether students possessed a separate mobile phone or not. Table 68 shows the details of the students' responses.

Table 68 Achievement by separate mobile phone possession

Separate mobile	Mean	N	Std. deviation	Std. error of mean
No	472.3741	338367	37.18873	0.06393
Yes	466.0726	92303	42.54937	0.14005
Not specified	439.8654	18174	50.41668	0.37398
Total	469.7619	448844	39.52154	0.05899

Table 68 presents 75% of students expressed that they did not possess a separate mobile, while 21% of students responded that they had their own mobile phone. This is presented in bar diagram form in Figure 56.

Figure 56 Achievement by separate mobile phone possession

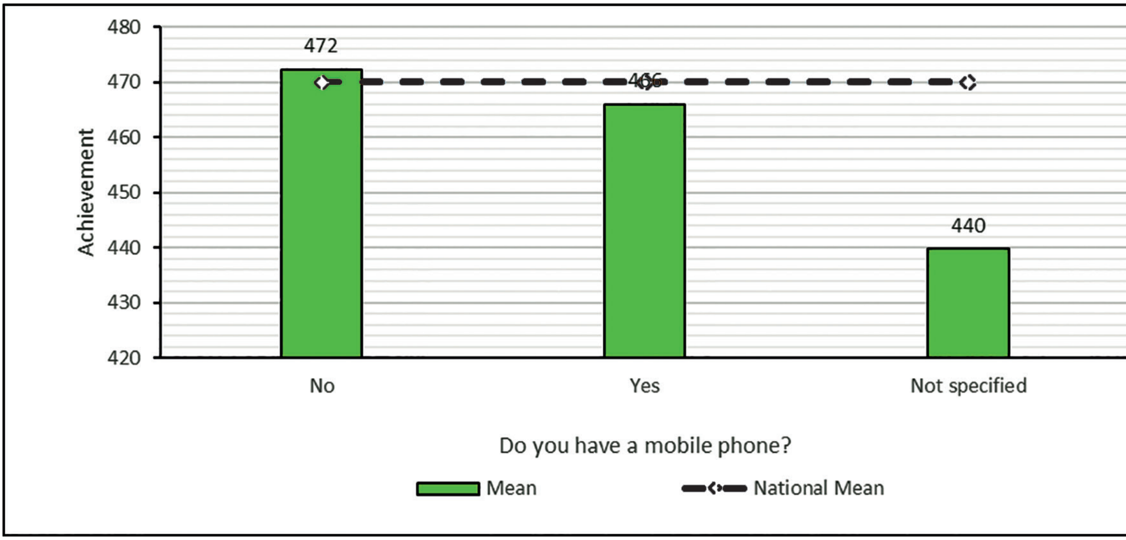


Figure 56 reveals that students who possess their own mobile phone achieved a score of 466, and those who do not possess a separate mobile phone scored 472. This result proves that there is a significant difference between the achievement of students who have a separate mobile phone compared to those who do not. Overall, the results show that students who possess their own separate mobile might waste time using their mobile rather than studying, and as a result their achievement has decreased significantly.

Achievement by access to social media

In NASA 2020, students were also asked to respond to a question regarding access to social media. The responses were categorised as 'Yes' and 'No'. Table 69 shows the data from respondents.

Table 69 Achievement by access to social media

Social media access	Mean	N	Std. deviation	Std. error of mean
Yes	478.8968	99133	43.44256	0.13798
No	469.0684	325572	36.36337	0.06373
Not specified	441.6009	24139	48.12807	0.30977
Total	469.7619	448844	39.52154	0.05899

Table 69 depicts that the mean score of the respondents who had access to social media was 479 scale score, whereas those who did not have access to social media scored 469. This is a difference of 10 scale score, and the data is presented as a bar diagram in Figure 57.

Figure 57 Achievement by access to social media

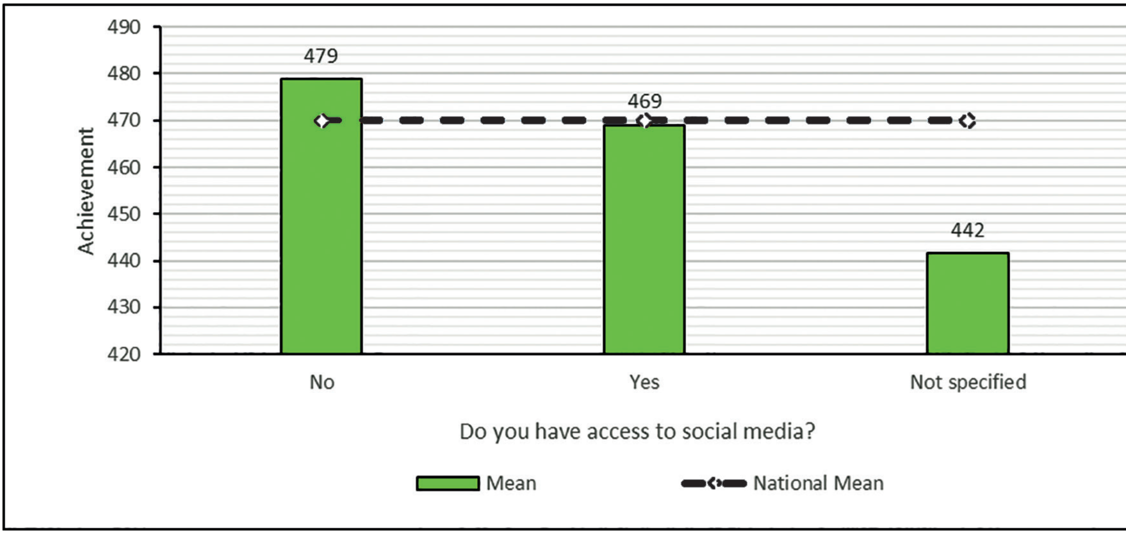


Figure 57 represents the achievements of respondents in relation to their access to social media. The data reveals that those students who did not have access to social media achieved 469, which is below the national average, whereas respondents who had access to social media achieved 479 which is 10 scale score higher. This data proves that there is a significant difference in the achievement of the respondents who had access to social media compared to those who did not. Overall, the results show that access to social media has a positive relation to the achievement of students.

Students' learning achievement by age group

Students were asked to state their age as a background variable so that the relationship between age and educational achievement could be analysed. The data shows that they were grouped into five different age strata: 12 and below, 13, 14, 15, and 16 or above. Students' learning achievements based on their age group were found to be different, as shown in Table 70.

Table 70 Students' learning achievement by age group

Age of students	Mean	N	Std. deviation
12 years and below	477.3821	29242	38.07874
13 years	478.4859	109043	39.30365
14 years	473.4079	150568	38.76435
15 years	464.6784	80276	35.29536
16 years and above	456.8385	34079	33.5086
Not specified	450.5971	45635	44.20649
Total	469.7619	448844	39.52154

Table 70 shows that students in the 13 years age group achieved the highest score (478), and students at the age of 16 or above achieved the lowest score (467), which is below the national mean. This indicates that maintaining enrolment at the proper age (12, 13 and 14 years) can have a positive effect on science learning for grade 8 students. This is presented in bar diagram form in Figure 58.

Figure 58 Students' learning achievement by age group

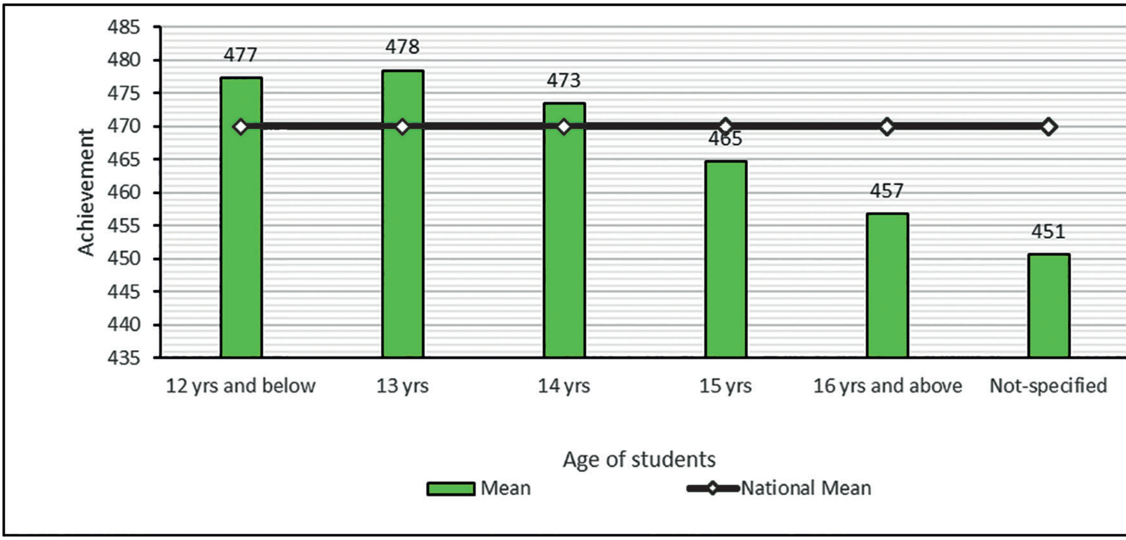


Figure 58 depicts a clear picture that students who were at the age of 12, 13 and 14 years have achieved above the national average, with scores of 477, 478 and 473 respectively, while students of age 15 and 16 or above have achieved distinctly below the national average, with scores of 465 and 457 respectively. This result shows that maintaining enrolment at the proper age (12, 13 and 14 years) can have a positive effect on science learning for grade 8 students. It implies that students within the appropriate age group learn science better than those who are over that age group.

Achievement by ethnic group

In NASA 2020, students were asked to state their ethnicity, which was then categorised into nine groups as Hill Brahmin, Hill Janajati, Hill Others, Hill Dalit, Madhesi Brahmin, Madhesi Janajati, Madhesi Others, Madhesi Dalit, Mountain Others. The result can be seen in Table 71.

Table 71 Achievement by ethnic group

Ethnic group	Mean	N	Std. deviation
Hill Brahmin	483.0765	114787	36.41606
Hill Janajati	475.9682	87628	32.09942
Hill Others	473.8398	16161	33.41931
Hill Dalit	471.499	29190	30.84722
Madhesi Brahmin	469.2421	38680	41.60668
Madhesi Janajati	464.1564	58441	37.04077
Madhesi Others	459.2145	33430	39.87711
Mountain Brahmin	457.2899	6236	38.60219
Madhesi Dalit	457.1877	16744	38.74211
Not specified	444.5192	47546	49.0648
Total	469.7619	448844	39.52154

Table 71 shows the mean score in science achieved by Hill Brahman, Hill Janajati and Dalit students to be 483, 475 and 471 respectively, which were found to be higher than the scores achieved by the other caste/ethnic categories and above the national average mean. Students from Madhesi Dalit and Mountain Brahman achieved the lowest scores, with 457 each, and along with Madhesi Brahman (469) and Madhesi Janajati (464), all the other groups were found to have scored less than the national average. The details are presented in bar diagram form in Figure 59.

Figure 59 Achievement by ethnic group

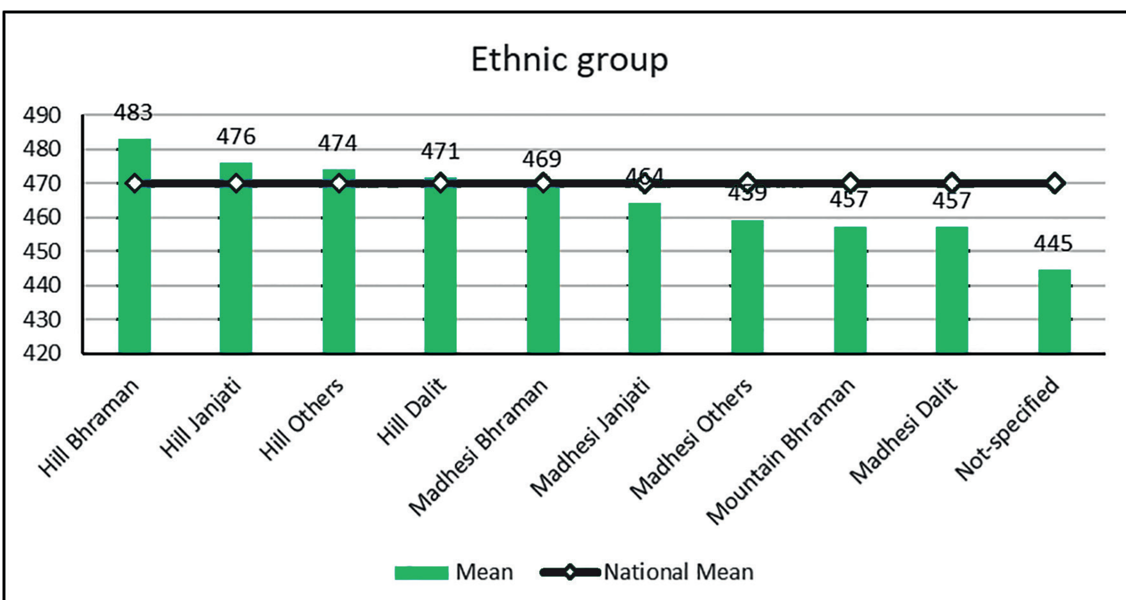


Figure 59 shows that Hill Brahman, Hill Janajati, Hill Others and Hill Dalit achieved 483, 476, 474 and 471 respectively, which are above the national average, whereas Madhesi Brahman, Madhesi Janajati, Madhesi Others, Mountain Brahman and Madhesi Dalit achieved distinctly below the national average mean. The results show that geographical and ethnic variations have also brought differences in the level of achievement among students.

Achievement by medium of instruction

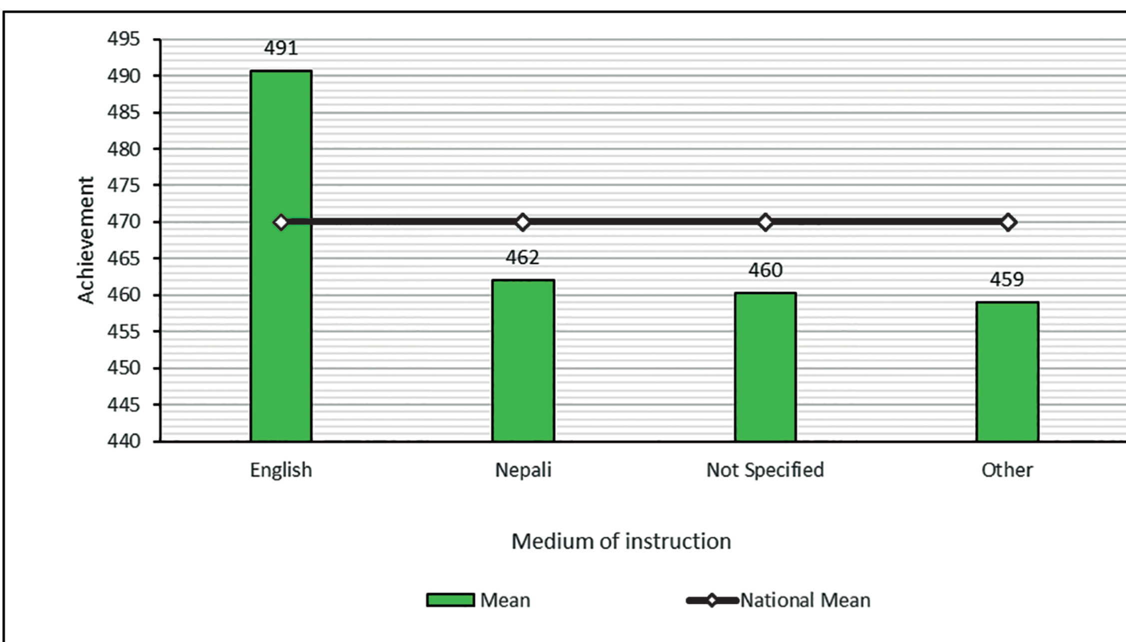
In NASA 2020, respondents were asked whether the medium of instruction was Nepali or English. The students' responses are presented in Table 72.

Table 72 Achievement by medium of instruction

Medium of instruction	Mean	N	Std. deviation	Std. error of mean
English	490.6735	128102	36.53529	0.10208
Nepali	462.0534	231630	35.16902	0.07307
Not specified	460.3146	49460	47.59668	0.21402
Other	459.0173	39653	36.4794	0.18319
Total	469.7619	448844	39.52154	0.05899

Table 72 indicates that 28% of the total students stated that the medium of instruction for teaching science was English, whereas 51% students responded that the medium of instruction in science was Nepali. The achievement of students based on medium of instruction is presented in Figure 60.

Figure 60 Achievement by medium of instruction



The bar diagram in Figure 60 depicts that students who were taught in English medium scored 491, whereas students taught in Nepali medium scored 462. The data reveals the fact that there is a significant difference in the achievement of students who were taught in English compared to those taught in Nepali.

Achievement by family size

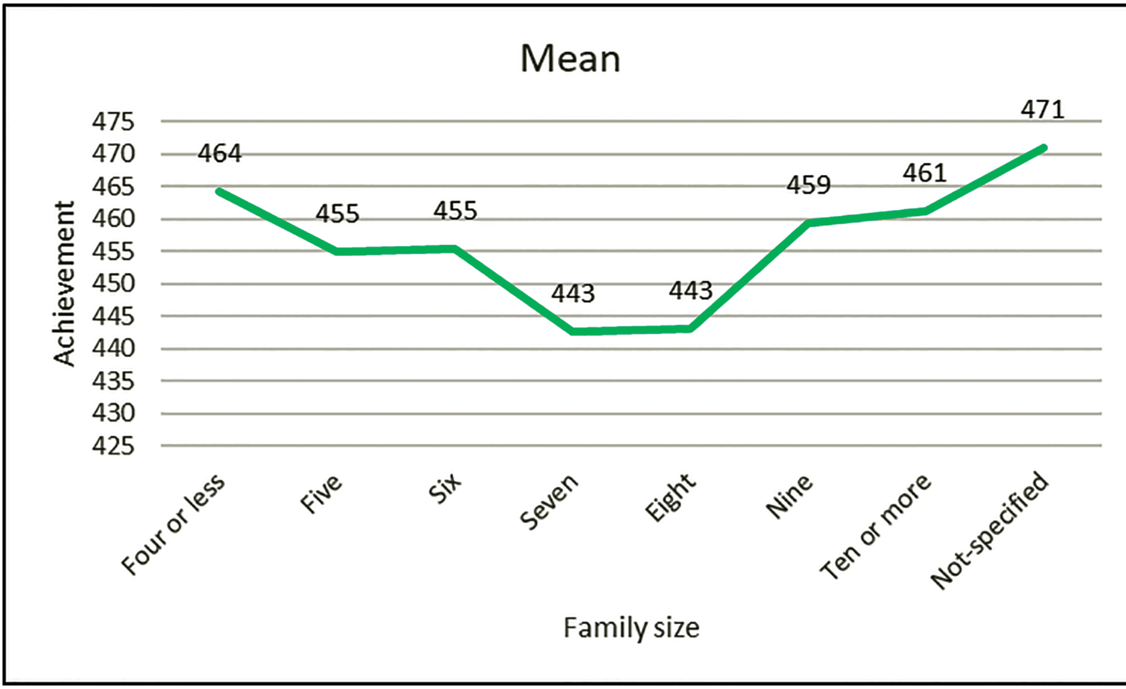
Family size is one of the factors that has a direct relation in the care, health, opportunity, facilities and study of children. The students were asked the number of members in their family and their responses were grouped into 4, 5, 6, 7, 8, 9, or 10 or more. The details are presented in Table 73.

Table 73 Achievement by family size

Size of the family	Mean	N	Std. deviation	Std. error of mean
Four or less	464.323	4168	44.09237	0.68296
Five	454.9256	5467	47.06612	0.63653
Six	455.3735	3508	44.57444	0.75255
Seven	442.6516	2350	48.93396	1.00947
Eight	443.0092	1829	37.64153	0.88021
Nine	459.373	1016	42.06709	1.31968
Ten or more	461.1444	22125	39.83921	0.26784
Not specified	470.9081	408381	39.00766	0.06104
Total	469.7619	448844	39.52154	0.05899

Table 73 indicates that 91% of students did not specify their family size. The number of students who expressed that there were four or less members in their family was 4,168. Table 73 shows that students with a family size of four or less, nine, or ten or more achieved 464, 459 and 461 respectively. However apart from the group who did not specify family size, none have scored at the level of the national average mean, which can be represented in a line graph as in Figure 61.

Figure 61 Achievement by family size



In Figure 61, the x-axis represents the size of the family and the y-axis shows the achievement scale. The line graph indicates the achievement score of the children of different family sizes. The students whose family size consists of four or less achieved a mean score of 464, which decreases drastically and reaches 443 when the family size becomes seven or eight members. Interestingly, when the family size increases to eight or nine, the mean score also increases, reaching 459 and 461 respectively, while the group which did not specify family size achieved 471, in line with the national average.

Achievement by availability of science textbook

In NASA 2020, the respondents were asked to comment on the timely availability of science textbooks. The students' responses of whether science textbooks were available to them or not are presented in Table 74.

Table 74 Achievement by availability of science textbook

Science textbook	Mean	N	Std. deviation	Std. error of mean
Yes	472.3612	400565	38.38026	0.06064
No	452.7664	37045	34.60052	0.17977
Not specified	433.1236	11234	58.30463	0.55008
Total	469.7619	448844	39.52154	0.05899

Table 74 illustrates that 89% of students responded that science textbooks were available to them, whereas 8% of students reported that they did not get science textbooks. The learning achievement of students in science is shown in Figure 62.

Figure 62 Achievement by availability of science textbook

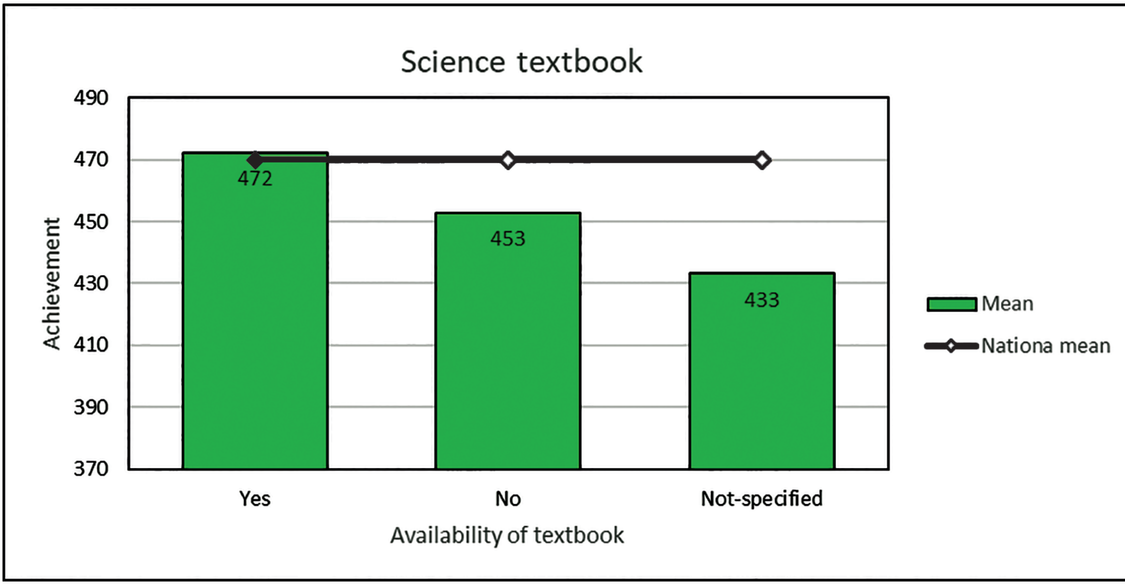


Figure 62 depicts the reality that students who obtained science books on time achieved a mean score of 472, which is above the national average, whereas those students who did not get science textbooks scored 453, which is below the national average. The difference of 19 scale score between them shows that there is a significant difference in the achievement of students due to the availability of science textbooks.

Achievement by medium of instruction in community schools

In NASA 2020, students were asked to state the medium of instruction. Out of 365,953 community school students, 224,025 of them responded that they had Nepali medium of instruction in science, whereas 63,222 of them informed that their medium of instruction for science is English. The achievements of the English- and Nepali-medium students are given in Table 75.

Table 75 Achievement by medium of instruction in community schools

Medium of instruction	Mean	N	Std. deviation	Std. error of mean
English	478.0417	63222	35.37253	0.14068
Nepali	461.2093	224025	34.80605	0.07354
Other	456.1692	36259	34.97979	0.1837
Not specified	454.5267	42447	47.07659	0.2285
Total	462.8428	365953	37.28316	0.06163

Table 75 shows that 17% of students, who studied in English medium, achieved a mean score of 478, whereas 61% students, who were taught in Nepali medium, scored 461 scale score, which is below the national average. The difference of 17 scale score between them is presented in Figure 63.

Figure 63 Achievement by medium of instruction in community schools

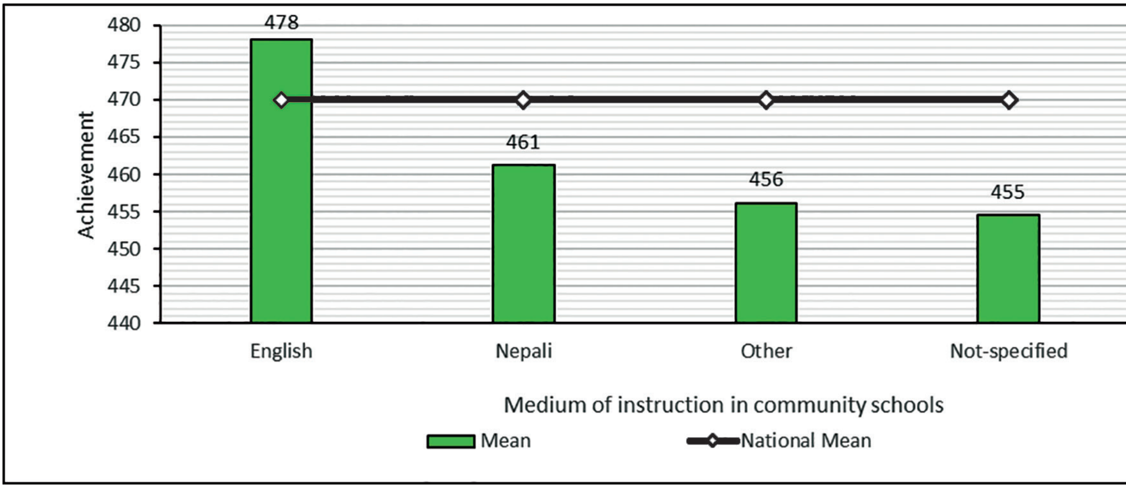


Figure 63 reveals that students studying in English-medium community schools achieved a higher score (478), compared to the score (461) of students in Nepali-medium community schools. The results indicate that there is a significant difference in achievement between English- and Nepali-medium teaching in community schools in Nepal.

Achievement by public and private school

Institutional schools are funded by the private sector, whereas community schools' sole investment is government funding. NASA 2020 collected data from 365,953 students studying in community schools and 82,808 students studying in institutional schools. The achievement of students based on the type of school is presented in Table 76.

Table 76 Achievement by type of school

Type of school	Mean	N	Std. deviation	Std. error of mean
Community	462.8428	365953	37.28316	0.06163
Institutional	500.3007	82808	34.31264	0.11924
Total	469.7547	448761	39.52112	0.059

Table 76 indicates that 82% of students study in community schools, whereas 18% of students study in institutional schools. The mean scores of students at community and institutional schools are 463 and 500 respectively. Students' achievement in relation to the type of school is presented in bar graph form in Figure 64.

Figure 64 Achievement by type of school

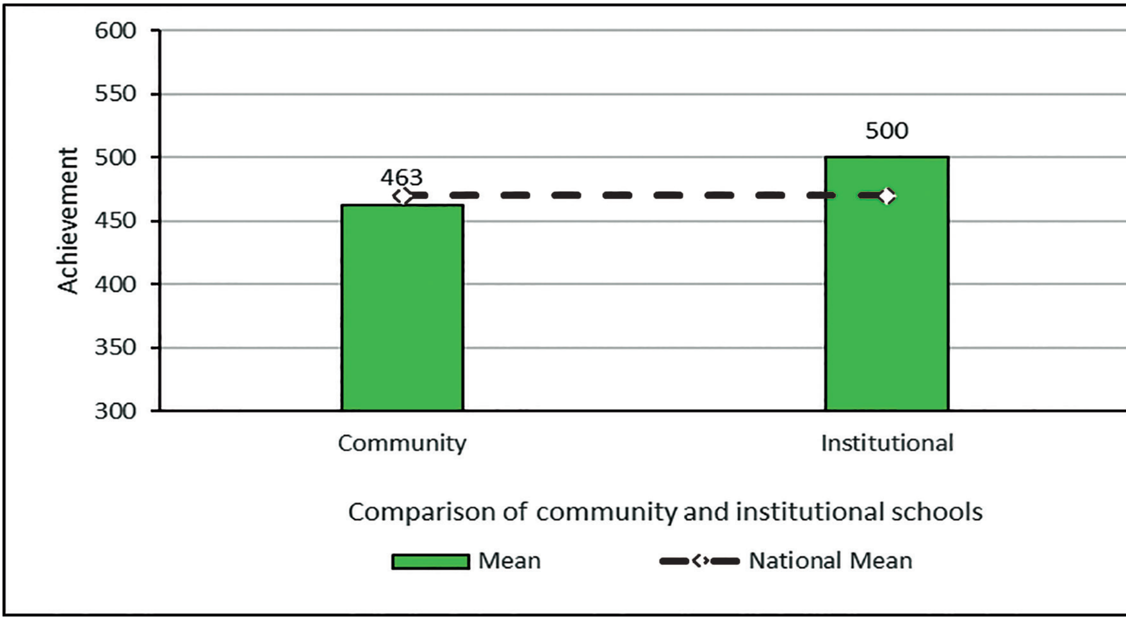


Figure 64 shows that the mean scores for students of institutional and community schools were 500 and 463 respectively. The score for community schools was lower than both the national mean and institutional schools' achievement. With a scale score variation of 37, the gap between the two types of schools was significantly different at $p < 0.05$. The difference is alarming, as the gap between the community and institutional schools is wide. The results indicate that the educational process in community schools, where the government has invested a huge amount of resources for meagre achievement, is a crucial area of concern. Private-sector-invested institutional schools have shown higher learning outcomes. The reasons behind such difference could be explained by looking at differences in time on task, school management, school location, facilities provided by parents and so on.

Status of scholarship

In NASA 2020, students were asked to state whether they had received a scholarship or not. The responses of students were given in Table 77.

Table 77 Number of students receiving scholarships

Receiving scholarship	N	Percentage
Yes	177535	39.6
No	258517	57.6
Not stated	12792	2.9
Total	448844	100

The data presented in Table 77 depicts that 40% of students expressed that they had received scholarships, while 58% of respondents stated that they had not got any opportunities of scholarship in schools. The results indicate that at least a third of students have obtained either full or partial scholarships at school level.

Scholarships provided in community schools

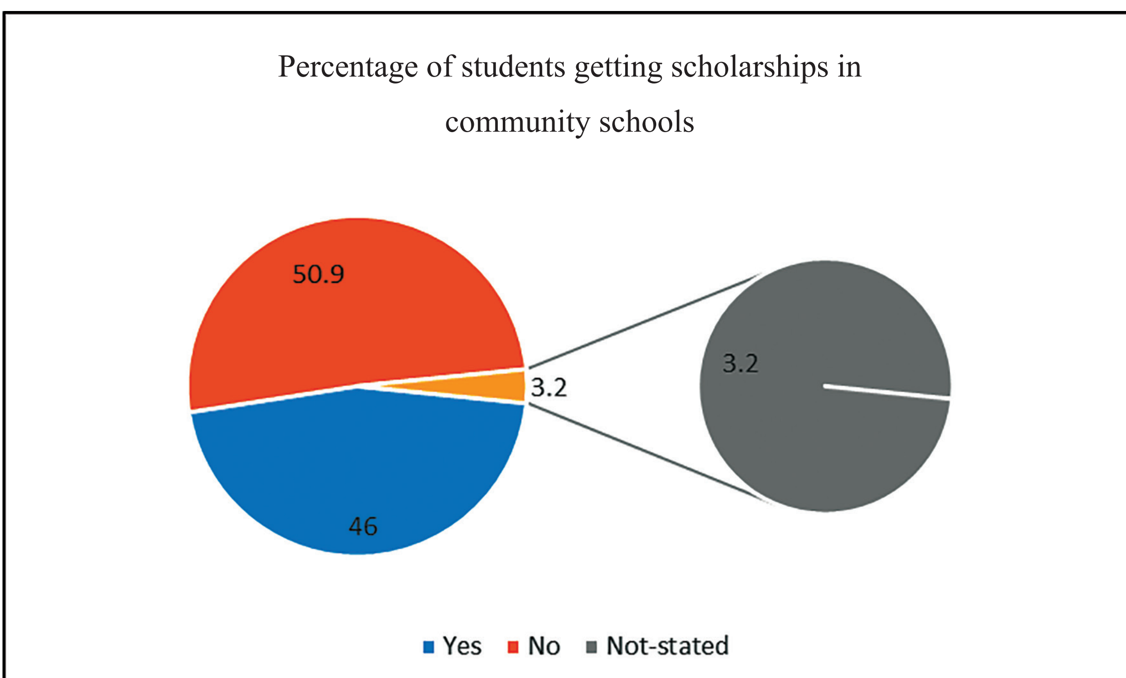
Students studying in community schools have opportunities to go for scholarship schemes in Nepal. In the background questions of NASA 2020, respondents were asked to state whether they received a scholarship or not. The responses provided by the students are presented in Table 78.

Table 78 Scholarships provided in community schools

Scholarship given or not	N	Percentage
Yes	168208	46
No	186214	50.9
Not stated	11532	3.2
Total	365953	100

Table 78 shows that 46% of students studying in community schools obtained scholarships, and 50% of them have not received any scholarship in school. This data is represented in Figure 65.

Figure 65 Percentage of students obtaining scholarships in community schools



The pie chart in Figure 65 reveals the percentage of students receiving scholarships in community schools in Nepal. The chart clearly shows that 46% of students got a scholarship, while 51% of students did not get any scholarships in grade 8.

Scholarships provided in institutional schools

Scholarships play a crucial role in institutional schools too. In the NASA 2020 study, students from institutional schools were asked to state their school scholarship status. The data given in Table 80 shows that 9,307 students have obtained scholarships, while 72,241 students have not received any scholarships at institutional schools.

Table 79 Scholarships provided in institutional schools

Scholarship given or not	N	Percent
Yes	9307	11.2
No	72241	87.2
Not-stated	1260	1.5
Total	82808	100

Table 79 shows that 11% of students in institutional schools have received scholarships and 87% of students studying in institutional schools have not received any scholarships in school. This is presented in Figure 66.

Figure 66 Scholarships provided in institutional schools

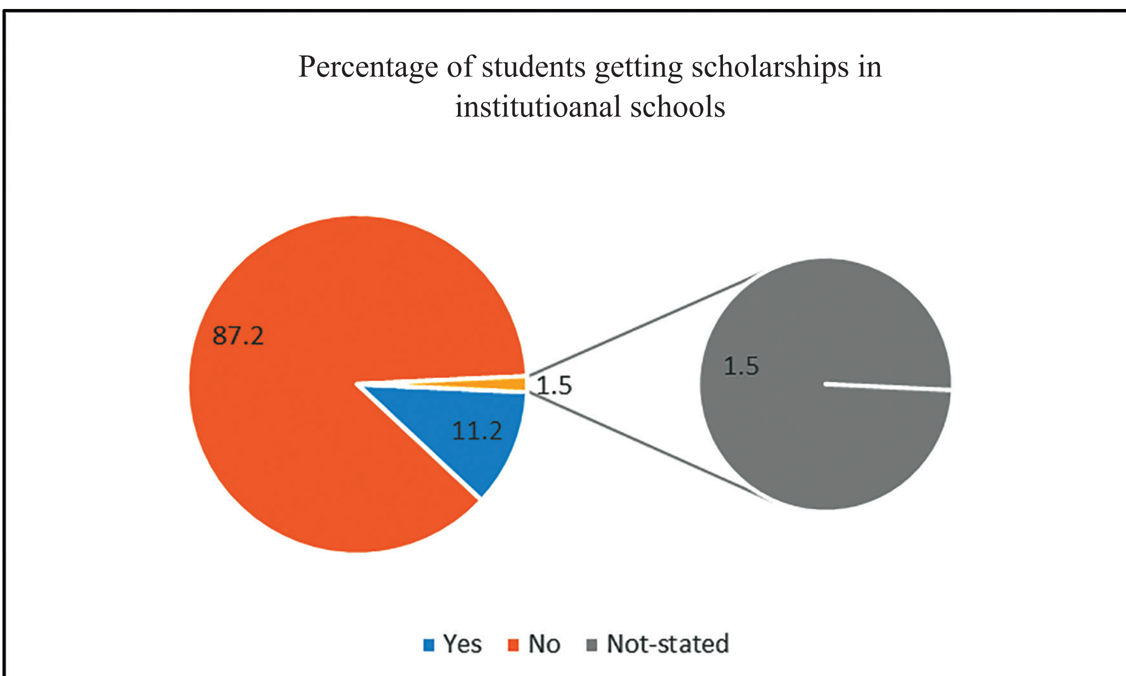


Figure 66 displays the percentages of students learning in institutional schools who have and have not received scholarships. The pie chart reveals that 82% of institutional students have not received any school scholarships. The data also shows that only 11% of students have got either full or partial scholarships in institutional schools in Nepal. Scholarship and overall achievement

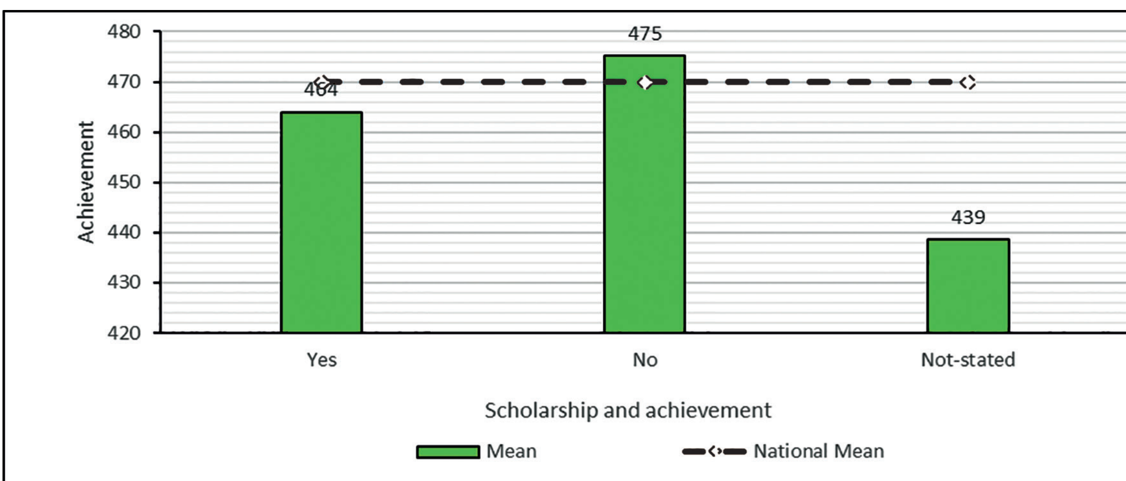
In NASA 2020, the relationship between scholarship status and students' achievements has been analysed. To the question regarding whether students were given a scholarship or not, 177,535 of them responded that they were given a scholarship and 258,517 students informed that they were not given a scholarship in school. The mean scores of these students are presented in Table 80.

Table 80 Achievement in science by scholarship

Scholarship given or not	Mean	N	Std. deviation	Std. error of mean
Yes	463.9787	177535	37.52023	0.08905
No	475.273	258517	38.47959	0.07568
Not stated	438.6497	12792	57.43406	0.50781
Total	469.7619	448844	39.52154	0.05899

Table 80 depicts that 40% of students have received scholarships, whereas 58% of them responded that they have had no opportunities of scholarships. The achievement of students is displayed in Figure 67.

Figure 67 Achievement in science by scholarship



The bar diagram in Figure 67 illustrates the mean score of students in relationship to their scholarship status. The students who obtained scholarships scored 464, which is below the national average, whereas those who have not had any opportunity for scholarship achieved 475, above the national average. The results indicate that these scholarships are provided to poor and marginalised groups rather than intelligent students.

Scholarship and the achievement of students in community schools

In NASA 2020, one of the areas of study was scholarship status and its relation to achievement. Students' responses are presented in Table 81.

Table 81 Scholarship and the achievement of students in community schools

Scholarship given or not	Mean	N	Std. deviation	Std. error of mean
Yes	461.8261	168208	36.16528	0.08818
No	465.6313	186214	35.94049	0.08329
Not stated	432.6442	11532	55.88348	0.52039
Total	462.8428	365953	37.28316	0.06163

Table 81 shows that 168,208 students stated that they were given scholarships, whereas 186,214 students responded that they were not given any scholarships. The relationship of their achievement with scholarship status is presented in Figure 68.

Figure 68 Scholarship and the achievement of students in community schools

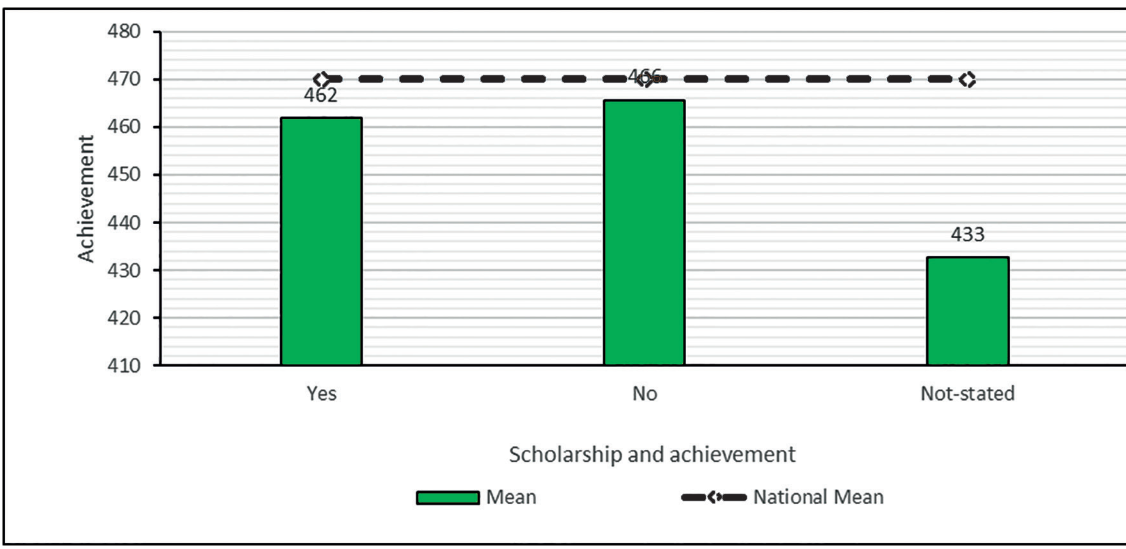


Figure 68 shows the mean score of students who had received scholarships compared to those who did not in community schools. The mean score of those who obtained scholarships was 662, and the score for those who did not receive scholarships was 466 scale score. The data shows that there is no significant difference in score between the students who received scholarship and those who did not. The results indicate that both mean scores for community schools are below the national average.

Scholarship and the achievement of students in institutional schools

One of the areas of study in NASA 2022 was the impact of scholarships on the learning achievement of students in institutional schools. Based on the data available, 9,307 students obtained a scholarship, while 72,241 students did not receive a scholarship. Table 82 presents the data for students and their achievement in institutional schools.

Table 82 Scholarship and the achievement of students in institutional schools

Scholarship given or not	Mean	N	Std. deviation	Std. error of mean
Yes	502.8095	9307	40.26877	0.41742
No	500.094	72241	33.34556	0.12406
Not stated	493.6162	1260	39.4532	1.11148
Total	500.3007	82808	34.31264	0.11924

Table 82 shows that 11% of students have received scholarships, whereas 82% of students have received no scholarships in institutional schools in Nepal. Students who received scholarships obtained a mean score of 503, while students who did not receive a scholarship achieved 500 scale score. The results indicate that the students who received a scholarship scored slightly higher than those who did not get a scholarship. This is shown in Figure 69.

Figure 69 Scholarship and the achievement of students in institutional schools

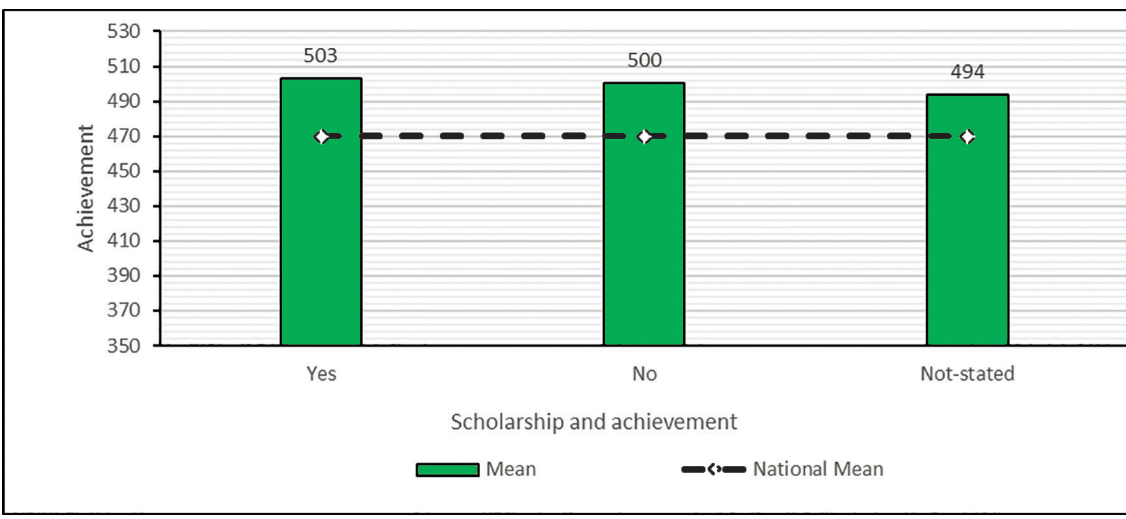


Figure 69 shows that the achievement score of students who had the opportunity to receive a scholarship was slightly higher than the mean score of those who did not get a scholarship. The mean score of those who received a scholarship was 503, and the mean score of those who did not receive a scholarship was 500, which are both above the national mean score.

Achievement by homework assigned

An important consideration in the teaching–learning process is the practice of teachers giving homework to students and then giving feedback on the homework. In NASA 2020, the sample students were asked how often their teachers give them feedback on their homework. Accordingly, three categories of students were identified: (i) those who receive feedback on their homework every day; (ii) those who receive feedback sometimes; and (iii) those who never receive feedback on their homework. The data is shown in Table 83.

Table 83 Achievement by homework assigned

Homework given by science teacher	Mean	N	Std. deviation	Std. error of mean
Always	470.3511	330909	38.03078	0.06611
Sometimes	473.0029	103767	38.08168	0.11822
Never	455.0969	3998	45.60482	0.72128
Not stated	423.2886	10171	62.43285	0.61907
Total	469.7619	448844	39.52154	0.05899

Table 83 gives the number of sample students in each of the three categories as 330,909, 103,767 and 3,998 respectively. The data shows that 74% of students have regular homework, 23% of students get homework sometimes and nearly 1% never get homework. This is presented in Figure 70.

Figure 70 Achievement by homework assigned

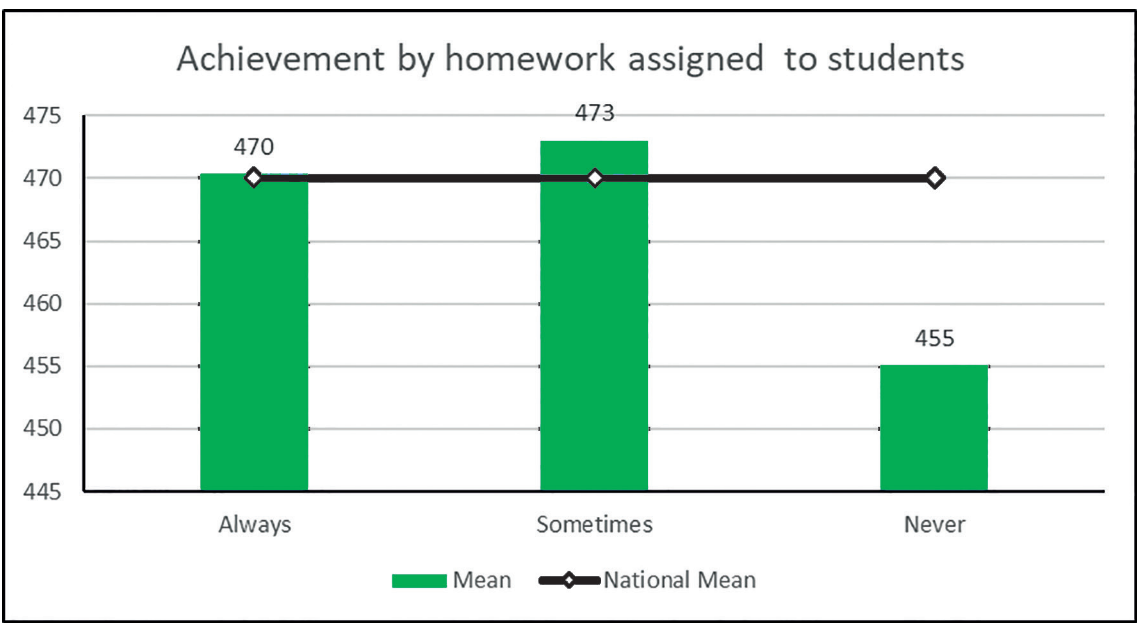


Figure 70 shows that students who always receive homework scored 470, whereas those who got homework only sometimes achieved a score of 473, slightly higher than the previous category. The bar diagram shows that students who never got homework achieved a mean score of 455, which is below the national score.

Achievement by homework correction and feedback

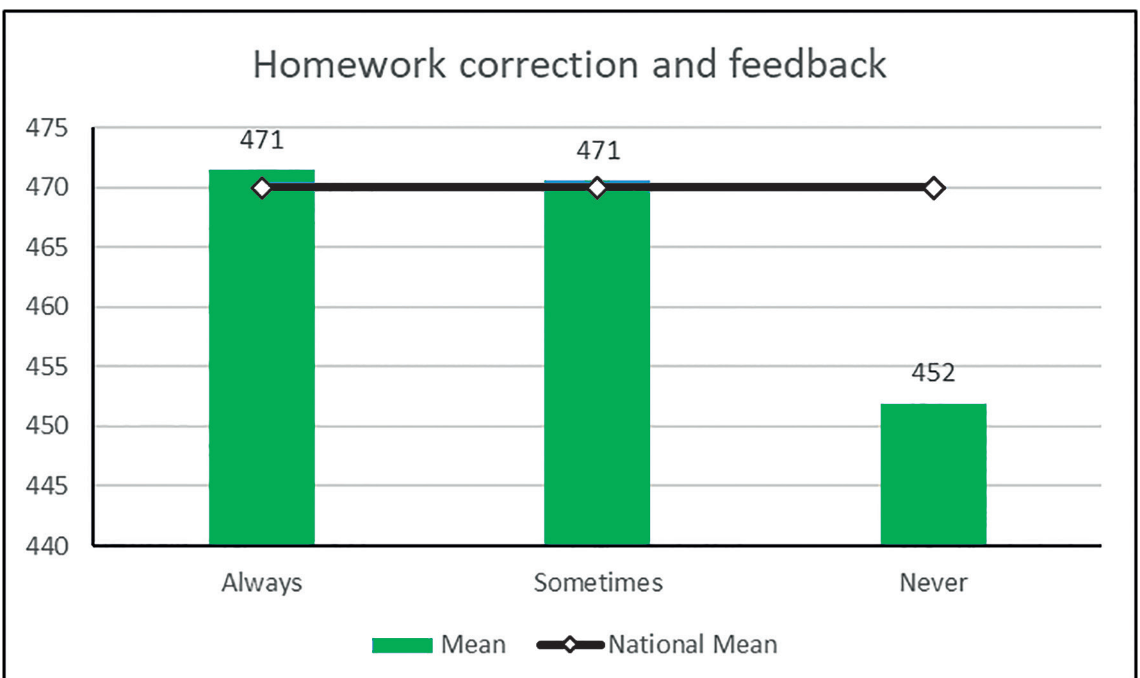
There is a relationship between homework and teachers' feedback on homework and the achievement of students. Research shows that teachers regularly checking and giving feedback on students' classwork, homework, project work and tests has a creditable role in improving students' learning performance (Dahal, 2019, p.76). Thus, NASA 2020 included the question: 'How often does your teacher give you feedback on homework?' The responses are plotted with the achievement score as shown in Table 84.

Table 84 Achievement by homework correction and feedback

Homework correction and feedback	Mean	N	Std. deviation	Std. error of mean
Always	471.492	321816	37.55732	0.06621
Sometimes	470.5774	108201	39.0453	0.1187
Never	451.9402	5823	45.75886	0.59967
Not stated	428.1444	13005	58.55188	0.51344
Total	469.7619	448844	39.52154	0.05899

Table 84 illustrates that students whose homework had been corrected regularly achieved a scale score of 471, whereas those students whose homework had been never been corrected achieved 452. The data can be shown as in the bar diagram in Figure 71.

Figure 71 Achievement by homework correction and feedback



The data presented in Figure 71 reveals the positive relationship between regular checking and feedback of homework and students' learning. The mean score of students who received feedback on their homework either regularly or occasionally was 471 for both. On the other hand, the mean score of students who never received feedback on their homework was 452. The difference between the highest mean score and the lowest mean score was 19 scale score, which is statistically significant at $p<0.05$. In conclusion, regular feedback on homework has a positive impact on achievement in science.

Students' attitude towards science teacher

One of the areas of concern in NASA 2020 has been to see whether the attitude of students towards their teachers has any correlation with student achievement, as the teaching–learning process is one that involves interaction between teacher and student. Thus, an attempt was made to record how positive (or negative) the students' attitudes towards teachers were. Attitudes towards teachers were determined on eight criteria, namely, the teacher loves, listens, gives physical punishment, treats equally, tells difficult things, gives homework, checks homework and gives feedback, and teaches for the whole period. The attitudes of teachers were measured on a Likert scale of strongly agree, partially agree, partially disagree and strongly disagree. Accordingly, the students' attitude was compared with their achievement. The results are depicted in Table 85.

Table 85 Attitude towards teacher in science

Statement	Strongly agree	Partially agree	Partially disagree	Strongly disagree	Not stated	Total
Teacher loves us	81	12	2	1	4	100
Most of the teachers do not listen	48	32	6	4	10	100
Teacher gives physical punishment	35	26	12	14	12	100
Teacher treats us equally	78	9	3	3	7	100
Teacher tells us to do beyond our ability	85	6	1	2	6	100
Teacher gives homework	83	7	1	2	7	100
Teacher checks homework and gives feedback	78	11	2	2	7	100
Teacher teaches for the whole period	64	19	4	3	10	100

Table 85 shows that 81% of students strongly agreed that their teacher loved them. About 48% of students strongly agreed that their teachers listen to them, and only 35% of them agreed that their teachers give physical punishment. According to the data, 78% of teachers treat students equally, and 85% of teachers tells us to do beyond our ability. Students revealed that 83% of teachers give homework, whereas 78% of them check homework and give feedback. As stated by students, only 64% of teachers teach for the whole period. As the data showed, 26% of students either partially or fully disagreed that their teacher gave physical punishment. However, 32% of students partially agreed that their teachers give physical punishment.

Attitude of students towards school

In NASA 2020, questions were also asked to measure the attitude of students towards school. The responses of students are presented in Table 86.

Table 86 Attitude of students towards school

Statement	Strongly agree	Partially agree	Partially disagree	Strongly disagree	Not stated	Total
I like to come to school.	90	4	1	1	4	100
Friends in school treat me well.	72	18	2	2	6	100
School have playing, drinking and toilet facilities	81	9	2	2	6	100
I participate in child club and children's programmes	54	27	4	6	9	100

Table 86 shows details of the responses from students. As shown in the table, 90% of the students agreed that they liked to come to school; 81% of students accepted that there are playing, drinking water and toilet facilities; 54% of them even participate in child club and children's programmes too.

Attitude of the students towards science subject

In the study, questions were asked in regard to students' attitudes towards science. The responses were categorised and are presented in Table 87.

Table 87 Attitude of the students towards science

Statement	Strongly agree	Partially agree	Partially disagree	Strongly disagree	Not stated	Total
Science helps in my daily work.	63	26	3	3	6	100
Science helps me to learn other subjects easily.	65	24	3	2	6	100
I like to learn science.	80	12	2	1	5	100
I should do better in science for my job.	75	15	2	2	6	100

Table 87 deals with the attitudes of the students towards science. As shown in the table, 80% of the students showed a positive attitude towards science, though only 63% accepted that science helps in their daily life.

Students' attitude towards science as a subject

Students' attitude towards science as a subject also plays an important role in achievement. Responses to the question of students' attitude towards science as a subject have been categorised and are presented in Table 88.

Table 88 Students' attitude towards science as a subject

Statement	Strongly agree	Partially agree	Partially disagree	Strongly disagree	Not stated	Total
I do better in science.	64	27	3	1	5	100
I like to learn science more.	79	14	2	1	5	100
Science is a joyful subject for me.	74	17	2	1	6	100
I can learn science quickly.	46	38	6	4	7	100
Science is difficult for me.	26	35	11	20	9	100

Table 88 indicates that 79% of students strongly agreed that they like to learn science; 74% of students agreed that science is a joyful subject for them. However, 36% of students agreed that science is a difficult subject.

Students' involvement in science activities

In NASA 2020, questions were asked enquiring about students' attitude towards different science activities, such as science activities with friends, solving science activities, using science equipment, doing homework, and asking the teacher difficult things. The responses were categorised and are presented in Table 89.

Table 89 Students' involvement in science activities

Statement	Almost all lessons	Some lessons	Never	Not stated	Total
We do exercises given in the textbook.	86	9	1	5	100
We do science activities with friends.	66	26	2	6	100
We solve science activities ourselves.	69	23	2	7	100
We use science equipment.	48	34	10	8	100
We start our homework in class.	36	29	26	9	100
We ask our teacher difficult things.	75	11	2	13	100

Table 89 displays that 86% of students strongly agreed that they do exercises given in the textbook; 75% of students agreed that they asked their teacher difficult questions. Only 48% of students expressed that they use science equipment. Overall, students had a positive attitude towards science as a subject and science activities.

Nepali Results

Introduction

In this section, the results of the responses of 21,578 students from 75 districts and 900 schools, who participated in NASA 2020 in Nepali language, taught as a core subject, are analysed. The results are presented in the form of proficiency levels, their description and comparison. Population estimates presented in this section are based on the ten plausible values drawn from WLE. The comparisons are based on groups formed from background information variables such as students' family background, socio-economic status, ethnicity, gender, home language, school types, home environment, province, etc.

In this assessment, IRT has been used for analysis, with 2PL used to analyse items of 1 point and GPCM used for questions with 2 or more points. TAM) was used to analyse the results.

The latent ability (theta) of each student is obtained from the analysis stated above. Ten plausible values were calculated to find out the ability of the population on the basis of the latent ability, and the dummy variables made on the basis of the variables asked in the students' background questionnaire. After transforming the average of those plausible values to zero (0), the following formula was used to present the results:

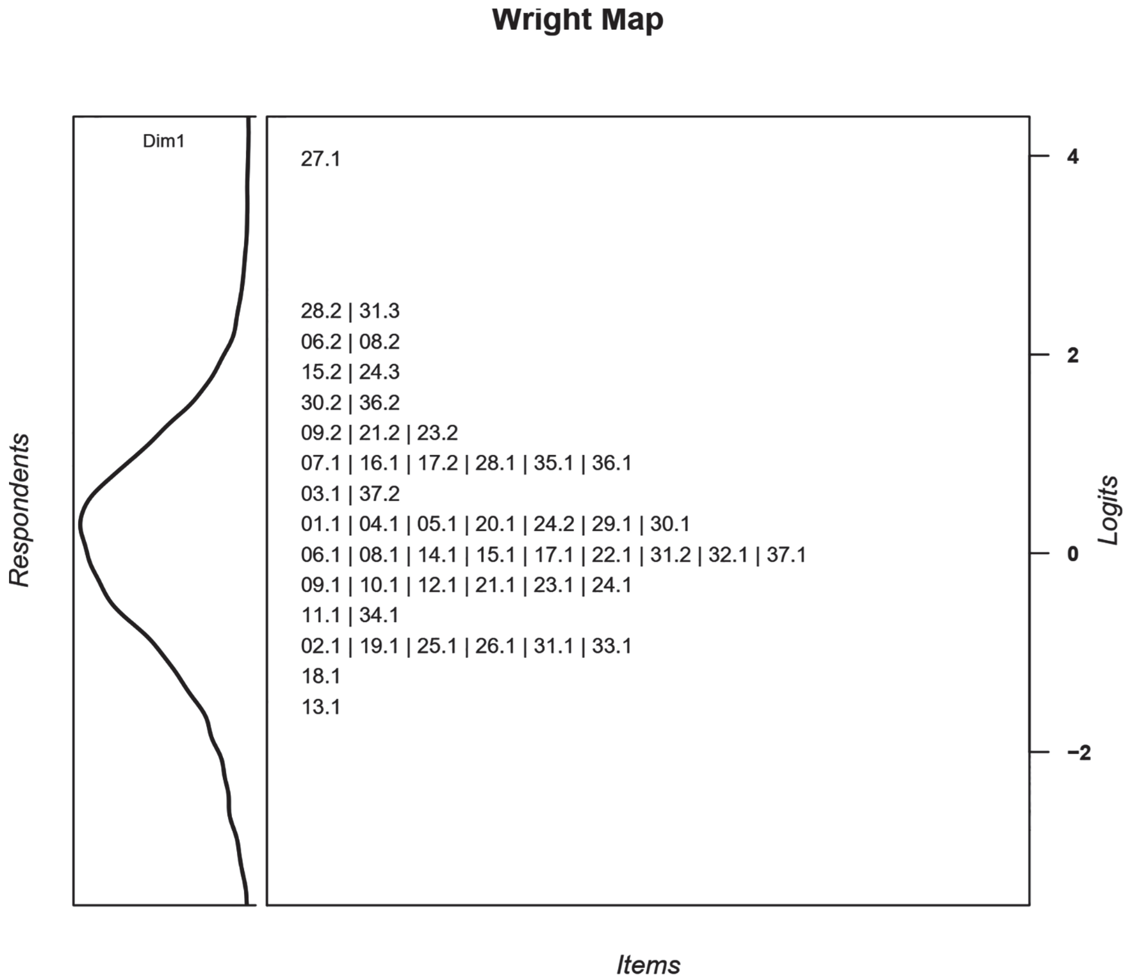
Student Achievement Score = 500 + PV × 50

Because the number of anchor items was insufficient and there were visible differential item functioning values, the results for Nepali language were not compared with the assessment of 2017, so the mean score is not compared with any other assessment. However, it can be compared with different variables within the year and the mean score becomes 500 at national level.

The Wright Map in Nepali

The Wright map is organised as a frequency curve and vertical histogram. The frequency curve on the left side shows the respondents, and the histogram on the right side shows the test items. The left side of the map shows the distribution of the measured ability of the respondents from most able at the top to least able at the bottom. The items on the right side of the map are distributed from the most difficult at the top to the least difficult at the bottom. In Figure 73, student ability (θ) on the left and NASA 2020 items to the right are plotted on the same scale. When a person and an item lie at the same level, the probability of that response by the particular respondent is 50%. Figure 72 presents the NASA 2020 Wright map in Nepali.

Figure 72 Wright map showing respondent and item on the same scale



The curve on the left side represents students; their latent ability is displayed in the logit scale ranging from -2 to +4. The distribution of students against the items administered (item numbers are shown to the right side) reveals that most of the items were difficult for the students. Although items were pre-tested and based on the grade 8 curriculum, most of the students were below the average latent ability '0'. This indicated that test items were difficult for the participating students. This further depicts that the performance level of grade 8 students in Nepali was not as expected by the curriculum.

Table 90 Mean score of plausible values in Nepali

Descriptive statistics					
	N	Minimum	Maximum	Mean	
				Statistic	Std. error
Mean of Nepali plausible values	21578	282.75	636.00	499.7723	.30992
Valid N (list-wise)	21578				

Descriptive statistics point to the fact that the national mean in Nepali is 499.7723 which can be rounded to 500 as the national average.

Note that some anchor (linking) items overlapped in NASA 2017 and NASA 2020. However, during the analysis, those anchor items showed large differential item functioning test values. The remaining usable anchor items were too few to compare results. Hence, this average score of 500 does not show the trend of achievement and is not compared with any previous assessments.

Student distribution in different proficiency levels in Nepali

The test scores were also analysed in terms of the six proficiency levels of students' achievement. The levels are Below basic, Basic, Proficient 1, Proficient 2, Proficient 3 and Advanced, from lowest to highest. The overall percentages of the students participating in the test are presented by proficiency level in Table 91.

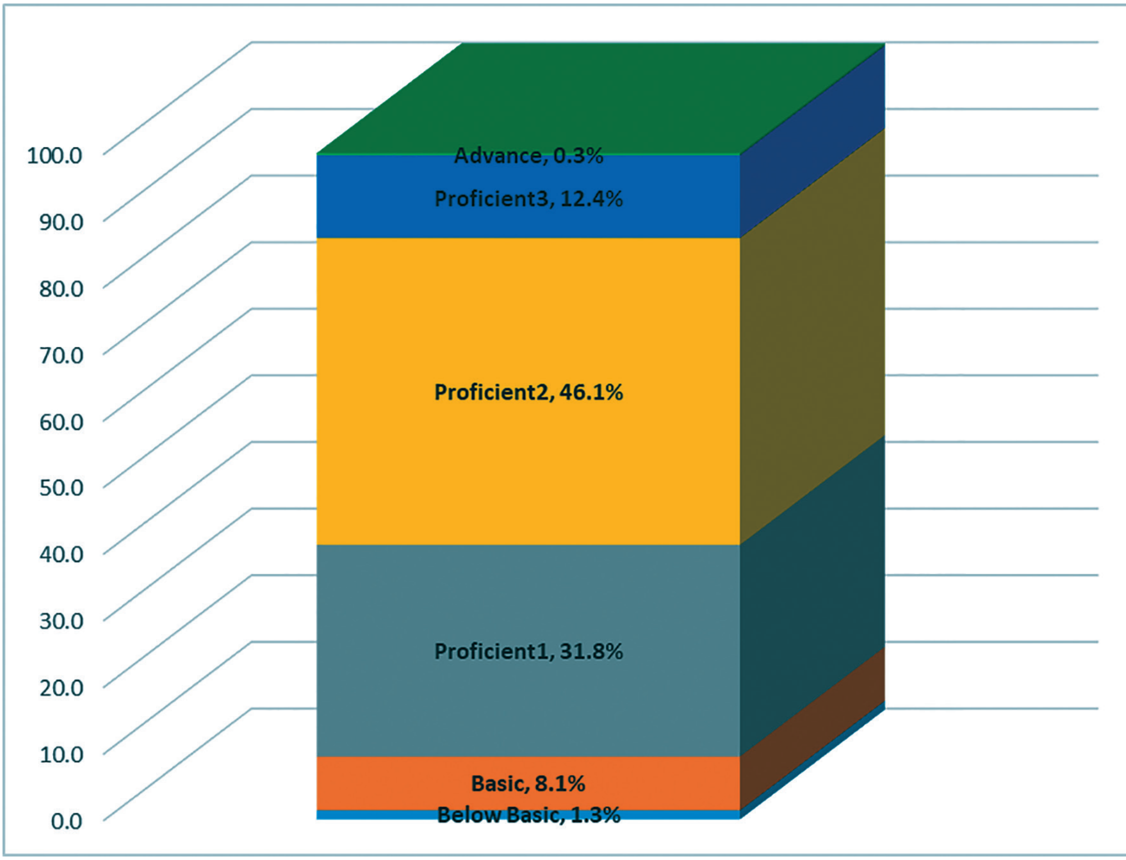
Table 91 Distribution of students at different levels of proficiency in Nepali

SN	Proficiency levels (ascending order)	N	Distribution N (%)
1	Below Basic	6041	1.3
2	Basic	36776	8.1
3	Proficient 1	144566	31.8
4	Proficient 2	209553	46.1
5	Proficient 3	56514	12.4
6	Advanced	1437	0.3
	Total	454887	100

Table 91 shows that of the 454,887 participants, 41.2% are Proficient 1 level or below, which means they do not achieve the minimum expected goals of the curriculum. Of the remaining students, 46.1% are in Proficient 2 level, 12.4 % in Proficient 3 level, while only 0.3% of students were placed at Advanced level. The results show that a low percentage of participants achieved the higher proficiency levels, which also indicates that the goals of the curriculum have not been achieved as expected.

The distribution of students across the proficiency levels in Nepali is presented in Figure 73.

Figure 73 Distribution of students across the proficiency levels in Nepali



The data in Figure 73 reveals that if students in level 3 or above are taken as achieving the minimum level of proficiency, a large proportion of students (41.2%) fall below this, with 58.8% of students having shown their performance to be above the minimum level of performance. The results indicate that learning opportunities in Nepali were not available to many students.

Proficiency level in Nepali by province

The distribution of grade 8 students across the proficiency levels varied by province in NASA 2020. Province-wise proficiency levels are presented in Table 92.

Table 92 Proficiency level by province

Provinces			Proficiency level						Total
			Below basic	Basic	Proficient 1	Proficient 2	Proficient 3	Advanced	
Province	Province1	Count	21	2742	19126	32455	10786	438	65568
		% within province	0.0%	4.2%	29.2%	49.5%	16.5%	0.7%	100.0%
	Mahesh	Count	5688	17305	26148	21309	3898	83	74431
		% within province	7.6%	23.2%	35.1%	28.6%	5.2%	0.1%	100.0%
	Bagmati	Count	41	2235	21658	41968	11282	138	77322
		% within province	0.1%	2.9%	28.0%	54.3%	14.6%	0.2%	100.0%
	Gandaki	Count	0	1082	12190	28710	12245	527	54754
		% within province	0.0%	2.0%	22.3%	52.4%	22.4%	1.0%	100.0%
	Lumbini	Count	146	5291	26516	33898	6435	21	72307
		% within province	0.2%	7.3%	36.7%	46.9%	8.9%	0.0%	100.0%
	Karnali	Count	83	5637	21951	21599	5050	124	54444
		% within province	0.2%	10.4%	40.3%	39.7%	9.3%	0.2%	100.0%
	Sudurpaschim	Count	62	2483	16978	29613	6817	107	56060
		% within province	0.1%	4.4%	30.3%	52.8%	12.2%	0.2%	100.0%
Total		Count	6041	36775	144567	209552	56513	1438	**454886**
		% within province	1.3%	8.1%	31.8%	46.1%	12.4%	0.3%	100.0%

Table 92 shows that Madhesh province has the highest percentage (7.6%) of students at Below basic level, while in Gandaki province 1% of students achieved Advanced level proficiency, which is the highest of all seven provinces. Gandaki province also scored 52.4% at Proficient 2 level and 22.4% at Proficient 3 level, which again are the highest levels of the seven provinces. In comparison to other provinces, Madhesh province achieved the lowest scores, as 65.9% of students achieved at Proficient 1 level or below, and only 28.6% of students were at Proficient 2 level and 5.2% at Proficient 3 level. Overall, the data indicates that Gandaki and Bagmati provinces are comparatively better than other provinces, though the expected results of the curriculum were not achieved.

Details of the province-wise distribution of participants across the proficiency levels are presented in Figure 74.

Figure 74 Distribution of students across the different levels of proficiency in Nepali by province

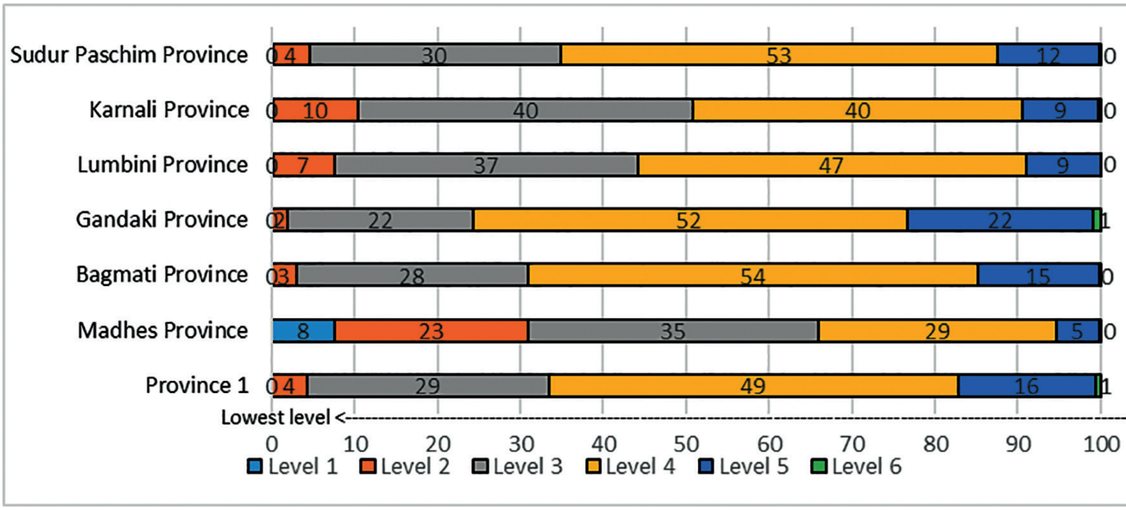


Figure 74 illustrates that 72% of students in Gandaki province passed the minimum level of proficiency whereas only 34% of students in Madhesh achieved this. Similarly, 49% of students in Karnali province passed the minimum level of proficiency. Students in Lumbini, Province 1 and Sudur Paschim, with 56%, 65% and 65% respectively, passed the minimum level of proficiency in NASA 2020. However, approximately 50% of students in Karnali province are at the basic level of proficiency or below, followed by Lumbini, where 44% have achieved only the basic proficiency level or below.

Proficiency by gender in Nepali

Based on the gender of the respondents, students were categorised into three sections – boys, girls and not specified – for uniform and proportionate results. Girls and boys should have equal opportunity and support in their studies; however, the results vary between them. To what extent this has been realised among students of Nepali in grade 8 has been an area of investigation. With gender as an implicit stratum, a comparison was made between the number of students in the six defined proficiency levels. The results indicate that girls scored higher than boys, which can be seen in the bar diagram in Figure 75.

Figure 75 Overall achievement in Nepali by gender

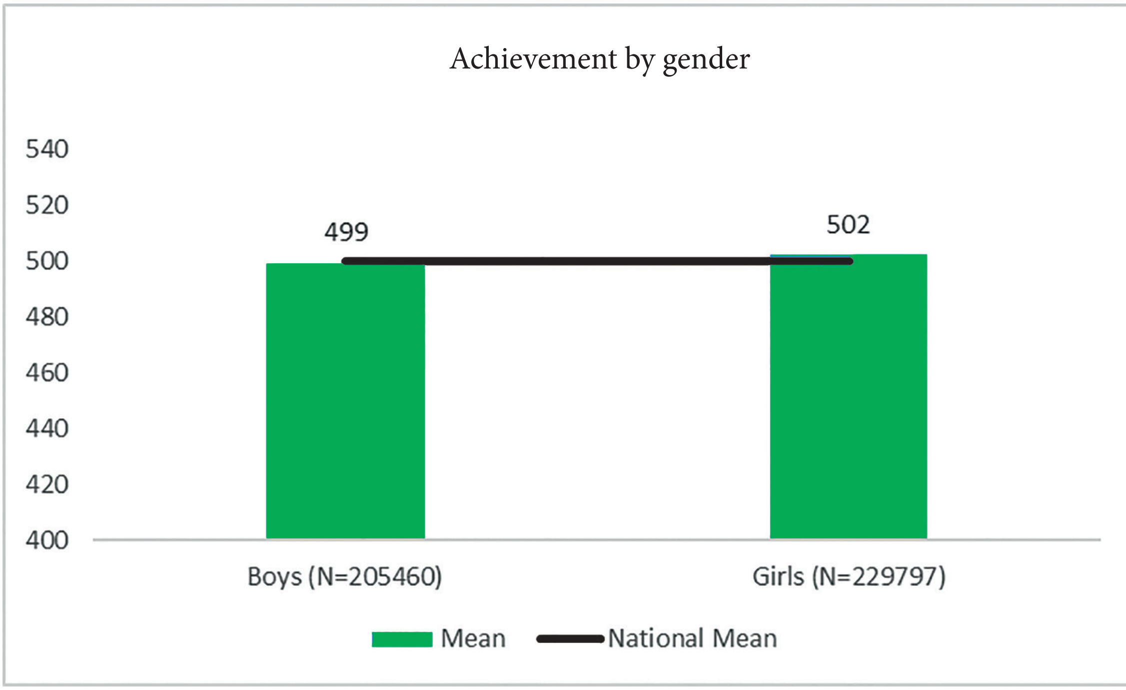


Figure 75 displays clear evidence that boys scored below the national average, whereas girls scored slightly above the national average. The results show that the score achieved by girls is 3 scale score higher than boys.

The distribution of students across the six performance levels by gender is presented in Table 93.

Table 93 Distribution of students across the different levels of proficiency in Nepali by gender

Gender	Proficiency levels						
	Below basic	Basic	Proficient 1	Proficient 2	Proficient 3	Advanced	
Boys	0.7%	7.7%	33.6%	46.6%	11.1%	0.2%	100.0%
Girls	1.6%	7.8%	29.5%	46.7%	14.0%	0.4%	100.0%
Total	1.2%	7.8%	31.4%	46.7%	12.6%	0.3%	100.0%

Table 93 presents that 40.4% of students are at Proficient 1 level or below, whereas 59.6% of students achieved at Proficient 2 level or above. In comparison to girls (38.9%), a higher percentage of boys (42%) scored at Proficient 1 level or below, while only 57.9% boys scored at Proficient 2 level and above. Meanwhile 61.1% of girls scored at Proficient 2 level and above. Thus, the results indicate that girls are comparatively better than boys. This result can be depicted as in Figure 76.

Figure 76 Distribution of students across the different levels of proficiency in Nepali by gender

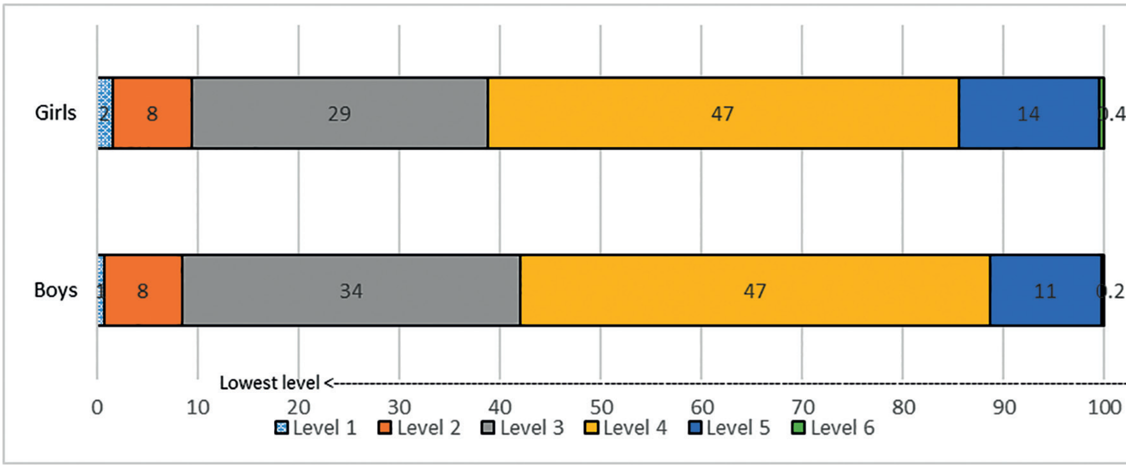


Figure 76 shows that 61.4% of girls have achieved above the minimum proficiency level, whereas 57.2% of boys have passed the minimum level of proficiency. Similarly, 39% of girls scored at the basic proficiency level or below, while 43% of boys scored at the basic level of proficiency or below. The results indicate that girls scored slightly higher than boys in NASA 2020.

Achievement by age in Nepali

In NASA 2020, students were asked to state their age as a background variable so that the relationship between age and educational achievement could be analysed. The data shows that students were grouped into five different age strata: 12 and below, 13, 14, 15, 16 and above. Students' learning achievements were found to differ as shown in Figure 77.

Figure 77 Achievement by age

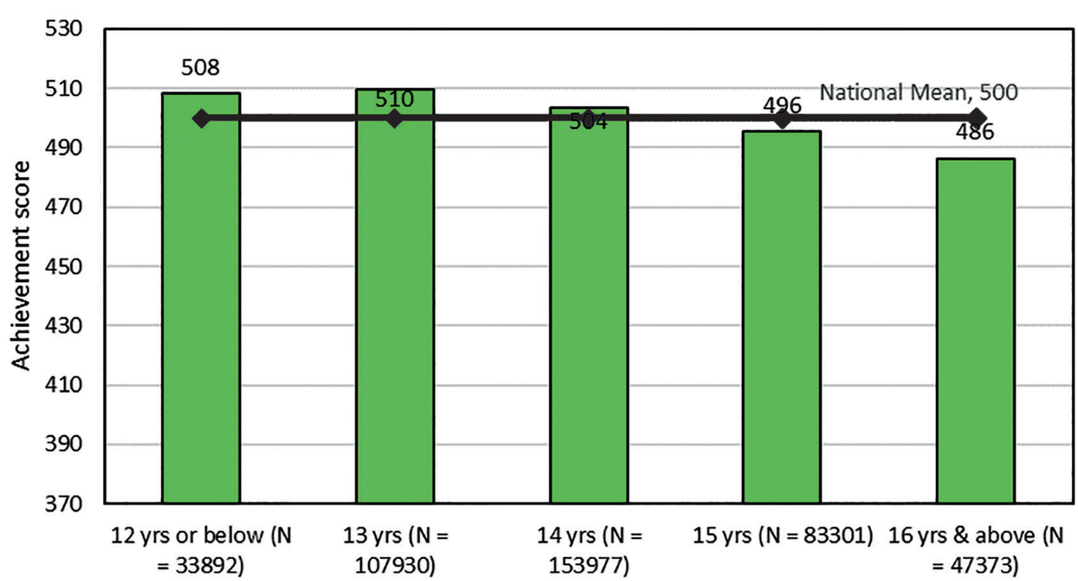


Figure 77 indicates that students of 13 years of age achieved the highest score (510), whereas students of age 16 and above scored the lowest (486), which is below the national average. Students of ages 12, 13 and 14 years scored 508, 510 and 504 respectively, which is above the national average, whereas students of ages 15 and 16 and above scored 496 and 486 respectively, which are below the national average. The results indicate that 13 years is the best age to study in grade 8, as students achieved the highest score.

Achievement in Nepali by home language

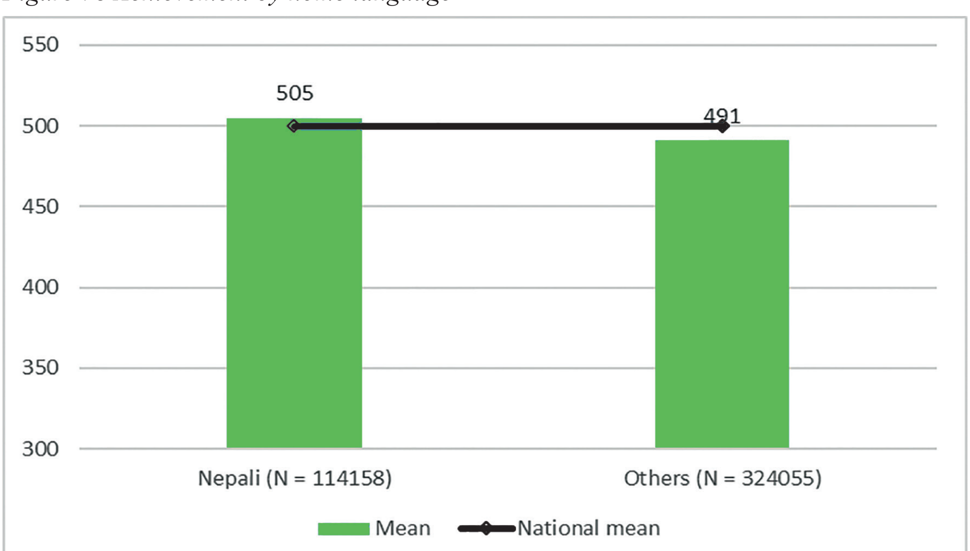
NASA 2020 included a question about students' home language as one of the issues looked into in relation to their achievement. Their responses disclosed that 73.94% of them spoke Nepali in their home, while 26.05% reported that they use other languages at home. Based on the students' responses, the achievement scores are presented in Table 94.

Table 94 Achievement by home language

Language spoken at home	N	Mean	Std. error of mean
Nepali	324055	505.2065	.07490
Others	114158	490.5211	.13819
Total	438213	501.3808	.06677

Table 94 shows that there is variation in the mean score in terms of students' linguistic background. Students with Nepali language spoken at home scored higher (505.2065) than the students who spoke other languages at home (490.5211). The results show that the medium of teaching and home language have a positive relationship with achievement, and are consistent with NASA findings (ERO, 2013; ERO 2015; 2016; ERO, 2018; ERO, 2019) that home language has shown an association with student learning outcomes. The results are displayed in Figure 78.

Figure 78 Achievement by home language



The bar diagram in Figure 78 reflects that those students whose home language is Nepali achieved 505, which is above the national average, and those students whose home language was other than Nepali achieved 491 scale score. The results show that there is a significant difference between the achievement of Nepali home language speakers and speakers of other languages at home in the NASA 2020 study.

Achievement based on students' time spent on different activities at home

In NASA 2020, students' engagement in different activities before and after school has seven categories: watch TV, play with friends and groups, household work, studying/homework, work for a wage, reading other books, and supporting siblings. These were measured from no time given to more than four hours. Table 95 provides the achievement of students based on their time spent at different activities.

Table 95 Achievement based on students' time spent on different activities

Time spent by students	No time given	Less than 1 hr	1–2 hr	2–3 hr	More than 4 hr
Watching TV/Internet/Mobile	488	506	508	499	478
Play with friend	488	507	504	493	486
Household chores	479	506	508	502	495
Study/homework	479	488	506	514	511
Work for wage	506	498	493	487	493
Reading other books	483	509	508	501	484
Support brother/sister to read	489	507	508	500	486

Table 95 shows that the achievement scores were found to be highest among those students in the habit of watching TV for 1–2 hours (mean = 508) or 2–3 hours (mean = 499), whereas the lowest scores were found among those who give more than 4 hours to watching TV (mean = 478) and not giving time to studying/homework (mean = 479). Overall, the results show that students spending 1–2 hours on any activities achieved higher scores, compared to those who spent 3–4 hours.

Achievement by use of leisure time in school

The NASA 2020 study asked students a question regarding their use of leisure time in school. The variables were categorised as classwork, homework, play, return home and not specified. Out of a total number of 454,887 students, 256,110 (56%) of them used their time to do classwork, whereas 9,831 (2%) of students returned home without engaging in any activities. The details are shown in Table 96.

Table 96 Achievement by use of school leisure time

Leisure-time activity	Mean	N	Std. deviation	Std. error of mean
Classwork	504.4136	256110	43.91363	0.08677
Homework	501.1061	145320	41.09753	0.10781
Play	487.7276	28272	46.56674	0.27695
Return home	449.6883	9831	45.16689	0.45554
Not specified	470.7222	15354	67.90184	0.54798
Total	500.0000	454887	45.43273	0.06736

Table 96 indicates that those students who spent their leisure time in classwork (mean = 504, SD = 43.91) and homework (mean = 501, SD = 41.09) have better performance in Nepali, while those who return home have the lowest performance (mean = 449, SD = 45.16). This data is also presented in diagram form in Figure 79.

Figure 79 Achievement by use of leisure time in school

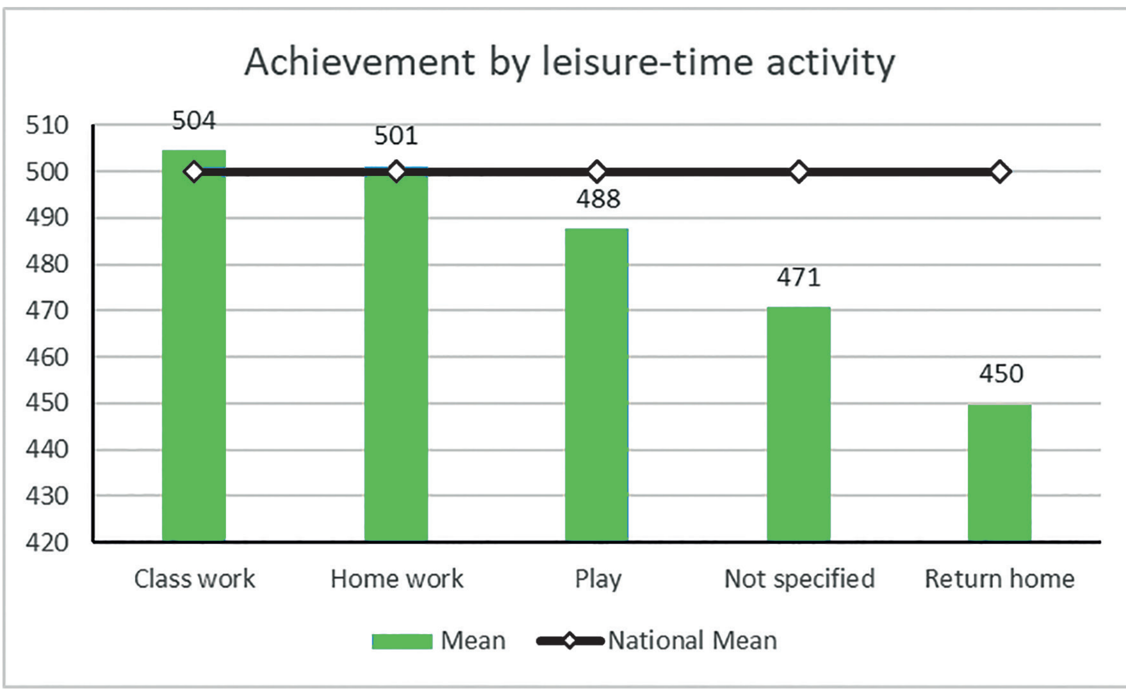


Figure 79 illustrates that students who engaged themselves in classwork achieved the highest score (504), whereas those who returned home scored 450 scale score in Nepali. Overall results show that students using their leisure time for classwork and homework achieved higher than the national score, whereas students who used their leisure time to play, to return home or who did not specify achieved below the national average in Nepali. The results indicate that being involved in classwork or doing homework are the best uses of students' leisure time.

Achievement and caste/ethnicity

In NASA 2020, one of the concerns has been the relationship between students' ethnicity and their achievement. The data has looked at the achievement of students who have different ethnic backgrounds. Broadly based on the geographical belts, i.e. Mountain, Hill and Madhesh, three important ethnicities have been considered for study. There are Brahmin/Chhetri, Janajati and Dalit. Students with ethnicities not falling into these three categories have been recognised here as 'Other'. Details of the ethnic categories and their mean scores are presented in Table 97.

Table 97 Achievement and caste/ethnicity

Ethnic group	N	Mean	Std. Error of Mean
Madhesi Brahman	37167	495.6296	.25471
Madhesi Janajati	55579	492.5796	.18701
Madhesi Dalit	14481	484.1788	.38465
Madhesi Others	31714	487.3964	.26240
Hill Brahman	121238	514.3009	.11218
Hill Janajati	82348	510.5414	.12721
Hill Dalit	30295	504.5823	.22012
Hill Others	15884	505.3397	.28898
Mountain Brahman	6535	495.9465	.62072
Mountain Janajati	4040	501.5532	.68898
Mountain Dalit	1292	492.4728	1.55374
Mountain Others	1337	472.2243	1.21441
Total	401910	503.8684	.06730

Table 97 depicts that there were 37,167 Brahmnn/Chhetri students, 55,579 Janajati students and 14,481 Dalits in Madhesh who participated in the study as samples. Their mean scores are 495.6296, 492.5796 and 484.1788 respectively. These results show that Madhesh Brahman/Chhetri have the highest mean score and Dalits have the lowest mean score of achievement. In the Hill region, the student numbers were Brahman 121,238, Janajati 82,348, Dalit 30,295 and Others 15,884. The mean scores of Hill Brahman/Chhetri, Janajati, Dalit and Other were 514.3009, 510.5414, 504.5823 and 505.3397 respectively. The results also show that Brahman/Chhetri has the highest score, whereas Dalit has the lowest mean score. In the Mountain region, the number of Brahman/Chhetri was 6,535 and their mean score was 495.9465; the number of Janajati was 4,040 and their mean score was 501.5532; the number of Dalit was 1,292 and their mean score was 492.4728, and the number of Others was 1,337 and their mean score was 472.2243. These results also depicted that Brahman/Chhetri had the highest mean score and Dalits had the lowest mean score.

Overall, the results showed that Brahman/Chhetri had the highest mean scores, and Dalits had the lowest mean scores, in all geographical regions. Hill Brahmans had the overall highest mean score of 514.3009 and Madhesh Dalit had the lowest mean score (484.1788).

Achievement in Nepali by students' family size

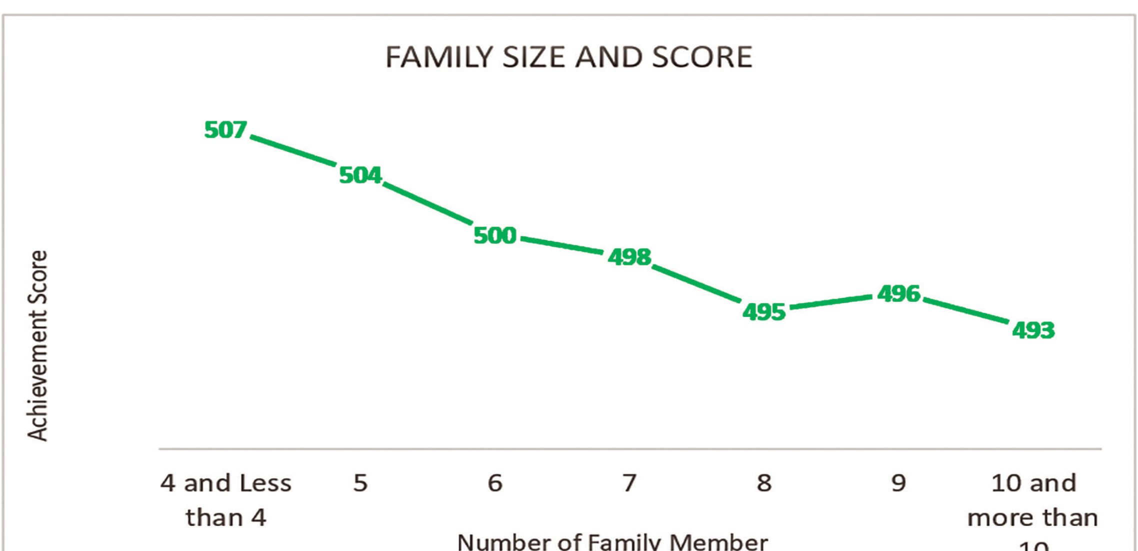
Family size is one of the factors that has a direct relation in the care, health, opportunity, facilities and study of children. Students were asked the number of members in their family, and the children were grouped into 4 or less, 5, 6, 7, 8, 9, and 10 or more family members. The achievement of students based on their family number is presented in Table 98.

Table 98 Achievement by family size

Number of family members	Mean	N	Std. deviation	Std. error of mean
4 or less	507.4584	119067	42.11057	0.12204
5	504.2425	106474	43.09624	0.13207
6	500.0048	79798	43.43908	0.15377
7	498.4563	45948	45.32729	0.21146
8	494.6103	26963	46.8552	0.28535
9	495.9035	15039	44.19299	0.36037
10 or more	493.3914	41148	46.55437	0.2295
Total	501.8193	434437	44.0152	0.06678

Table 98 shows that 119,067 students responded that their family size was 4 or less members, and their achievement was the highest of all (507); 41,148 students stated that their family size was 10 or more members, and they achieved a score of 493 which is the lowest of all. Overall, the results illustrate that students with 4, 5 or 6 family members achieved higher than or equal to the national average, whereas students with family sizes of 7, 8, 9 and 10 or more members achieved below the national average. This is presented as a line graph in Figure 80, where the x-axis represents the size of the family and the y-axis shows the achievement scale. The line indicates the achievement score of children of different family sizes.

Figure 80 Achievement by family size



On the basis of family size, the majority of students had five family members, and a few of them had twelve family members. Students with four or less family members achieved the highest score (507), and five family members scored 504. When the number of family increased, the average achievement score decreased. This is a negative association between family size and the student learning. This clearly shows that small families can give a proper care environment and opportunities for study to children. When the family size increases, the students' quality decreases.

Achievement by parents' education

In the 2020 NASA study, one of the key concerns has been to see the extent to which family-related factors, including parental education and occupation, are associated with students' learning achievement. The data has been studied from this angle for students' achievement in Nepali as well. The association between parental education and students' achievement is presented first, with the role of parental occupation presented thereafter.

In looking at the role of parental education in students' achievement, the role of mother's education and father's education in students' achievement was analysed separately.

Achievement by mother's education level

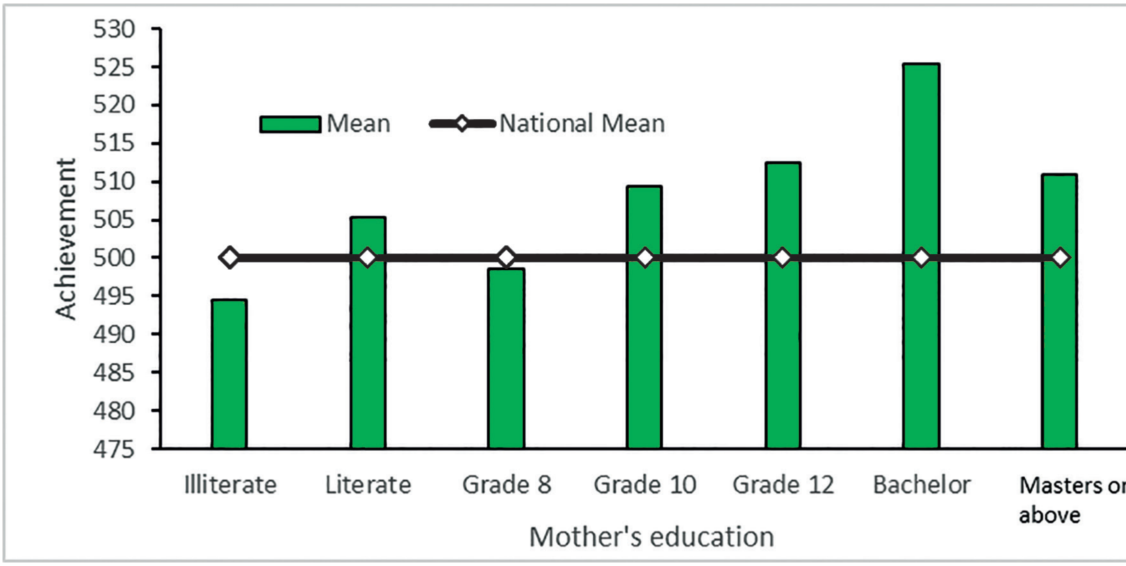
The educational status of a student's mother has a bearing on the learning achievement of that student. Of the 438,420 mothers sampled in NASA 2020, 142,241 were illiterate, while 112,945 were literate. The sample were categorised as grade 8, grade 10, grade 12, Bachelor's, and Master's or above, and these levels numbered 83,128, 53,105, 30,853, 10,610 and 5,538 respectively. Table 100 displays the mean score of students by mother's education level.

Table 99 Achievement by mother's education level

Mother's education level	N	Mean	Std. error of mean
Illiterate	142241	494.5108	.11301
Literate	112945	505.2435	.11773
Grade 8	83128	498.6029	.14908
Grade 10	53105	509.2678	.19337
Grade 12	30853	512.4622	.26579
Bachelor's	10610	525.3623	.45606
Master's or above	5538	510.8511	.64298
Total	438420	502.0554	.06532

Table 99 reveals the variation in students' achievement based on the educational level of mothers. The data shows that students whose mothers are illiterate achieved the lowest score (494.5108), whereas the achievement of students whose mothers are literate up to grade 12 increases gradually. The highest score (525.3623) was achieved by students whose mother had a Bachelor's degree, with students whose mother had a Master's degree or above scoring 510.8511. The data is shown as a bar diagram in Figure 81.

Figure 81 Achievement by mother's education level



Achievement by father's education

The achievement by father's education is presented in table 100.

Table 100 Achievement by father's education level

Father's education level	N	Mean	Std. error of mean
Illiterate	61954	488.0217	.17818
Literate	87184	504.4231	.13344
Grade 8	108180	496.5655	.13019
Grade 10	98550	502.3075	.13831
Grade 12	50807	510.7804	.18882
Bachelor's	21753	519.4928	.29503
Master's or above	11251	526.1279	.41184
Total	439680	501.7401	.06548

The results is also presented in the bar graph in figure 82.

Figure 82 Achievement by father's education level

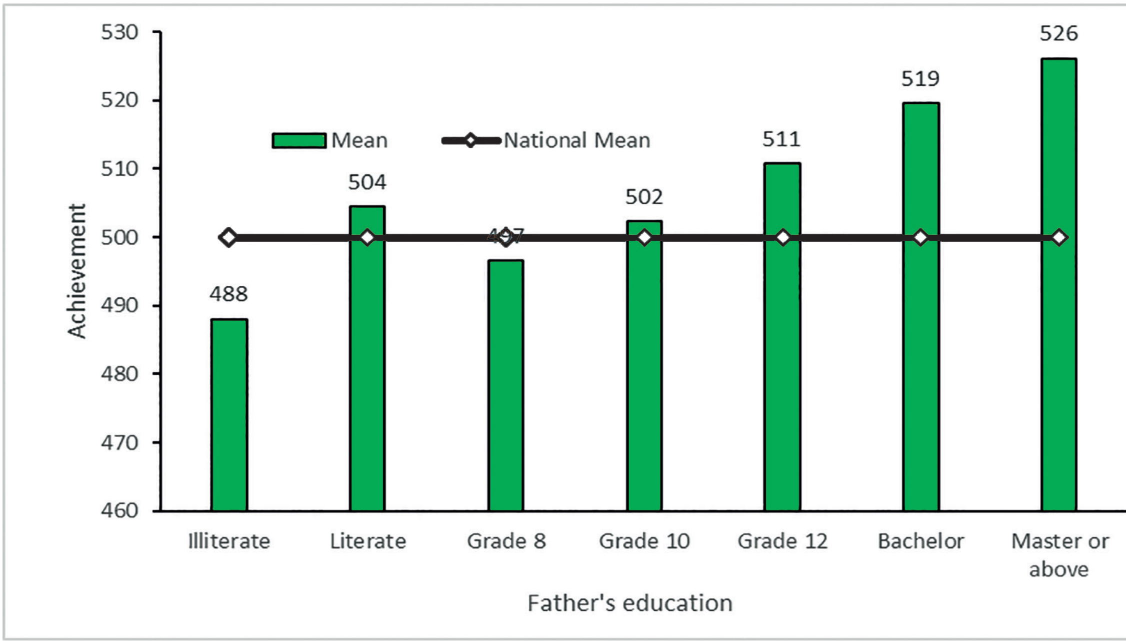


Figure 82 shows that the highest achievement score was 526, and the lowest was 488. It is clear that the achievement score decreases as the level of the father's education decreases. With this descending hierarchy, the lowest achievement is seen in students whose fathers are illiterate, as was the case with mothers. This tendency has been applicable to the students in any of the categories.

Parental occupation

NASA 2020 also analyses students' learning achievement from the perspective of parental occupation. Consequently, the achievement of students having mothers engaged in various occupations was studied. In the same way, the association between father's occupation and students' achievement has also been analysed.

Achievement by mother's occupation

Mother's occupation is an important consideration in relation to students' learning achievement in the NASA 2020 study, including in relation to Nepali as a subject. For this purpose, nine different categories of mother's occupation were established for data analysis: Agriculture (262,975 samples), Household work (96,559 samples), Work in other's house (5,071 samples), Labourer (5,505 samples), Foreign employment (7,990 samples), Teaching (12,059 samples), Business(31,043 samples), Government service (8,812 samples), and Other (9,050 samples). According to the responses given by sample students, the number of students having mothers working in these various categories is indicated in Table 101.

Table 101 Achievement by mother's occupation

Mother's occupation	N	Mean	Std. error of mean
Agriculture + household	262975	498.3774	.08374
Household	96559	502.6147	.14069
Work in other home	5071	494.7146	.68450
Labour	5505	503.8391	.54979
Foreign employment	7990	506.9841	.51233
Teaching	12059	524.9976	.39845
Business	31043	512.8840	.21806
Government job	8812	514.4833	.46155
Others	9050	512.7386	.44648
Total	439064	501.8681	.06545

Table 101 shows that the mean score of students in Nepali varies as the mothers' occupations differ. Students whose mothers work in others' homes achieved the lowest score (495), whereas those whose mothers work as teachers achieved the highest score (525). Students whose mothers work in agriculture and housework, household work, labour, foreign employment, business, government and other work achieved 498, 503, 504, 507, 513, 514 and 513 respectively in round figures. Students whose mothers are engaged in government jobs are in the second highest position, with a score of 514. This data is shown in Figure 83.

Figure 83 Achievement by mother's occupation

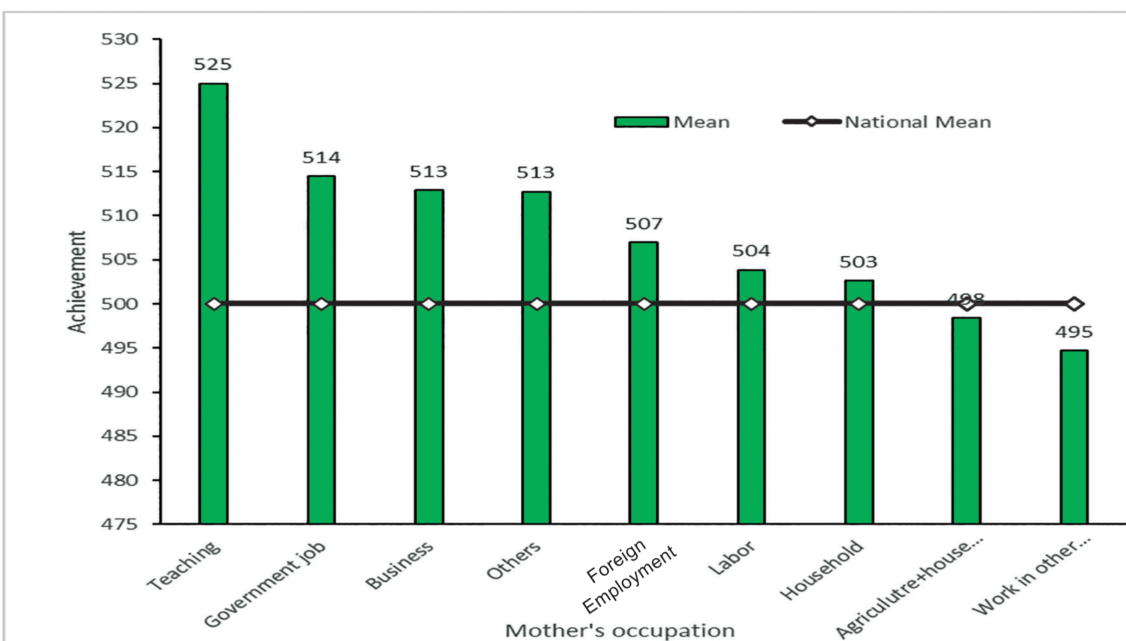


Figure 83 shows that mother's occupation has some association with students' achievement in Nepali as well, with the data clearly depicting that children whose mothers work in others' homes have the poorest achievement, the achievement of children whose mothers work in business or government job looks encouraging, and students having mothers in teaching have the highest achievement score of all.

Achievement by father's occupation

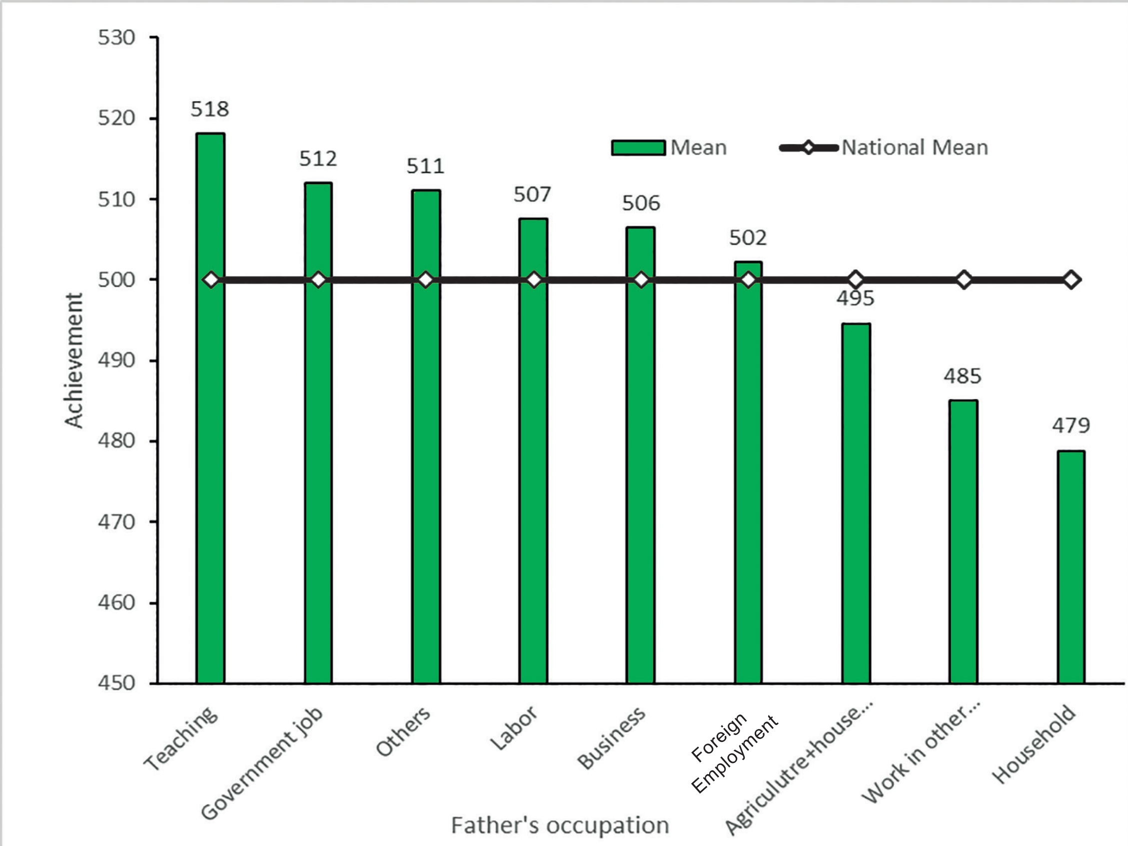
In NASA 2020, data on father's occupation has also been studied in relation to students' learning achievement to see how far this is applicable in Nepali subject. Using the same frame of reference as for mother's occupation, students were grouped into nine different categories on the basis of their father's occupation. The results are presented in Table 102.

Table 102 Achievement by father's occupation

Father's occupation	N	Mean	Std. error of mean
Agriculture + household	130001	494.5917	.12265
Household	10962	478.8157	.42951
Work in other home	9063	485.0791	.43945
Labour	33620	507.4919	.21553
Foreign employment	104663	502.1244	.13147
Teaching	14296	518.0918	.37754
Business	65455	506.4851	.16389
Government job	32528	511.9535	.22484
Others	32238	511.0435	.23275
Total	432827	501.9214	.06598

The data in Table 102 shows that the achievement score varies as per the variation in father's profession. For instance, the achievement of the students whose parents were involved in household work only was the lowest (479) and that of students whose parents were in the teaching profession was the highest (518). The rounded figures of the achievement by the father's occupation are presented in figure 84.

Figure 84 Achievement by father's occupation



Achievement by possession of a separate mobile phone

In the NASA 2020 study, one of the questions asked was whether students possessed a separate mobile phone, and the data has also been studied to see if owning a mobile phone has a bearing on students' learning achievement in Nepali subject. Of the 427,798 students surveyed, 333,945 of them stated that they did not possess a mobile phone, and only 93,853 of them stated that they possessed a mobile. The data is shown in Table 103.

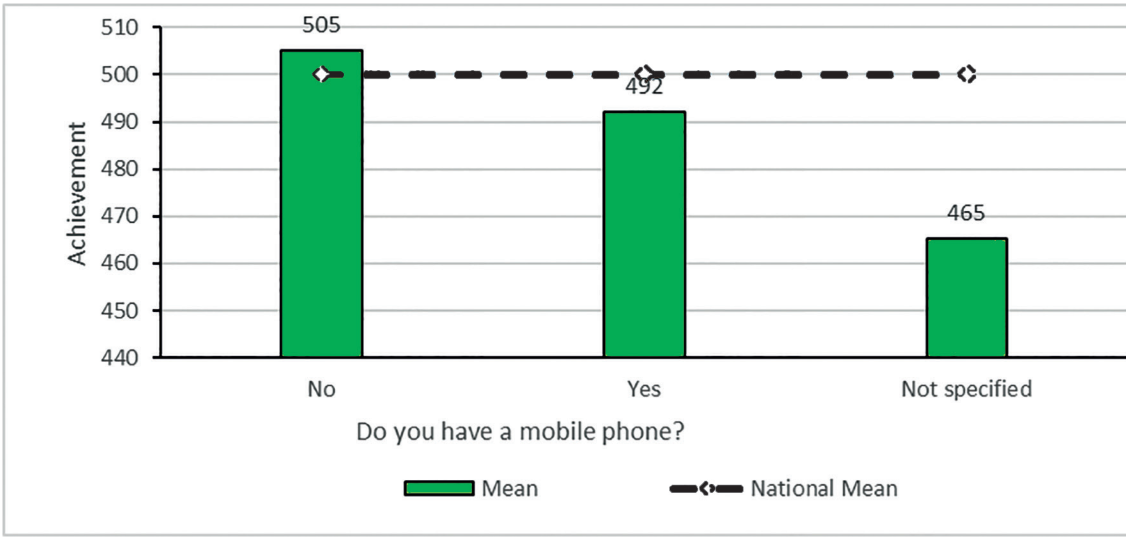
Table 103 Achievement by possession of a separate mobile phone

Do you have a mobile phone?	N	Mean	Std. error of mean
No	333945	505.0621	.07363
Yes	93853	492.0528	.15279
Total	427798	502.2081	.06704

Note: Non-responses are not included in the analysis in this table.

Table 103 shows that those students who did not possess a mobile phone scored more highly (505) than students who possess a mobile phone (492). Overall, the average achievement score was 502. The findings indicate that students who possessed a mobile phone achieved lower in Nepali than those who did not possess it. This data is presented in bar diagram form in Figure 85.

Figure 85 Achievement by possession of a separate mobile phone



The data in Figure 85 reveals that students who possessed a mobile phone achieved below the national average (492), while those who did not possess a separate mobile phone achieved above the national average (505), . The results indicate that owning a separate mobile phone does not mean that students progress in achievement.

Achievement by availability of Nepali textbook

Availability of textbooks is one of the reasons for better achievement in the study of NASA 2020. The results of how far possessing a textbook is applicable in relation to students' learning achievement in Nepali subject are shown in Table 104.

Table 104 Achievement by availability of Nepali textbook

Do you have a Nepali textbook ?	N	Mean	Std. error of mean
Yes	415783	502.5752	.06793
No	32070	479.5815	.26071
Total	447853	500.9287	.06636

Table 104 presents data revealing that 415,783 students had a Nepali textbook, whereas 32,070 students did not possess a Nepali textbook. The mean score of the students who possessed a Nepali textbook is 502, while those who did not possess a Nepali textbook scored 479. The overall achievement of students possessing Nepali textbooks exceeds that of those who did not possess them, providing evidence that timely availability of textbooks results in better achievement.

Achievement by regularity of homework and feedback

Homework also plays a critical role in the students' achievement. In NASA 2020, the data showed that 79% of students responded that their teacher provided regular homework, whereas 19.7% of students agreed that they were given homework only sometimes. Nearly 1% of students never got homework from their teacher. Table 105 shows the mean scores of students in relation to how often they received homework from their teacher.

Table 105 Achievement in Nepali by regularity of being given homework

Does your teacher give you homework ?	N	Mean	Std. error of mean
Always	354693	502.4667	.07304
Sometimes	87780	497.6860	.15139
Never	4069	463.1063	.91877
Total	446543	501.1683	.06603

Table 105 demonstrates that students who received homework regularly obtained a score of 502.56, whereas students who received homework from their teacher only sometimes scored 497.68. Meanwhile, students whose teacher never gave them homework scored 463.10. This evidence proves that giving homework has a direct relationship with achievement. Overall, results show that regular homework results in better achievement.

Achievement by teachers' feedback

The NASA 2020 study investigated feedback on homework from teachers and its relationship with achievement. Among the sampled respondents, about 72.25% of students agreed that they received regular feedback from their teachers, whereas about 21.58% of students stated that they received feedback from their teachers only sometimes. The data in Table 106 reveals the variations in achievement between students who received feedback all the time and those who received feedback only sometimes.

Table 106 Achievement by regularity of feedback on homework

Feedback from teacher	N	Mean	Std. error of mean
Always	343332	503.2128	.07369
Sometime	95905	496.8883	.14453
Never	5172	469.1285	.76069
Total	444409	501.4513	.06584

Table 106 shows that students who always received feedback on their homework scored 503, whereas students who received feedback only sometimes scored 496. Those students who never received feedback on their homework scored only 469. Overall, the data in Table 107 illustrates that regular feedback from teachers on students' homework has a positive effect on their achievement.

Achievement by type of reference book used for Nepali

In the NASA 2020 study, one of the areas of enquiry was availability of reference books to students.

In the study, 69.43% responded that they had the old question set, whereas 8.13% of students had the guess paper. Meanwhile, 22.42% of students expressed that they possess study guides. The mean score students achieved in NASA 2020 is shown in Table 107.

Table 107 Achievement by type of reference book used for Nepali

Do you have a reference book for Nepali ?	N	Mean	Std. error of mean
Old question set	289447	503.9990	.08109
Guess paper	33930	485.3266	.25106
Guide	93493	497.2747	.13942
Total	416870	500.9711	.06809

Table 107 indicates that students who possessed the old question set scored 503.99, while students with a guide scored 497.27. Meanwhile students possessing guess paper achieved 485.33. The results prove that the best reference materials for students are the old question sets. Compared to the guess paper, guides are better reference materials for students as they achieved 12 scale score more than those who had guess paper.

Achievement by support provided at home for reading/writing

Parental help to read and write at home is another area of the NASA 2020 study. To the question of who helped students to read and write at home, 16% of respondents replied that their father helped them, while 7.33% agreed that their mother helped. The majority of students (50.65%) responded that their brothers and sisters helped them to read and write. Alongside this, 7.58% of students were supported by friends, 10.05% by tutors, 3.19% were supported by others, and 4.65% responded that they were supported by nobody at home. The differences in achievement according to the support received are presented in Table 108.

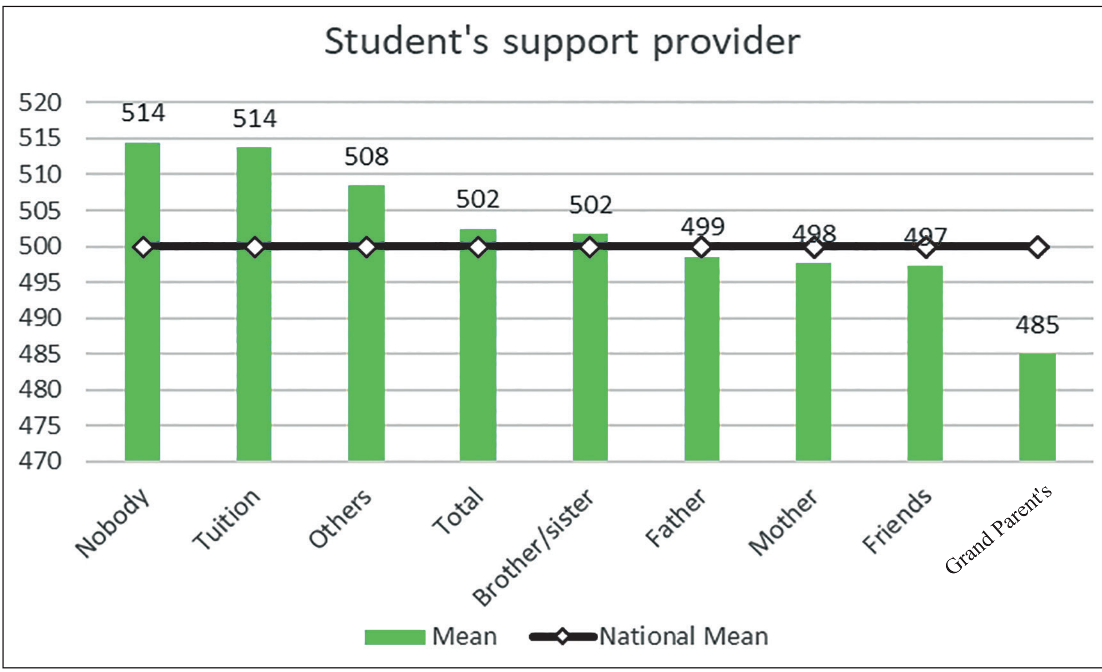
Table 108 Achievement by support provider at home for reading/writing

Support provider at home for reading/writing	N	Mean	Std. error of mean
Father	63970	498.5411	.18044
Mother	29343	497.5799	.26501
Brother/sister	202541	501.7034	.09077
Grandmother/Grandfather	2561	484.9957	.93218
Friends	30306	497.2430	.24467
Others	12766	508.4332	.38181
Tuition	39777	513.6772	.20729
Nobody	18613	514.4115	.30385
Total	399876	502.4473	.06746

Table 108 shows an interesting result. The mean scores of students who were supported by their father, mother and friends, were 498.54, 497.57 and 497.24 respectively. Students supported by tutors achieved 513.6. However, 514.41 was the mean score of students whom nobody supported at home.

The lowest mean score was 484.99, achieved by students who were supported by their grandfather or grandmother. Overall, the results show that family support has no effect for exceptional students, as the highest score was achieved by students who received no support by either family members or tutors. The results are presented in bar diagram form in Figure 86.

Figure 86 Achievement by support provider at home for reading/writing



The bar diagram in Figure 86 illustrates that students who receive support provided by tutors achieved the highest score in Nepali (514), whereas students supported by their grandparents scored the lowest (485). The results show that students who received support from tutors, brothers and sisters, and others scored above the national average, whereas students getting support from their father, mother or friends scored below the national average. Interestingly, students who received no support from anybody scored highest of all (514) which indicates that student's own efforts are more valuable in their success compared to other's support.

Achievement by future career goal

The future aims of students was another study area of NASA 2020. The possible future aims of students were categorised as teacher, government job, private job, business, work abroad, farmer, doctor/engineer and other. In total, 434,177 responses to the question of future aim were received, and the mean score was 502.33, which was slightly above the national mean. The details of the mean scores are presented in Table 109.

Table 109 Achievement by future career goal

Your future aim	N	Mean	Std. error of mean
Teacher	104440	489.0072	.13242
Government job	58043	506.2185	.16845
Private job	5196	490.9150	.61828
Business	17153	495.3250	.29204
Work abroad	14068	484.5736	.34545
Farmer	11961	480.4875	.41468
Doctor/Engineer	153577	511.0286	.10884
Other	69740	509.8295	.14830
Total	434177	502.3361	.06529

Table 109 shows that students whose future aim was to be doctor/engineer scored highest (511.02), and students whose aim was to be a farmer scored lowest (480.48). Students whose aim was to work in government sector scored 506.21, which was slightly higher than for students whose future aim was teaching (489.00), a private sector job (490.91), business (495.32) and working abroad (484.57). Overall, the results show that students whose future aim was more challenging, such as becoming a doctor/engineer or working in government, scored higher than for other jobs. These results can be seen as a bar diagram in Figure 87.

Figure 87 Achievement by future career goal

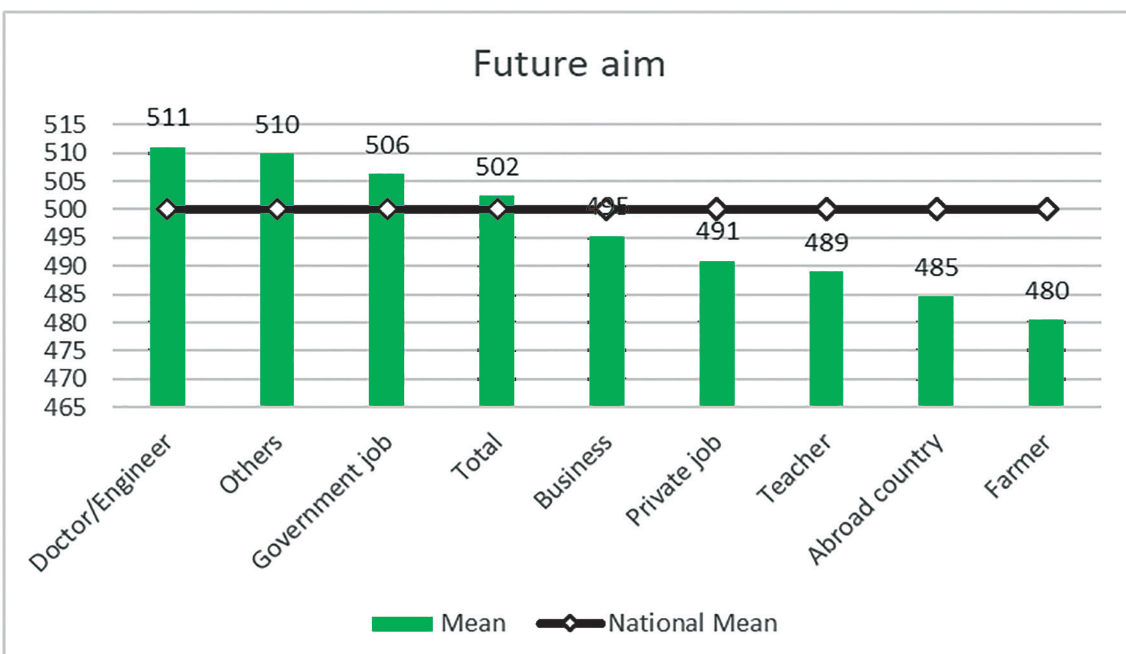


Figure 87 shows that students who wished to be a doctor/engineer scored the highest mean (511), whereas those respondents whose future aim is to be a farmer obtained the lowest score (480). Similarly, students whose intended career aims were to be a doctor/ engineer, other jobs and government jobs achieved above the national average, while those wishing to be in business, to work in private jobs, to be teachers, or to work abroad achieved below the national average.

Achievement by access to social media

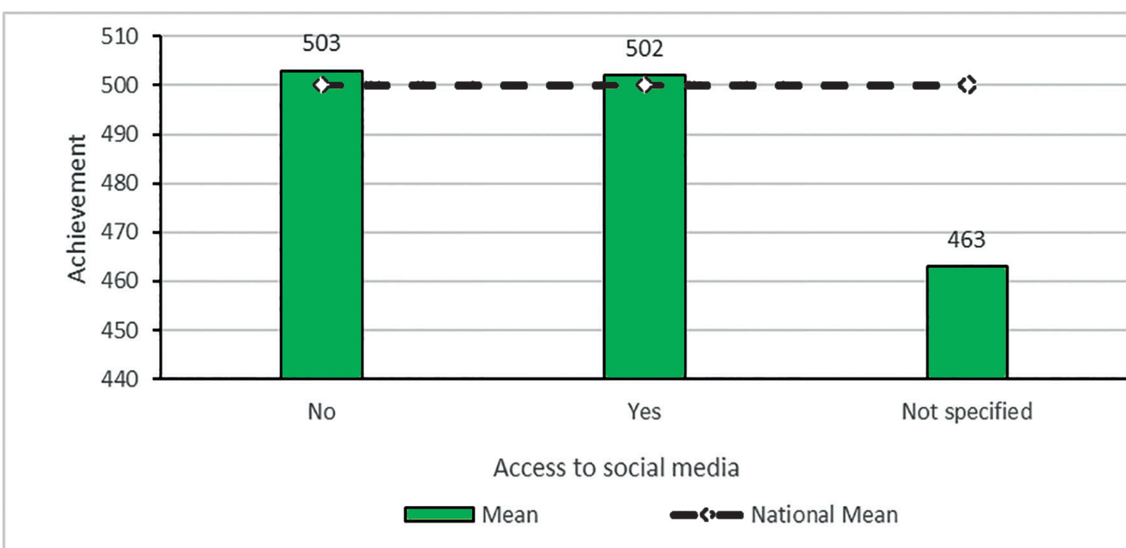
One of the questions students were asked was regarding access to social media. The details are presented in Table 110.

Table 110 Achievement by access to social media

Do you have access to social media?	Mean	N	Std. deviation	Std. error of mean
Yes	502	101416	46.37081	0.14561
No	503	323076	42.4925	0.07476
Not stated	463	30395	55.116	0.31614
Total	500	454887	45.43273	0.06736

Table 110 shows that 22% of students indicated that they had access to social media, while 71% of students did not. The results show that students who had access to social media achieved a score of 502, while the mean score of those who stated that they did not have access to social media was 503. The results indicate that there is no significant difference in achievement in Nepali between students having access to social media and not. This is shown as a bar diagram in Figure 88.

Figure 88 Achievement by access to social media



The data in figure 88 shows that those students who had social media scored 502, while those who did not scored 503. Both scores are above the national average. The results indicated that there is no significant difference in achievement between having access to social media or not.

Achievement by attitude towards teacher

Students' attitude towards teachers' behaviour was also explored in NASA 2020. The results are presented in Table 111 as percentages.

Table 111 Students' Attitude towards teacher (N percentage)

Statements	Totally agree	Somewhat agree	Somewhat disagree	Totally disagree	Not specified	Agree	Disagree
Teacher teaches with love and care.	81.8	12	1.2	1.1	3.9	94	2
Most of the teachers really listen to me.	47.7	33.7	5.4	3.7	9.6	81	9
Teachers do not give corporal punishment.	35.4	26.2	12.2	14.5	11.7	62	27
Teachers treat us equally and care about us.	79.2	9.1	2.6	2.2	6.8	88	5
Teachers answer when we ask when we don't understand.	87	5.1	1	1.1	5.8	92	2
Teachers give us homework.	84.3	6.6	1.1	1.5	6.5	91	3
Teachers give and check homework.	79.0	10.5	1.9	1.5	7.0	90	3
Teachers teach for the full time of the period.	65.2	18.9	4	2.5	9.3	84	7

To simplify the results shown in Table 111, the four categories were reduced to two, with the first two merging into AGREE and the second two merging into DISAGREE. The results in Table 112 are clearer.

Table 112 Students' Attitude towards teacher (N percentage)

Statements	Agree	Disagree	Not specified
Teacher teaches with love and care.	94	2	4
Most of the teachers really listen to me.	81	9	10
Teachers do not give corporal punishment.	62	27	11
Teachers treat us equally and care about us.	88	5	7
Teachers answer when we ask when we don't understand.	92	2	6
Teachers give us homework.	91	3	6
Teachers give and check homework.	90	3	7
Teachers teach for the full time of the period.	84	7	9

Table 112 indicates that 94% of students agreed that their teachers teach them with love and care, 91% agreed that their teacher gives homework and, similarly, 90% agreed that their teachers check homework. Overall, the results show that students have a positive attitude towards teachers of Nepali.

English Results

Introduction

In this section, the results of the responses of 21,683 students from 75 districts and 900 schools, who participated in NASA 2020 in English language, are analysed. The results are presented in the form of proficiency levels, their description and comparison. Population estimates presented in this section are based on the ten plausible values drawn from WLE. The comparisons are made on the basis of groups formed from background information variables such as students' family background, socio-economic status, ethnicity, gender, home language, school types, home environment, province, etc.

In this assessment, Item Response Theory has been used, with 2PL used to analyse items of 1 point and GPCM used for questions with 2 or more points. TAM was used to analyse the results.

The latent ability (theta) of each student is obtained from the analysis stated above. Ten plausible values were calculated to find out the ability of the population on the basis of the latent ability, and the dummy variables made on the basis of the variables asked in the students' background questionnaire. After transforming the average of those plausible values to zero (0), the following formula was used to present the results:

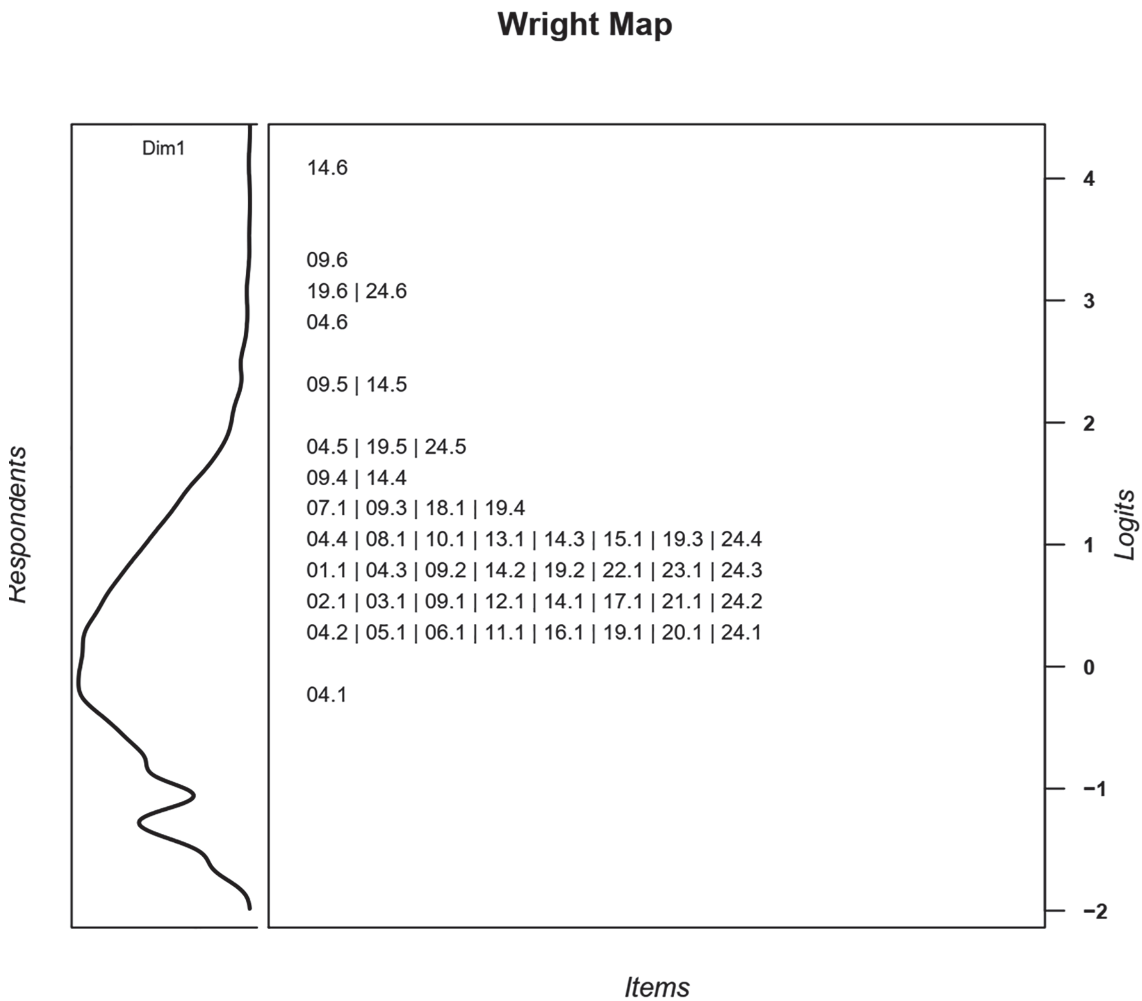
Student Achievement Score = 500 + PV * 50

Because there was an insufficient number of anchor items and visible differential item functioning values, the English language result was not compared with the assessment of 2017. The mean score is not compared with any other assessment, but it can be compared within the year with different variables, and the mean score becomes 500 at the national level.

The Wright map in English

The Wright map is organised as a frequency curve and vertical histogram. The frequency curve on the left side shows the respondents, and the histogram on the right side shows the test items. The left side of the map shows the distribution of the measured ability of the respondents from most able at the top to least able at the bottom. The items on the right side of the map are distributed from the most difficult at the top to the least difficult at the bottom. In Figure 93, student ability (θ) on the left and NASA 2020 items to the right are plotted on the same scale. When a person and an item lie at the same level, the probability of that response by the particular respondent is 50%. Figure 89 presents the NASA 2020 Wright map for English.

Figure 89 Wright-map showing respondent and item on the same scale



The curve on the left side represents the number of students; their latent ability is displayed in the logit scale ranging from -2 to +4. The distribution of students against the items administered (item numbers are shown to the right side) reveals that most of the items were difficult for the students. Although items were pre-tested and based on the grade 8 curriculum, most of the students were below the average latent ability '0'. This indicated that test items were difficult for the participating students. This further depicts that the performance level of grade 8 students in English was not as expected by the curriculum.

The results for English have been presented skill-wise, first Reading and Writing and then Listening and Speaking. Achievement in reading and writing skills has been provided in combination and separately.

Reading and writing

Proficiency level-wise student distribution

The assessment framework for NASA 2020 set students' proficiency standards into six different levels. The data presented in Table 114 shows the overall distribution of sample students across the six proficiency levels: Below basic, Basic, Proficient 1, Proficient 2, Proficient 3, and Advanced. The percentage of students falling into each of the six levels is shown in Table 114.

Table 114 Proficiency level-wise distribution of the student population

	N	Percentage of students
Below basic level	2360	0.5
Basic level	46814	10.2
Proficient 1 level	172885	37.8
Proficient 2 level	152000	33.3
Proficient 3 level	75707	16.6
Advanced level	7367	1.6
Total	457133	100

Table 114 shows that the smallest numbers of students were to be found in the Below basic (0.5%) and Advanced (1.6%) levels. The number of students with Basic (10.2%) and Proficient 3 (16.60%) levels was comparatively lower than for Proficient 1 (37.80%) and Proficient 2 (33.3%) levels. Results indicate that the proficiency level of most students (87.50%) are around the average, and the proficiency level of around half of the students (51.30%) was above the national average score (500). Furthermore, only a tenth (10.8%) of the students were found to be at the Basic level or below in terms of proficiency. The data also can be seen in bar diagram form in Figure 90.

Figure 90 Distribution of students across the different proficiency levels in English

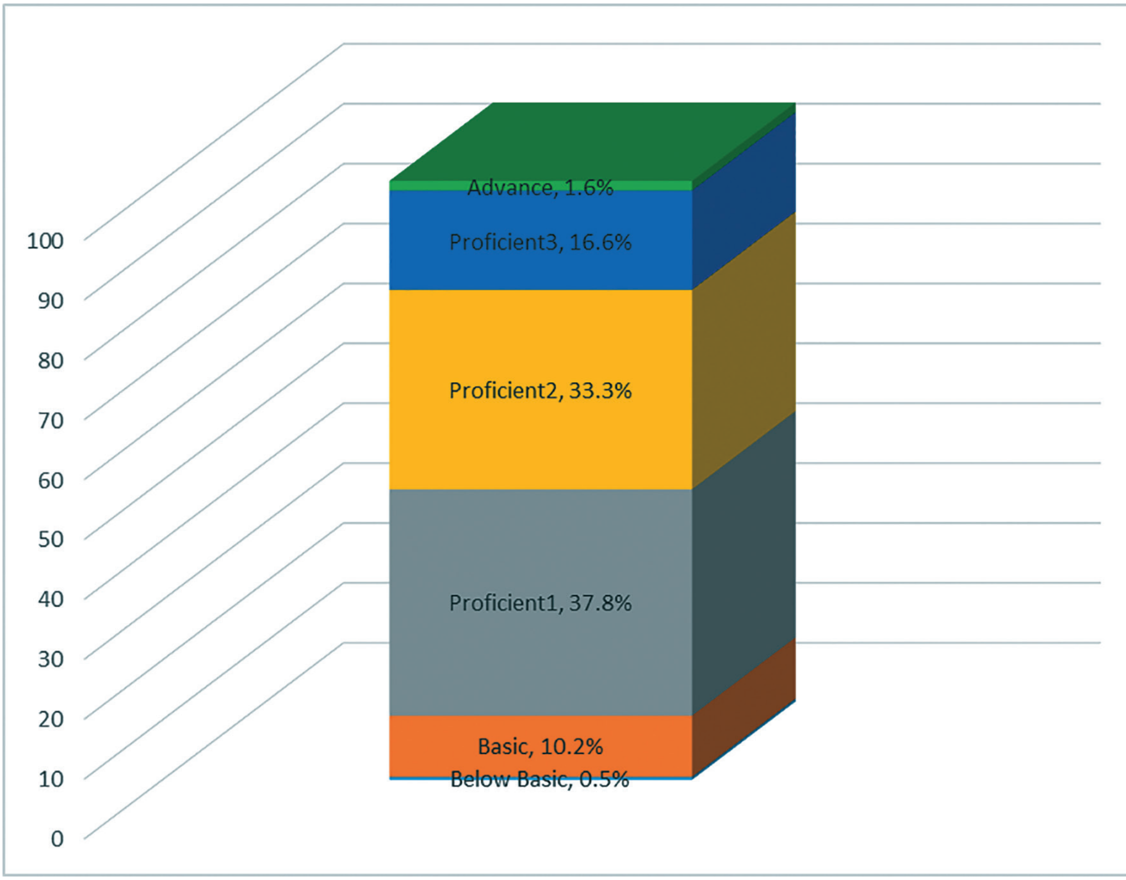


Figure 90 shows a histogram of the students' proficiency levels. The data reveals that 51.5% of students passed the minimum level of proficiency, whereas 48.5% were still struggling to achieve the minimum level of proficiency in English.

Distribution of students by proficiency level in the seven provinces

As the Federal Democratic Republic of Nepal is divided into seven provinces and their local levels, the achievements of students in English were also analysed by province, namely: Province 1, Madhesh, Bagmati, Lumbini, Gandaki, Karnali and Sudurpaschim. The distribution of students according to their levels of proficiency in English varies by province. The province-wise distribution of respondents for English, based on their number and percentage, are presented in Table 115.

Table 115 Distribution of students by proficiency level across the seven provinces

	Below basic		Basic		Proficient 1		Proficient 2		Proficient 3		Advanced	
Provinces	N	Per cent	N	Per cent	N	Per cent	N	Per cent	N	Per cent	N	Per cent
Province 1	0	0	6046	9.2	25600	38.7	22413	33.9	11216	17	792	1.2
Madhes Province	2027	2.7	12904	17.3	28449	38.2	22682	30.5	7544	10.1	826	1.1
Bagmati Province	0	0	961	1.2	17167	22.2	29789	38.5	26662	34.5	2793	3.6
Gandaki Province	0	0	557	1	15718	28.5	22666	41.1	14199	25.7	2048	3.7
Lumbini Province	0	0	6257	8.6	31984	44.1	24755	34.1	9035	12.5	522	0.7
Karnali Province	84	0.2	9480	17.3	28075	51.1	14984	27.3	2174	4	103	0.2
SudurPaschim Province	249	0.4	10610	18.7	25893	45.7	14711	26	4878	8.6	283	0.5

Table 115 shows that the proficiency level for the majority of students was found to be lower than the national average in Sudurpaschim (64.8%), Karnali (68.6%) and Madhesh (58.20%), but around the national average in Province 1 (47.9%) and Lumbini (52.70%). Furthermore, the proficiency level was found to be higher than the national average in Bagmati (76.6%) and Gandaki (70.5%). Additionally, these results show that the proficiency level of students in English was found to be better in Bagmati and poorer in Karnali Province when all provinces are compared. The data from Table 119 is presented as a bar diagram in Figure 91.

Figure 91 Distribution of students by proficiency level across the seven provinces

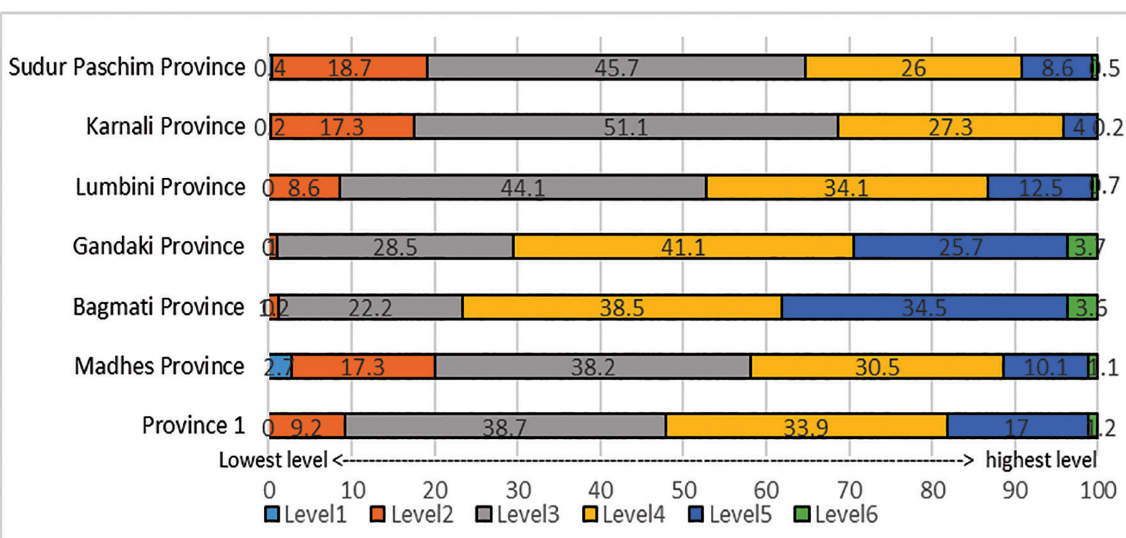
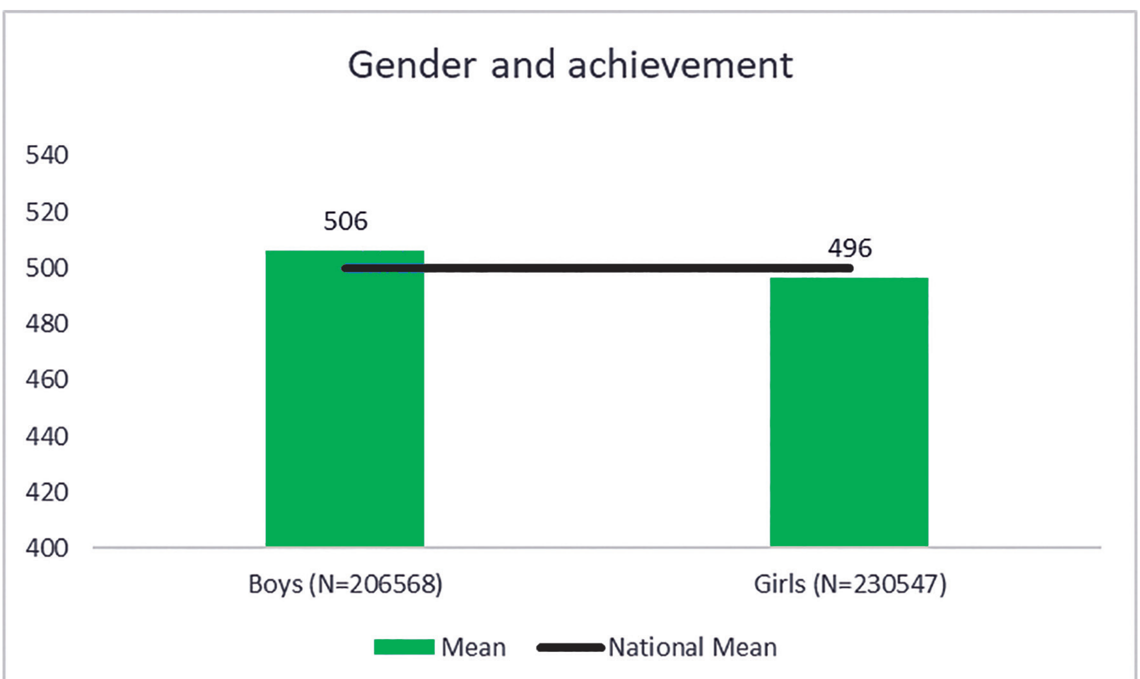


Figure 91 shows that the number of students achieving Advanced level was found to be below 1% in Sudurpaschim (0.5%), Karnali (0.2%) and Lumbini (0.7%). However, the percentage of students obtaining that level is also very poor in the remaining provinces. Furthermore, the number of students found to be Below basic level in Lumbini, Gandaki, Bagmati and Province 1 was zero, and was below 1% in Sudurpaschim and Karnali Province. The percentage of level 1 respondents in Madhesh is 2.7, which is the highest percentage achieving at the lowest level. Similarly, the percentage of students at level 2 (Basic) in Sudurpaschim, Karnali and Madhesh was 18.7, 17.3 and 17.3 respectively. Likewise, Lumbini, Sudurpashchim and Karnali provinces had percentages of 44, 45.7 and 51.1 respectively at proficiency level 3, whereas Bagmati and Gandaki province had 34.5% and 25.7% respectively at proficiency level 5. Overall, the results show that Karnali, Sudurpaschim and Madhesh have the lowest percentage at level 5 and level 6, but Province 1, Gandaki and Bagmati have higher achievement percentages.

Achievement by gender

In NASA 2020, achievement by gender has also been analysed. Looking at the percentage of boys and girls at different learning levels, it appears that there is inequality in learning. The bar diagram in Figure 92displays the achievement by gender in English.

Figure 92 Achievement in English by gender



The bar diagram shows the national average (500) achievement score from the NASA 2020 study against the actual results based on gender. The data shows that boys achieved 506, which is above the national average, whereas girls achieved 496, which is below the national average. The results indicate that girls need more learning opportunities compared to boys.

Figure 93 Gender-wise distribution of students by proficiency level

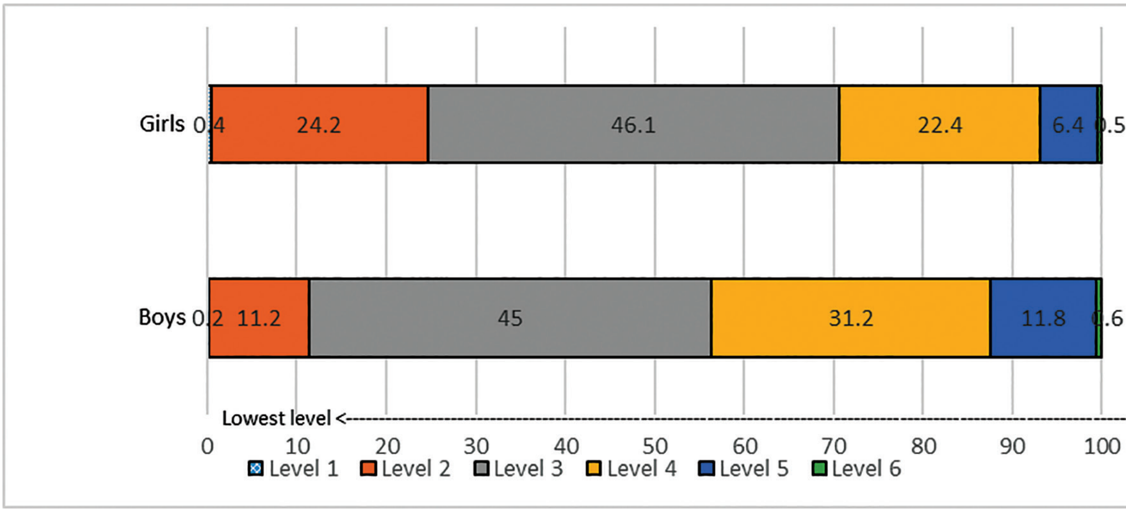


Figure indicates that 29.3% of girls crossed the minimum level of achievement, whereas only 43.6% of boys crossed the minimum level of proficiency. The results also show that 56.4% of boys are below the minimum level of proficiency, while 70.7% of girls achieved below the minimum score level in NASA 2020.

Achievement by caste/ethnicity

In NASA 2020, one of the areas of enquiry has been students' ethnic background and its relation to achievement. The data has been analysed to see the achievement of students from different ethnic backgrounds. Broadly based on the geographical belts, i.e. Mountain, Hill and Madhesh, three important ethnicities have been considered for study: Brahmin/Chherti, Janajati and Dalit. Students with ethnicities not falling into these three categories have been recognised here as 'Other'. The details are presented in Table 116.

Table 116 Achievement by caste/ethnicity

Ethnic group	Mean	N	Std. error of mean
Madhesi Brahmin	505.2750	37300	.24999
Madhes Janajati	496.5299	55662	.18168
Madhesi Dalit	486.4111	14564	.35453
Madhesi Others	494.9069	31776	.24860

Hill Brahmin	511.6426	121950	.13390
Hill Janajati	506.1659	82727	.14796
Hill Dalit	490.6801	30468	.21784
Hill Others	503.2637	16013	.34615
Mountain Brahmin	490.6657	6576	.55997
Mountain Janajati	501.3341	4060	.67275
Mountain Dalit	495.2164	1292	1.13366
Mountain Others	478.5478	1337	1.48902
Total	503.0992	403725	.07100

Table shows that Hill Brahmin/Chhetri had the highest mean score (512), whereas Madhesi Dalits achieved the lowest mean score (486). The national average for caste/ethnic group is 503. As can be seen from the table, the mean achievement of Madhesi Janajati (496), Madhesi Other (494), Hill Dalit (490), Mountain Dalit (495), Mountain Brahman (490) and Mountain Others (478) were below the national average, whereas Hill Brahman (512), Hill Janajati (506) and Madhesi Brahman (505) obtained above the national average. Overall, results indicate that students from Brahman/Chhetri communities achieve a higher score when compared to the mean score of students from Janajati and Dalit communities. The caste/ethnic group of students can be categorised into three: Madhesi All, Hill All, Mountain All. These have been presented in Table 117.

Table 117 Achievement by caste/ethnicity

Ethnic group	Mean	N	Std. error of mean
Madhesi All	497.4434	139302	.12067
Hill All	506.7615	251158	.08923
Mountain All	493.1529	13265	.39695

The data in Table 117 shows that when the results are compared, Mountain All achieved the lowest score (493), while Hill All achieved the highest score (506). The data reveals that Madhesi All (497) achieved higher than Mountain All (493), although both results are below the national average (503).

Achievement by students' age

In the NASA 2020 study, the respondents were asked to state their age as a background variable so that the relationship between age and educational achievement could be analysed. The ages of the students were categorised on five strata: 12 years or less, 13 years, 14 years, 15 years and 16 years or above. The mean scores of their educational achievement by age are presented in Table 118.

Table 118 Achievement by students' age

Age group	Mean	Std. error of mean	N
12 and below	503.1433	0.24509	34133
13	507.4663	0.14461	108266
14	506.7094	0.11844	154684
15	493.5777	0.14591	83596
16 and above	485.3063	0.17985	47551
Total	501.6764	0.07000	428229

Table 118 illustrates that students aged 12 and below achieved a mean score of 503. Students aged 16 or above achieved 485, which is the lowest score achieved in the study. The national mean score by students' age group is 501, and the higher scores were achieved by students aged 13 and 14 with mean scores of 507 and 506 respectively. This can be represented as a bar diagram as in Figure 94.

Figure 94 Achievement by students' age

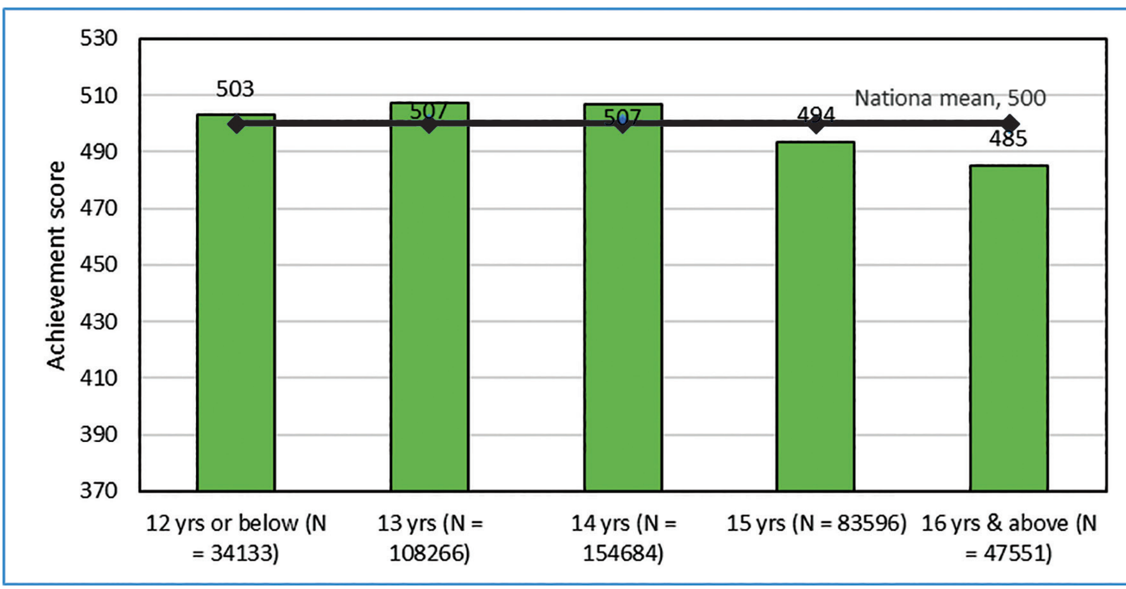


Figure 94 reveals that the 13 years and 14 years age groups achieved the highest mean score (507), and the lowest mean score (485) was for the age group of 16 years or above. The 12 years, 13 years and 14 years age groups scored higher than the national average, whereas the 15 years and 16 years or above age groups scored below the national mean. Overall, the results indicate that 12, 13 or 14 years of age is the best period for higher achievement.

Achievement by home language

The NASA 2020 study investigated the association of students' home language with their educational achievement. In the study, 325,646 respondents stated that they spoke Nepali as their home language, whereas 114,466 stated that they spoke other languages at home. The achievement of students by home language is presented in Table 119.

Table 119 Achievement by home language

Language spoken at home	N	Mean	Std. deviation	Std. error of mean
Nepali	325646	504.4749	45.83345	0.08032
Others	114466	491.9608	44.14757	0.13049

Table 119 shows that students who speak Nepali at home scored 504, whereas speakers of other languages at home scored 492. There is a significant difference in the achievement of Nepali speakers at home compared to other languages.

Figure 95 Achievement by home language

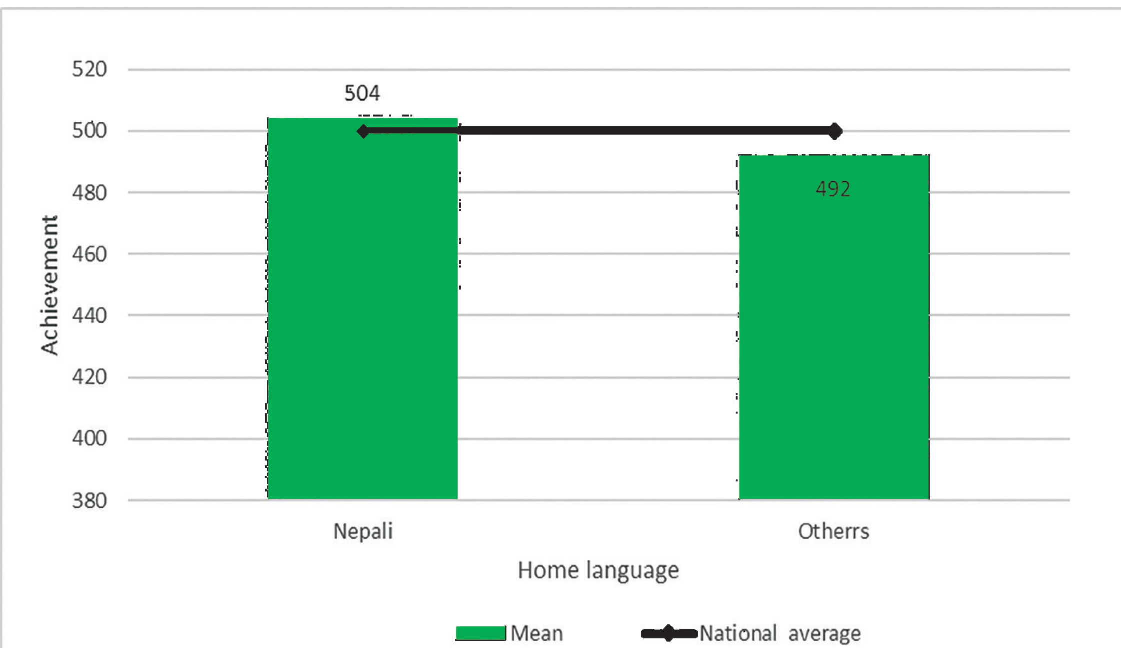


Figure 95 shows that students who speak Nepali language at home have better performance (504) than the speakers of other languages (492) in English.

Achievement by family size of the students

In NASA 2020's background questionnaire, the respondents were asked to state the size of their family. The family size ranged from four to ten members. Table 120 illustrates family size and its association with students' educational achievement.

Table 120 Achievement by family size of the students

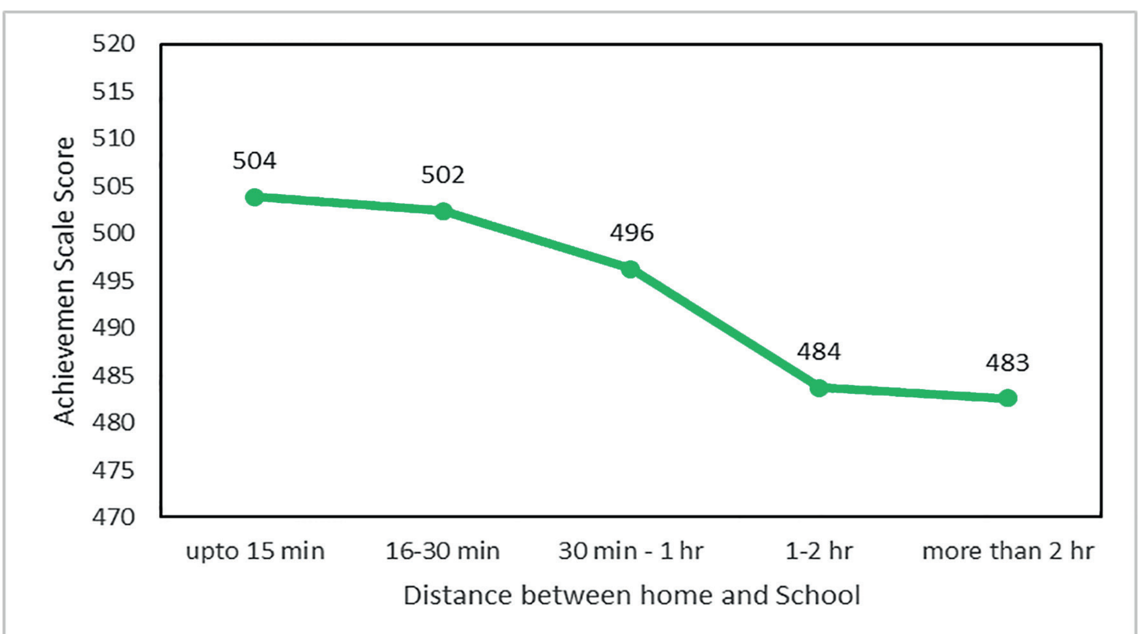
Number of family members	Mean	N	Std. error of mean	Std. deviation
4 or less	517	119563	0.132	45.8
5	504	106777	0.136	44.6
6	495	80162	0.154	43.5
7	492	46155	0.200	42.9
8	488	27046	0.262	43.0
9	487	15087	0.336	41.3
10 or more	489	41441	0.215	43.8
Total	501	436232	0.069	45.6

Table 120 shows that when the family size was four members or less, the educational achievement of students was highest (517), and when the family size was six members or more, students' achievement in English was below the national average. This shows that there is a negative correlation between the number of family members and a student's achievement, so children from families with fewer members have higher achievement scores in English.

Achievement by distance to school

In NASA 2020, one of the question areas was distance between school and home. To analyse the effect of school distance on achievement, distances were categorised into five groups: up to 15-minute walk, 16–30 minutes, 30 minutes to 1 hour, 1–2 hours, and more than 2-hour walking distance. The results are shown in Figure 96.

Figure 96 Achievement by distance between school and student's home



From Figure 96, it can be seen that school distance and students' achievement in English have a negative relationship. The results show that students whose educational institution is up to 30 minutes' walk away have a higher than national average score, whereas distances that take more than 30 minutes to walk result in scores that are below the national mean, i.e. 496, 484 and 483 respectively. The data revealed that up to a 15-minute walk between school and home is the ideal distance, as these students scored highest (504), whereas a distance of more than two hours affects performance, as these students achieved the lowest score (483).

Achievement by medium of instruction

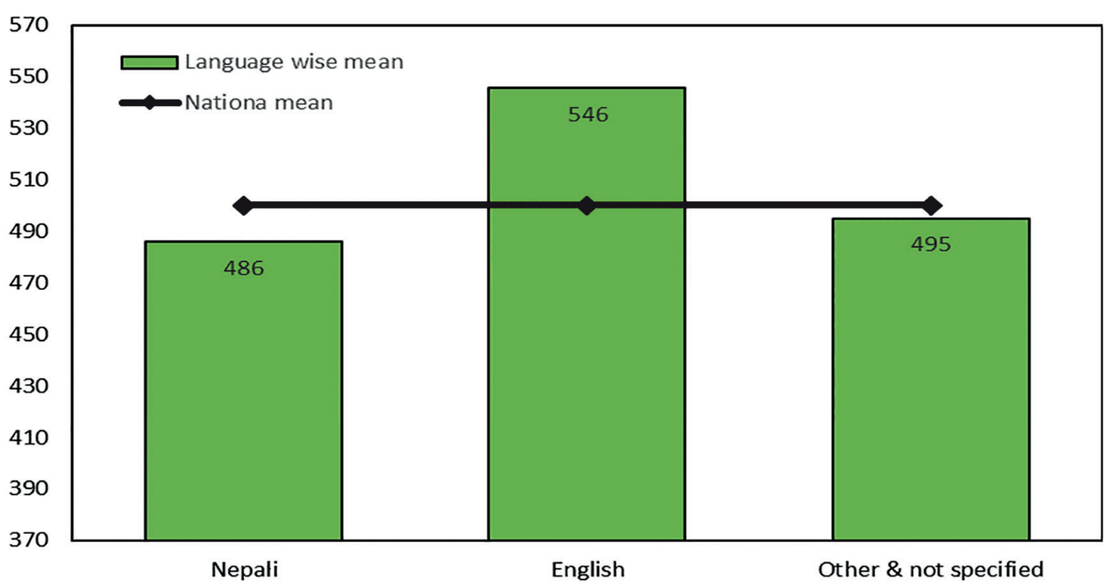
NASA 2020 also investigated the medium of instruction. Respondents were asked to state their medium for teaching English language. The respondents' answers were categorized as Nepali, English, and Other or not specified. Details of achievement based on the medium of instruction are presented in Table 121.

Table 121 Achievement by medium of instruction

Medium of instruction	Mean	N	Std. error of mean	Std. deviation
Nepali	485.8579	279931	0.06902	36.51737
English	545.5363	96113	0.12789	39.64958
Other or not stated	494.8475	81090	0.17267	49.1699

Table 121 shows that students whose medium of teaching was English scored highest (545.53), whereas students whose mediums of instruction were Nepali or Other and not stated scored 485.85 and 494.84 respectively, below the national mean score. This is represented in Figure 101.

Figure 97 Achievement by medium of instruction



From Figure 97, it can be seen that English as the instructional language results in higher achievement than the national average in the subject of English, whereas the result is poor for other languages. Hence, schools should follow English-medium instruction to achieve higher scores, although regarding quality concerns about the instructional language, it does not necessarily mean that English is the best medium of instruction. The results indicate that further research into the selection of the medium of instruction seems necessary.

Achievement by time spent at home on different activities

Students' engagement in different activities before and after school were categorised as: watching TV/internet/mobile, play with friends and groups, household work, study/homework, work for a wage, reading other books, and supporting siblings to read. These were measured from no time given to more than 4 hours. The results are presented in Table 122.

Table 122 Achievement by time spent on different activities

Time spent by students	No time given	Less than 1 hr	1–2 hr	2–3 hr	More than 4 hr
Watching TV/Internet/Mobile	480	507	525	522	501
Playing with friend	491	506	509	509	491
Household chores	483	512	506	499	490
Study/homework	472	489	508	516	509
Working for wage	511	496	492	487	484
Reading other books	488	511	509	499	485
Supporting brother/sister to read	499	508	507	497	484

Table 122 shows that those students in the habit of spending their time using digital resources, such as watching TV/internet/mobile, for more than one hour perform better in English compared to the other activities, except for study/homework 2–3 hour per day and reading other books. Additionally, achievement was found to be poorest among those students not given time for watching TV/internet/mobile. Therefore, guardians and teachers should encourage students to use these resources for a limited time period, to enhance learning achievement.

Achievement by support provider at home for writing/reading

In NASA 2020, one of the questions students were asked related to the support provided to them at home with reading and writing. The responses were categorised into nine groups, including father, mother, sibling, friends, grandparents and tutors, and these are presented in Table 123.

Table 123 Achievement by support provider at home for writing/reading

Support provider at home for writing/reading	Mean	N	Std. error of mean
Father	498	64304	0.18064
Mother	504	29550	0.28089
Brother/sister	498	203551	0.09696
Grandparents	489	2561	0.87398
Friends	496	30368	0.23848
Others	512	12835	0.40616
Tuition	514	39881	0.20856
Nobody	515	18724	0.32737
Total	501	401774	0.0704

Table 123 shows that students who have nobody supporting them at home achieved the highest average scale score (515). Of people providing support, tutors were the most effective as the students achieved a scale score of 514. Similarly, students who were supported by their mother or others achieved well, scoring 504 and 512 respectively, both of which were above the national average. Students who were supported by grandparents and friends achieved 489 and 496 respectively, which were below the national average and were less effective. The results are presented as a bar diagram in Figure 98.

Figure 98 Achievement by support provider at home for writing/reading

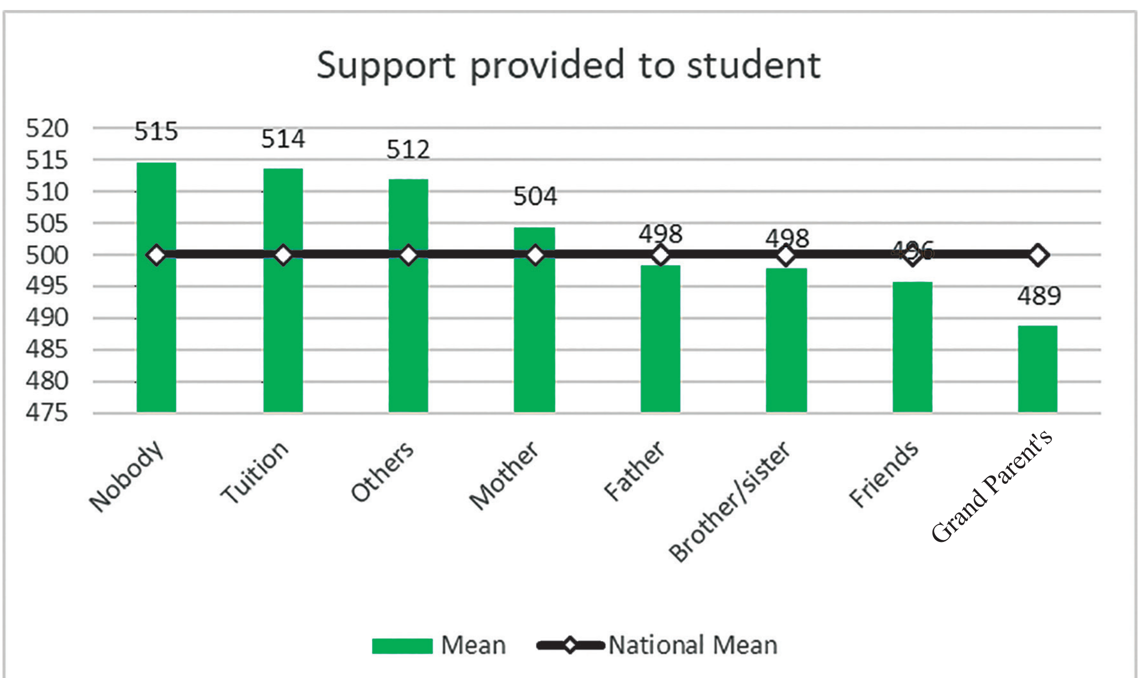


Figure 98 shows that the achievement scores of students supported by their mother, others, tutors or nobody were higher than the national average. However, the results are poorest for students getting support from grandparents. The highest achievement is found among learners not getting help from anybody, indicating that self-learners perform better in English. Therefore, students should be motivated towards self-learning at this age.

Achievement by career aspiration

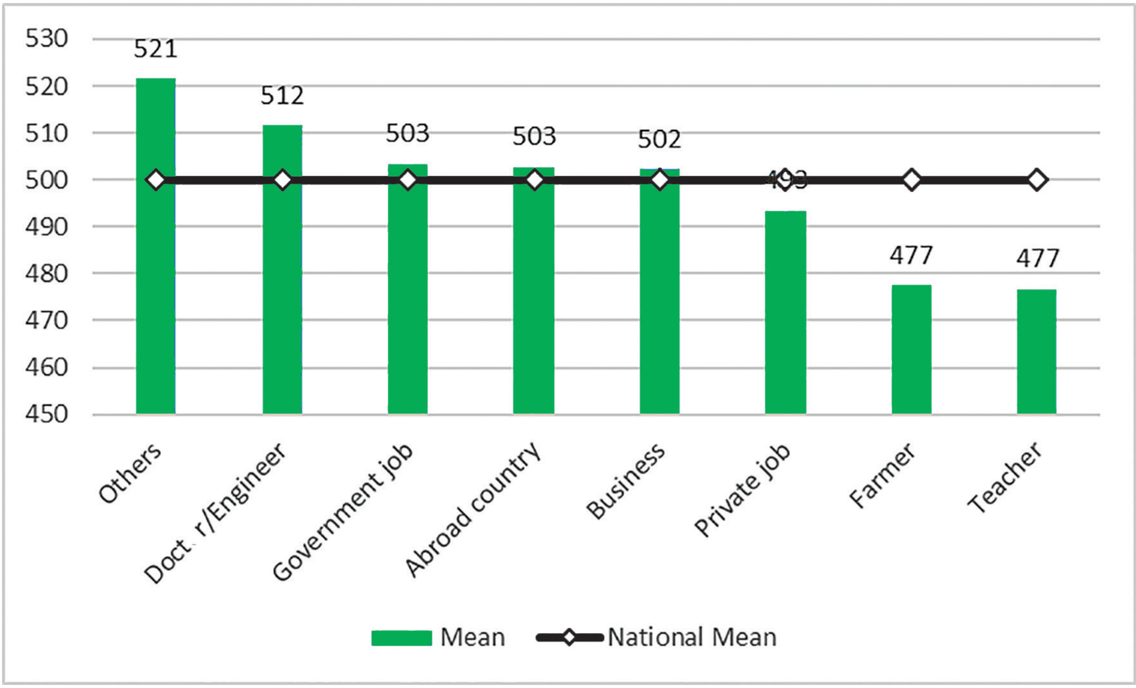
In NASA 2020, students could indicate their career aspiration, although most students (the highest frequency) did not respond to this. Their responses were categorised as teacher, government official, private job, business, work abroad, farmer, and doctor and engineer as in Table 124.

Table 124 Achievement by future career goal

Your future aim	Mean	N	Std. error of mean
Teacher	476.6765	105035	0.11291
Government job	503.4034	58213	0.17493
Private job	493.4151	5217	0.60383
Business	502.1675	17313	0.33307
Work abroad	502.7628	14130	0.36156
Farmer	477.467	12106	0.37862
Doctor/Engineer	511.6341	153867	0.11421
Others	521.4977	70157	0.16105
Total	501.8716	436038	0.06831

Table 124 shows that students wishing to go into teaching achieved the lowest, as their mean score was 476, whereas those who hope to be doctors and engineers achieved a score of 511. Interestingly, the students whose aspiring careers came under 'Others' had the highest achievement (521), which indicates the necessity of including other careers in the study. The overall mean score was 501.87 based on students' choice of future career goal in relation to their achievement. The results are displayed in Figure 99.

Figure 99 Achievement by career aspiration



From Figure99it can be seen that those students interested in becoming a doctor/engineer or working for the government, abroad or in business perform better in English than the national average. However, those interested in being farmers or teachers have lower achievement compared to others and the national average.

Achievement and use of leisure time in school

In NASA 2020, students were asked to state their leisure-time activities in school. Their engagement was categorised as classwork, homework, play, and return home. The details are given in Table 125.

Table 125 Achievement by use of leisure time

Leisure-time activity	Mean	N	Std. error of mean	Std. deviation
Classwork	501.4336	256975	0.0871	44.15535
Homework	500.9414	146014	0.12160	46.46486
Play	505.5959	28673	0.28769	48.71408
Return home	465.1174	9893	0.38021	37.81717
Not specified	479.3812	15579	0.49392	61.64876
Total	500	457133		46.25077

Table 125 shows that those students who spent their leisure time playing (mean = 505.59), doing classwork (mean = 501.43) or doing homework (mean = 500.94) perform better in English. However, those who did not specify any types of activities had the lowest performance (mean = 460.15). Interestingly, Table 132 reveals that those who played during leisure time achieved the highest in English. The results are displayed in Figure100.

Figure 100 Achievement by use of leisure time

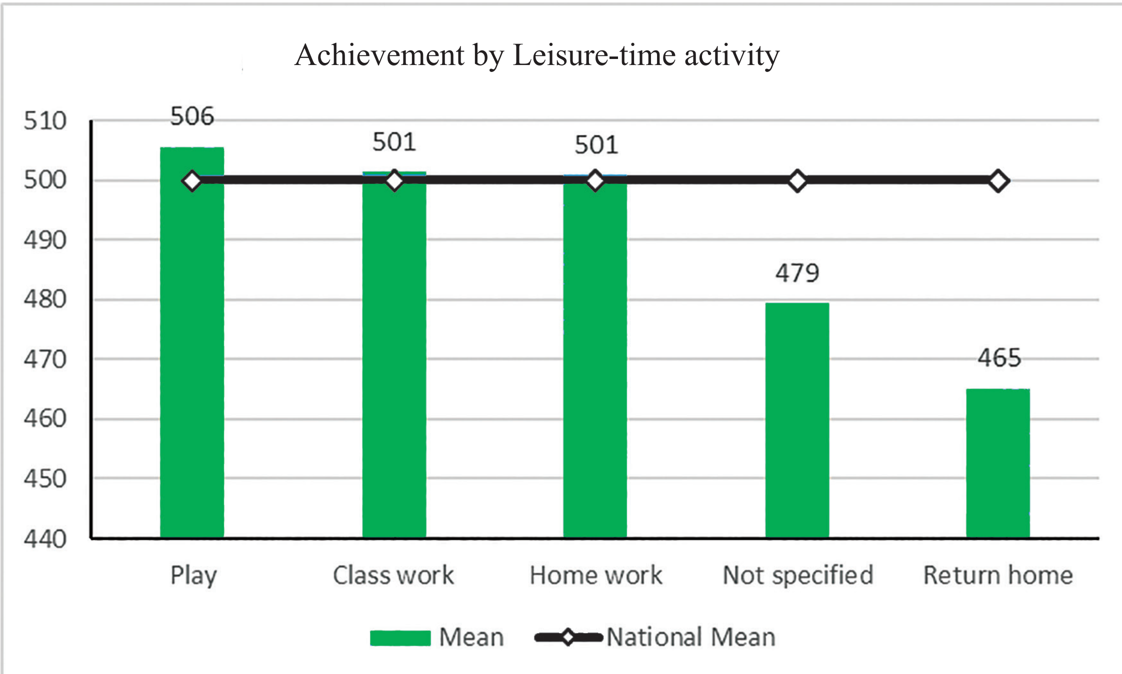


Figure 100 shows that those students in the habit of playing during leisure time achieved the best results, and those returning home had the poorest achievement in English. The results further show that students who played, did classwork or did homework in leisure time achieved 506, 501 and 501 respectively, which are higher than the national average, whereas those who returned home achieved 465, which is below the national average.

Achievement by parents' education

In NASA 2020, one of the key areas of enquiry was to see how far family-related factors, including parental education, could be associated with students' learning achievement. The role of both the mother's and the father's education level in students' learning was analysed.

Achievement by mother's education

Students were asked about their mother's education level in this study. Their responses were categorised under seven headings: Illiterate, Literate, Grade 8, Grade 10, Grade 12, Bachelor's, and Master's or above. The responses from students and their achievements are presented in Table 126.

Table 126 Achievement by mother's education

Mother's education level	Mean	N	Std. error of mean
Illiterate	484.2887	142957	0.10302
Literate	495.9103	113357	0.11774
Grade 8	503.0973	83566	0.1491
Grade 10	524.3687	53395	0.19143
Grade 12	534.4456	30915	0.26699
Bachelor's	549.3149	10692	0.43445
Master's or above	545.549	5580	0.65742
Total	501.5817	440463	0.06814

Table 126 indicates that students whose mothers were illiterate or just literate achieved below the national average, whereas students whose mothers passed grade 8, grade 10, grade 12, or had a Bachelor's or Master's degree or above achieved well above the national average. This is shown in Figure 101.

Figure 101 Achievement by mother's education

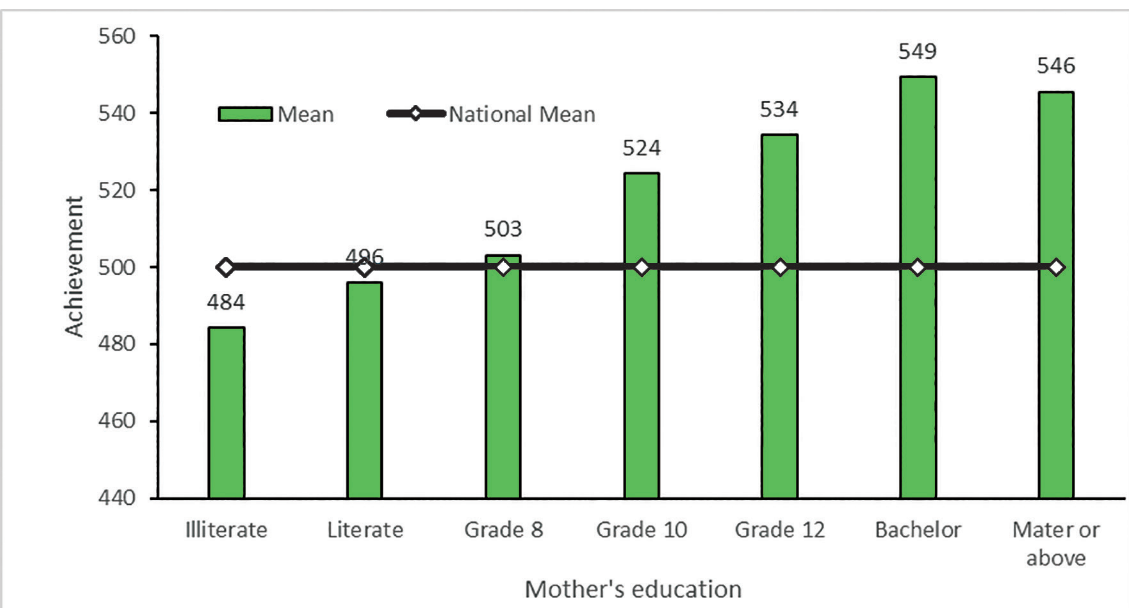


Figure 101 shows that the achievement scores of students with mothers whose education is Grade 8 or higher were found to be better than the national average. However, the scores are poorer than the national average for students whose mothers are illiterate or just literate. Furthermore, the achievement score for English was found to be highest among those students whose mothers have higher than school-level qualifications. The results may occur because these mother can support their children in learning English. Hence, relevant stakeholders should focus on enhancing the qualification of women.

Achievement by father's education

In the same way, students were asked about their father's education level in this study. Their responses were categorised under the same seven headings: Illiterate, Literate, Grade 8, Grade 10, Grade 12, Bachelor's, and Master's or above. The responses from students and their achievements are presented in Table 127.

Table 127 Achievement by father's education

Father's education level	Mean	N	Std. error of mean
Illiterate	478.1452	62413	0.1542
Literate	491.4871	87457	0.13059
Grade 8	491.8153	108584	0.12157
Grade 10	508.9681	99042	0.13979
Grade 12	524.2916	51055	0.2012
Bachelor's	537.1468	21857	0.3074
Master's or above	555.5081	11251	0.40727
Total	501.2852	441660	0.06816

Table 127 indicates that those students whose fathers were illiterate or at Grade 8 level or below achieved below the national average, whereas those students whose fathers passed grade 10, grade 12, or had a Bachelor's or Master's-level education achieved well above the national average. This is shown as a bar diagram in Figure 102.

Figure 102 Achievement by father's education

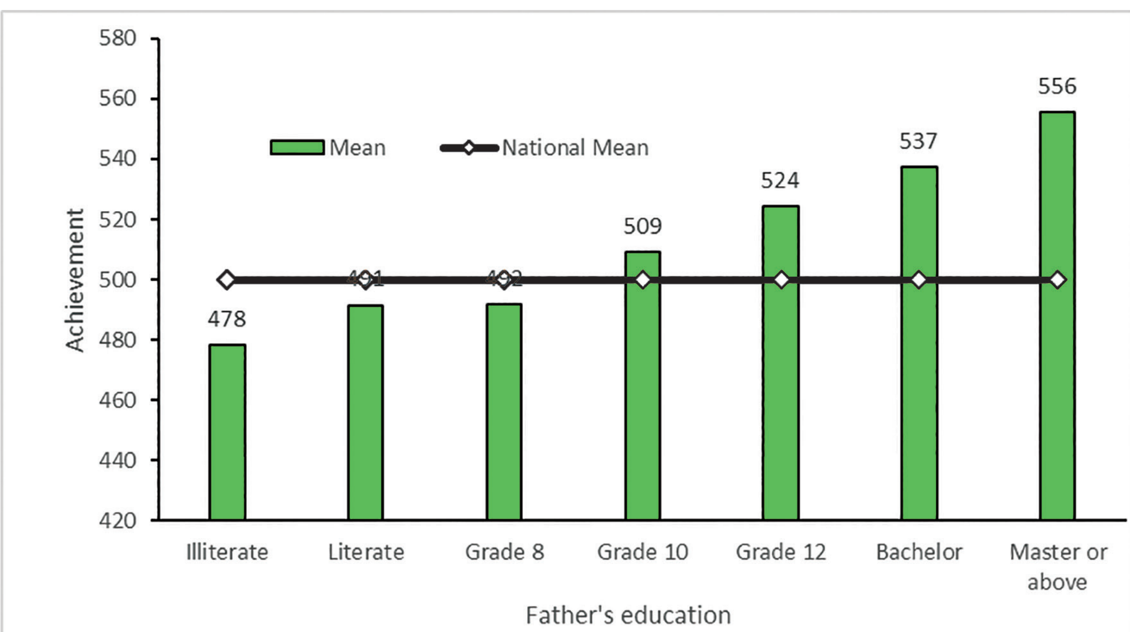


Figure 102 shows that the performance of students was found to be higher than the national average among students whose fathers had passed grade 10 or had a higher than school level education, with the best result being achieved by students whose fathers had a Master's qualification. The poorest results were achieved by students whose fathers were illiterate. It is clear that students' achievement increases as their father's level of education gets higher.

Achievement by parental occupation

In NASA 2020, students' learning achievement is considered from the perspective of parental occupation as well. The achievements of students having mothers engaged in various occupations were studied. In the same way, the association between father's occupation and the student's achievement was analysed.

Achievement by mother's occupation

In NASA 2020, the mother's occupation has been an important consideration in relation to students' learning achievement in English. For this purpose, nine categories of occupation were established for data analysis. According to the responses given by sample students, the number of students having mothers in these various categories is indicated in Table 128.

Table 128 Achievement by mother's occupation

Mother's occupation	Mean	N	Std. error of mean
Agriculture + household	488.5497	264344	0.07737
Household	517.7676	96854	0.14742
Work in other's home	495.5851	5113	0.66968
Labour	502.7175	5505	0.58857
Foreign employment	507.8091	8073	0.49622
Teaching	537.048	12121	0.41816
Business	524.9169	31188	0.24753
Government job	529.4933	8853	0.47024
Others	538.6567	9099	0.44458
Total	501.3339	441148	0.06807

Table 128 presents the distribution of students across the nine categories of their mothers' occupations. These were: Agriculture (264,344 samples), Household work (96,854 samples), Work in other’s home (5,113 samples), Labour (5,505 samples), Foreign employment (8,073 samples), Teaching (12,121 samples), Business (31,188 samples), Government service (8,853 samples) and Other (9,099 samples). The details of the students' achievements in relation to their mothers' occupation are shown in Figure 103.

Figure 103 Achievement by mother's occupation

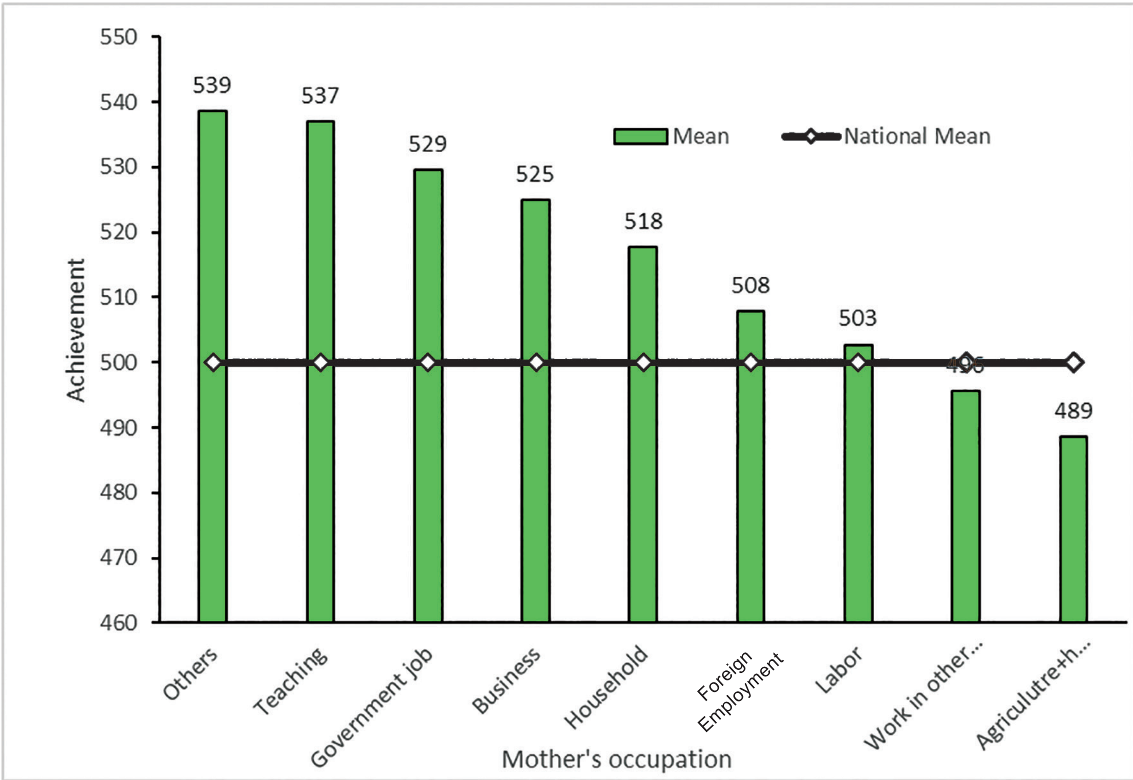


Figure 103 shows that students whose mothers work in others' home or agriculture have poor performance in English, which may be because those jobs in Nepal are mostly filled by people who are illiterate or have a low level of education, and as such they cannot support their children in content and care. The results also show that students whose mothers have teaching or government jobs perform better in English.

Achievement by father's occupation

In NASA 2020, the father's professional background has a direct impact on students' learning achievement. The father is assumed to be the family breadwinner in Nepalese society, and he plays a dominant role in maintaining the financial status, which is related with providing better options and educational opportunities. The fathers' occupations were grouped into agriculture & household, only household, work in other home, labour, work abroad, teaching, business, and government job. The details are presented in Table 129.

Table 129 Achievement by father's occupation

Father's occupation	Mean	N	Std. error of mean
Agriculture + household	481.8486	130654	0.10555
Household	482.4612	11066	0.38045
Work in other home	478.2208	9126	0.38541
Labour	496.3194	33641	0.20942
Foreign (Work abroad)	502.1785	105171	0.13269
Teaching	524.0573	14317	0.40475
Business	519.7127	65749	0.17913
Government job	524.1564	32652	0.24685
Others	527.4834	32452	0.24185
Total	501.5226	434828	0.06867

Table 129 illustrates that students whose father's profession was in teaching, business, government, or others achieved mean scores of 524.05, 519.71, 524.1 and 527.48 respectively. Students whose fathers worked in agriculture and household work or who worked in others' homes achieved 481.84 and 478.22 respectively. This data is represented in Figure 104.

Figure 104 Achievement by father's occupation

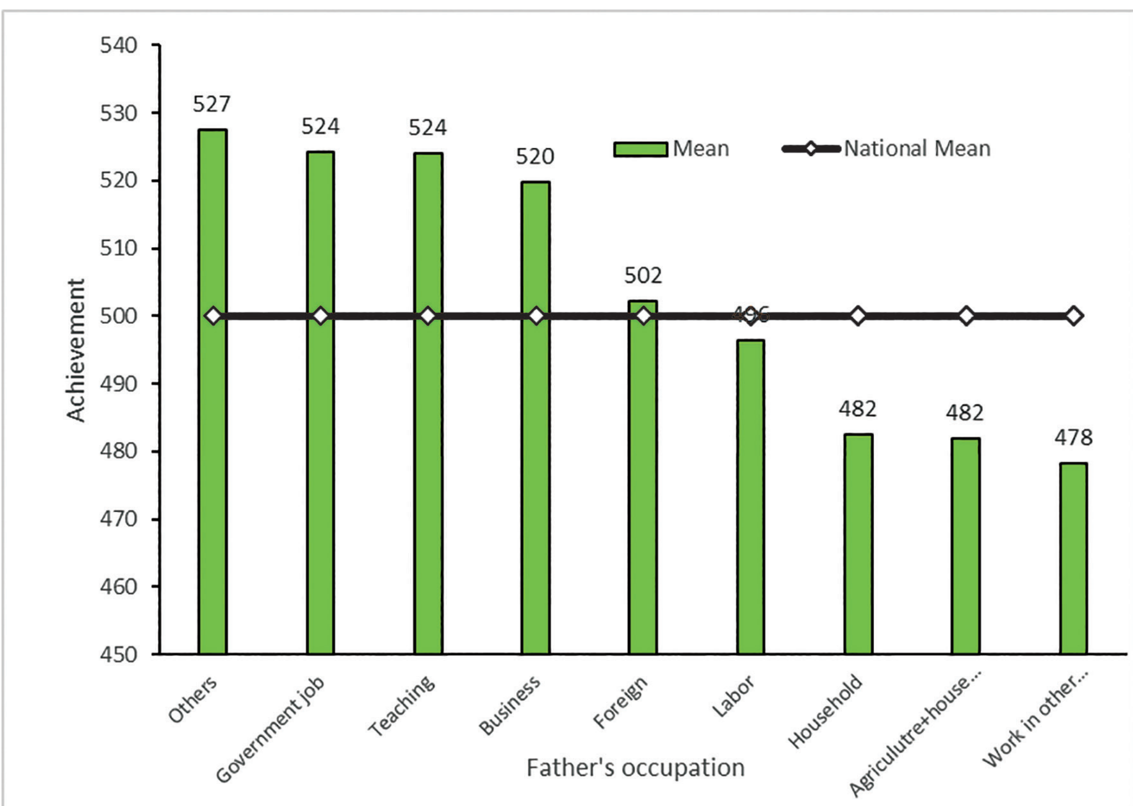


Figure 104 shows that students whose fathers have jobs in government, teaching, business, abroad and others perform better in English than other students and the national average. The results were found to be poorest for students whose fathers had household jobs or worked in others' homes, which may be because in the Nepalese context these types of jobs are filled by people who are either illiterate or less qualified, because those having high levels of qualification are engaged in other jobs.

Achievement by number of reference books at home

This study also enquired about the number of reference books available to students for study at home. The numbers are presented in Table 130.

Table 130 Achievement by number of reference books at home

Number of books other than textbooks	Mean	N	Std. deviation	Std. error of mean
None	487	110095	39.76494	0.11984
1–50	508	272189	45.2761	0.08678
50–100	507	28376	48.57098	0.28834
More than 100	497	12844	50.6454	0.44689

Table 130 indicates that the achievement score of students is found to be poor among students who do not have books at home (mean = 487). However, the score was found to be comparatively good for those having below 50 books (mean = 508) or 50–100 books (mean = 507) at home, but the achievement score was found to be higher than the national average in all cases except where there were no books at home. This is presented as a graph in Figure 105.

Figure 105 Achievement by number of reference books at home

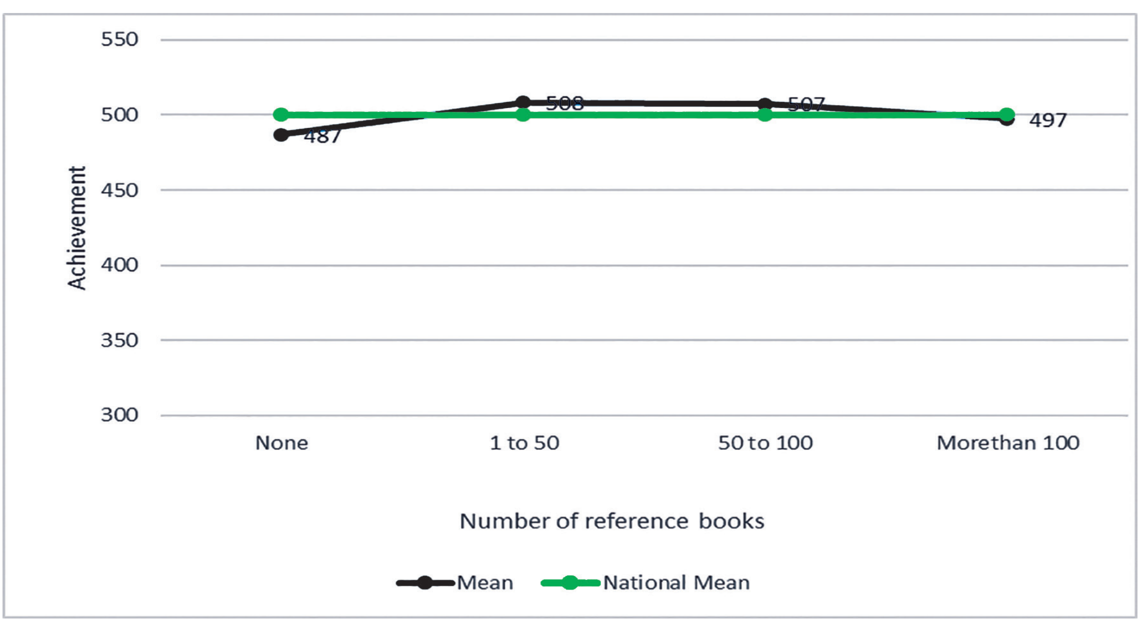


Figure 105 shows that students who have no books or more than 100 books at home have an achievement score below the national average. However, those having between 1 and 100 books at home perform better in English when compared to others and the national average. The results may be obvious, because most school-level students may not have time to study more books at home.

Achievement by home accessories

Based on the information given by students, data is tabulated regarding the various accessories found in their homes, such as a Table for study, Separate study room, Peaceful space to study, Computer for schoolwork, Children's magazine, story/poetry and pictures, Reference book for schoolwork support, Internet, and dictionary. They are categorised into number of facilities: none, one, two, three, four, five, six, seven and eight. Table 131 presents how many home facilities students did or did not have and their achievement in English.

Table 131 Achievement by home accessories

Number of home accessories	Mean	N	Std. error of mean
0	472.0262	9879	0.48907
1	488.9444	210180	0.08689
2	494.4942	52744	0.19007
3	500.2596	64443	0.17153
4	511.8695	51145	0.20423
5	523.8580	30937	0.27292
6	533.6113	17729	0.36493
7	548.3377	11430	0.43971
8	564.5498	6618	0.49344
Total	500.1034	455106	0.06850

Table 131 depicts that when the number of home facilities increases, the achievement of students in English also increases. These facts are shown as a graph in Figure 106.

Figure 106 Achievement by number of home accessories

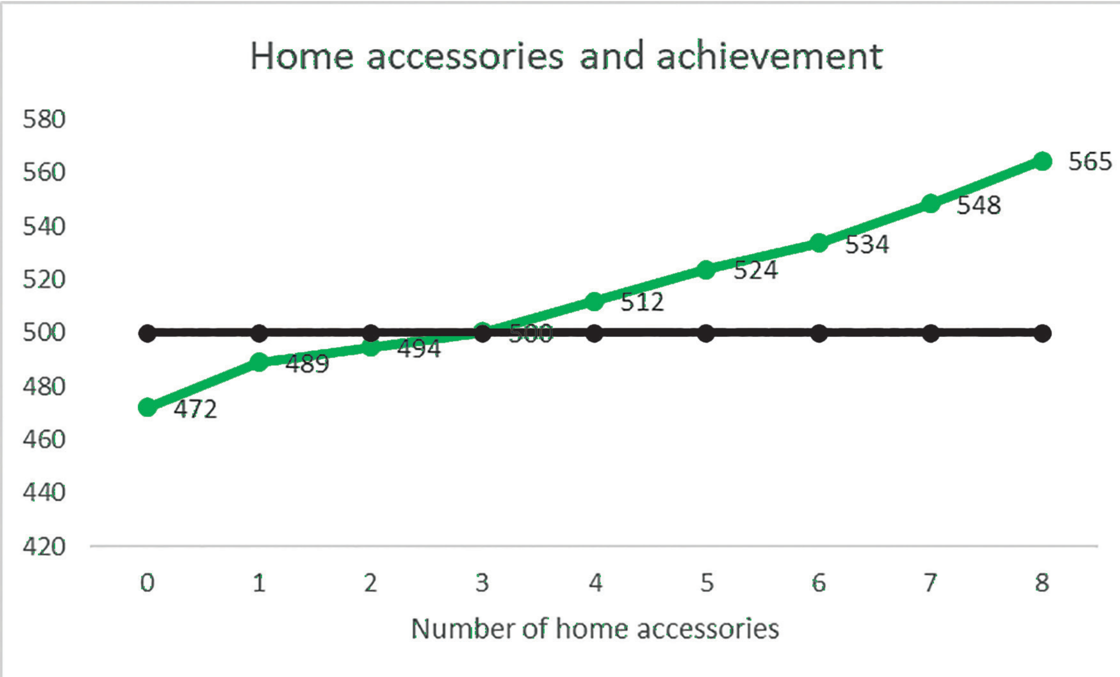


Figure 106 shows that students who have access to more than three home accessories perform better than the national average. Additionally, the number of home facilities and achievement are positively correlated. Therefore, relevant stakeholders should focus on increasing home accessories for students at home to enhance their achievement score in English.

Achievement by availability of a personal mobile phone

In the NASA 2020 study, one of the questions concerned students' possession of mobile phones. Data has also been studied concerning possession of a mobile in relation to students' learning achievement, and how far this is applicable to the subject of English. Of the 457,133 responses, 335,059 students expressed that they did not possess a mobile phone, and only 94,439 of them stated that they possessed a mobile. The data is shown in Table 132.

Table 132 Achievement by availability of a personal mobile phone

Do you have a mobile ?	Mean	N	Std. deviation	Std. error of mean
No	501	335059	44.57785	0.07701
Yes	504	94439	48.94375	0.15927
Not stated	470	27636	46.38918	0.27905
Total	500	457133	46.25077	0.06841

Table 132 shows that students who do not possess a mobile phone achieved a slightly lower score (mean score 501) than students who possess a mobile phone (mean score 504). The findings indicate that students who possess a mobile achieve better in English than those who do not possess a mobile. This data is presented as a bar diagram in Figure 107.

Figure 107 Achievement by availability of a personal mobile phone

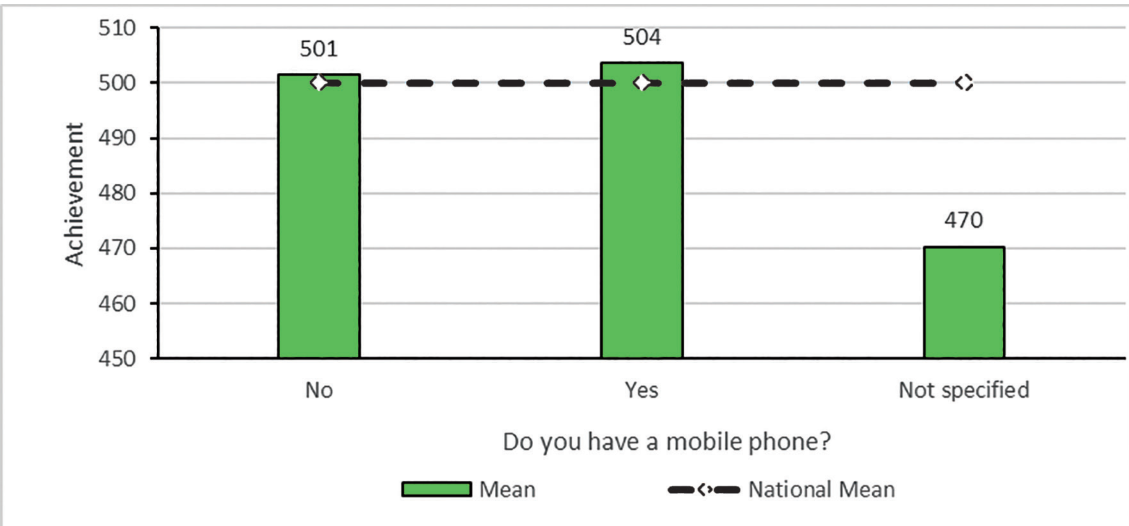


Figure 107 shows that students who have a mobile perform better in English than the national average. A similar result was found for students who do not have a mobile. However, achievement was found to be poor for students who did not specify whether they had a mobile available to them or not. In the Nepalese context, most mobile users can access the internet, through which learners can immediately get help with problems at any time and anywhere. Therefore, mobile devices are important for all learners in this age.

Achievement by access to social media

Social media has also been one of the factors that impact students' achievement. Therefore, one of the questions students were asked was regarding access to social media. The details are presented in Table 133.

Table 133 Achievement by access to social media

Do you have social media ?	Mean	N	Std. deviation	Std. error of mean
Yes	523	102024	47.34623	0.14823
No	496	324525	42.80314	0.07514
Not Stated	466	30584	43.88307	0.25093
Total	500	457133	46.25077	0.06841

Table 133 shows that 22% of students agreed that they had access to social media, whereas 71% of students do not have access to social media. The results show that students who had access to social media achieved a mean score of 523, while the mean score of those who did not have access to social media was 496. The results indicate that there is a significant difference between the achievement in English of students who have access to social media and those who do not. This can be seen from the bar diagram in Figure 108.

Figure 108 Achievement by access to social media

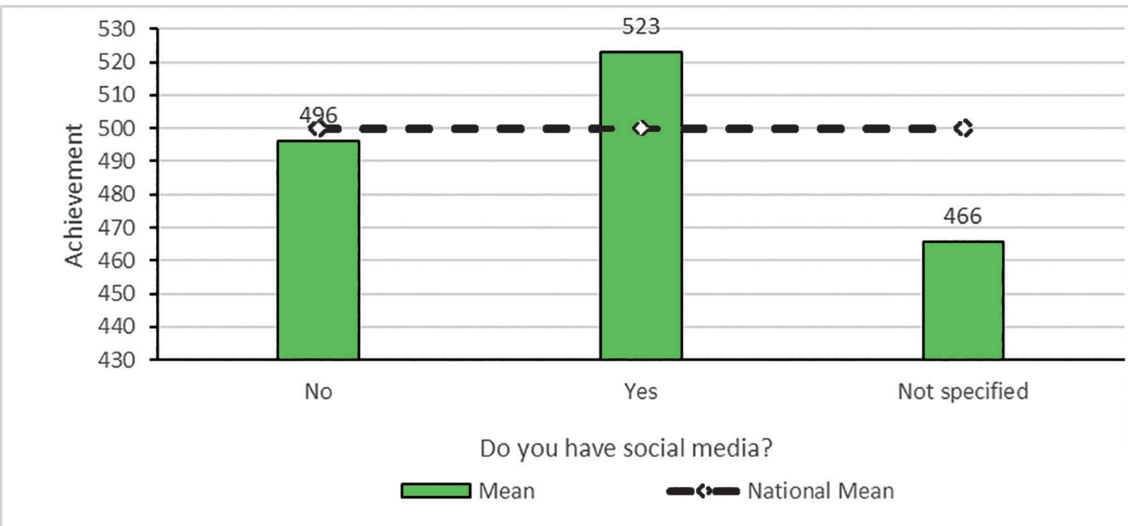


Figure 108 clearly shows that students who have access to social media perform better in English than the national average and the other options. However, those not having access to social media perform poorly in English. This may be because students who can connect on social media can communicate and share their problems with their peers and experts. Furthermore, students can solve their problems by means of different social sites like YouTube, Facebook and other subject-related online resources.

Students' attitude towards teachers

In NASA 2020, questions were also asked measuring the attitude of students towards their teachers. Thus, an attempt was made to record how positive (or negative) the students' attitudes towards teachers were. Their attitudes were determined on eight criteria: teacher loves, listens, gives physical punishment, treats us equally, tells us difficult things, gives homework, checks and gives feedback, teaches for the whole period. The attitudes of students toward teachers were measured on a Likert scale of totally agree, somewhat agree, somewhat disagree, and totally disagree. Accordingly, their attitude was compared with their achievement. The responses of students are presented in Table 134.

Table 134 Students' attitude towards teacher

Statements	Totally agree	Somewhat agree	Somewhat disagree	Totally disagree	Not specified
Teacher teaches with love and care	81.8	12	1.2	1.1	3.9
Most of the teachers really listen to me	47.6	33.7	5.4	3.7	9.6
Teachers do not give corporal punishment	35.4	26.2	12.2	14.5	11.8
Teachers treat us equally and care about us	79.2	9.1	2.6	2.2	6.8
Teachers answer when we ask when we do n't understand	87	5.1	1	1.1	5.8
Teachers give us homework	84.2	6.5	1.1	1.5	6.6
Teacher check and give feedback	79	10.5	1.9	1.5	7.1
Teachers teach for the full time of the period	65.1	18.9	4	2.5	9.4

The four categories were merged into two variables, agree or disagree, to generate a new table. Thus, whether the students' attitude towards the teacher is positive or negative is presented in the aggregated form in Table 135.

Table 135 Students' attitude towards school in aggregated form

Statements	Agree	Disagree	Not specified
Teacher teaches with love and care	94	2	3.9
Most of the teachers really listen to me	81	9	9.6
Teachers do not give corporal punishment	62	27	11.7
Teachers treat us equally and care about us	88	5	6.8
Teachers answer when we ask when we don't understand	92	2	5.8
Teachers give us homework	91	3	6.5
Teacher check and give feedback	90	3	7.0
Teachers teach for the full time of the period	84	7	9.3

Table 135 shows that the attitude of the learners towards the teachers was positive because most students agreed with the statements. However, the results show that around a quarter (27%) of students disagreed that the teacher does not give corporal punishment. Hence, the habit of corporal punishment should be avoided by some teachers.

Achievement by availability of English textbooks

In NASA 2020, students were asked about the availability of English textbooks. In response, 404,703 students agreed that they got English textbooks, while 37,539 students stated that they did not get a textbook. The achievement of students by textbook availability is shown in Table 136.

Table 136 Achievement by availability of English textbooks

Do you have English textbooks ?	Mean	N	Std. error of mean
No	471	37539	0.194
Yes	504	404703	0.07093
Total	501.4391	442242	0.06843

Table 136 indicates that students who got textbooks achieved a mean score of 504, which was above the national average, whereas those who did not get textbooks scored 471, which was 33 scale score below the former and below the national average. The results indicate that availability of textbooks remains crucial to achieving higher scores. This has been illustrated as a bar diagram in ~Figure 109.

Figure 109 Achievement by availability of English textbooks

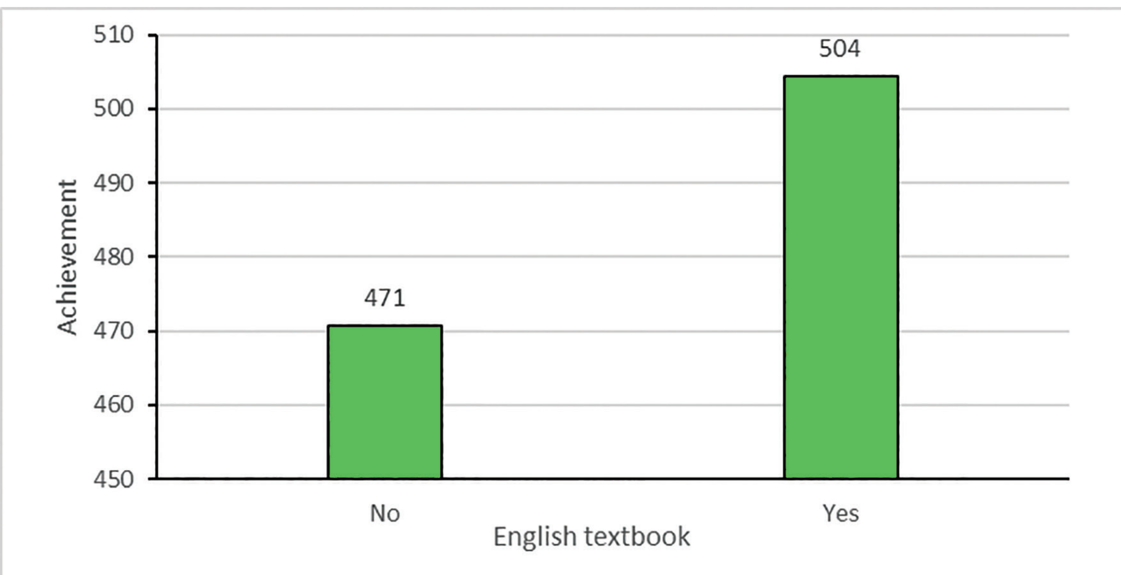


Figure 109 shows that students who had textbooks have higher achievement (504) than the national average. However, the score is poor (471) among students who did not have English textbooks. Therefore, relevant stakeholders should focus on timely provision of textbooks.

Achievement by homework provided

The practice of giving homework to students and teachers giving feedback is essential in the teaching–learning process. In NASA 2020 stuy, the sample students were asked whether their teacher gave homework. Accordingly, three categories of student response were identified: (i) those who always received homework from teachers; (ii) those who sometimes received homework; and (iii) those who never received homework. The data is shown in Table 137.

Table 137 Achievement by frequency of homework

Does your teacher give homework ?	Mean	N	Std. error of mean
Always	502	346439	0.07586
Sometimes	501	93314	0.15555
Never	477	5538	0.67876
Total	501.2743	445291	0.06807

Table 137 shows that 346,439 got regular homework from their teachers, whereas 93,314 students were given homework only sometimes by their teachers. However, 5,538 students never got homework from their teachers. The mean scores of the students based on their homework are shown in Figure 110.

Figure 110 Achievement by the frequency of homework given

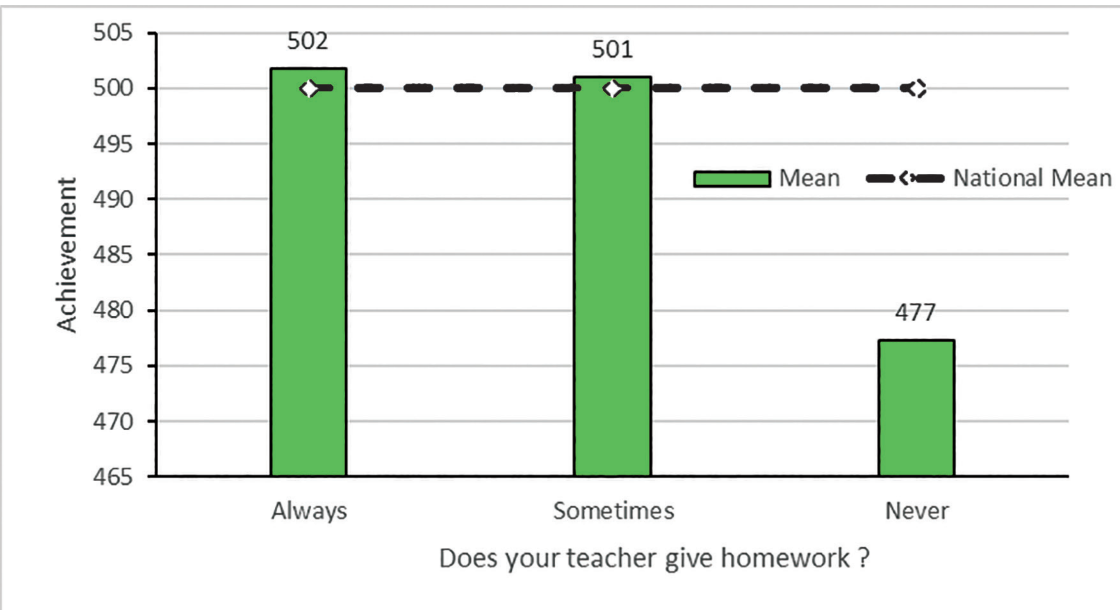


Figure 110 shows that those students getting homework from their teacher either always or sometimes perform better than the national average. However, those never getting homework have poor performance. Hence, every teacher at school level should give appropriate homework to their students for enhancing English performance.

Achievement by feedback on homework

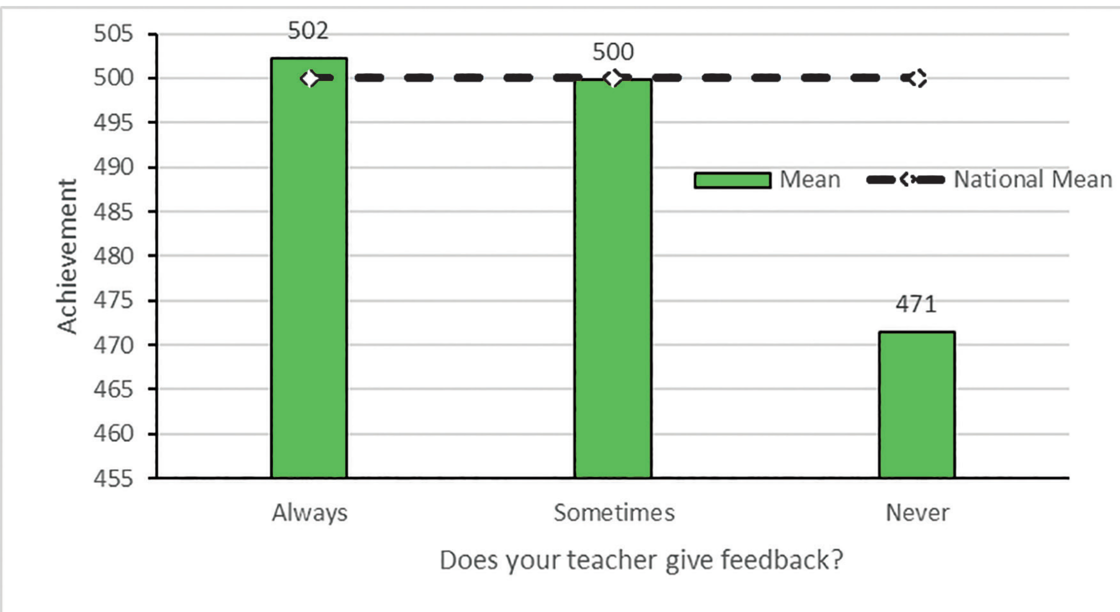
NASA 2020 also enquired about teachers' feedback on homework in relation to students' achievement. Research indicates that teachers regularly checking and giving feedback on the students' classwork, homework, project work and tests plays a creditable role in improving students' learning performance (Dahal, 2019, p.76No). Thus, students were questioned about how often their teacher gave feedback on homework. The responses and the associated achievement scores are shown in Table 138.

Table 138 Achievement by frequency of feedback on homework

Does your teacher give feedback ?	Mean	N	Std. error of mean
Always	502.2594	348150	0.07616
Sometimes	499.7781	90187	0.15367
Never	471.4774	5920	0.63347
Total	501.3455	444257	0.06809

Table 138 shows that 348,150 students regularly got feedback from their teachers, while 90,187 students were given feedback by their teachers sometimes. However, 5,920 students never got feedback from their teachers. The mean score of the students based on feedback on their homework is shown in Figure 111.

Figure 111 Achievement by frequency of feedback on homework



The data presented in Figure 111 reveals the positive relationship between regular checking and feedback of homework and students' learning. The mean scores of students who received feedback on their homework either regularly or occasionally were 502 and 500 respectively. In contrast to this, the mean score of students who never received feedback on their homework was 471. The difference between the highest and lowest mean scores is 31 scale score, which is statistically significant at $p < 0.05$. In conclusion, regular feedback on homework has a positive impact on achievement in English.

Use of different types of reference materials

Students were asked to state their use of different types of reference materials other than textbooks, in NASA 2020. The students' responses are presented in Table 139.

Table 139 Frequency of use of different types of reference materials

Type of reference materials used	Frequency	Percent
Old questions	205407	44.9
Guess paper	22579	4.9
Guide	55831	12.2
Dictionary	125709	27.5
Total	409526	89.6
System missing	47607	10.4
	457133	100.0

Table 139 shows that around half (44.9%) of students were in the habit of using old questions, and around a quarter (27.5%) used a dictionary as reference material in English. However, the number of students using guess paper was found to be very low (4.9%).

Students attitude towards school behaviour

Students' attitude towards school behaviour can also have an impact on the achievement of students. Students' responses were categorized as whether the student likes to come to school or not; whether peers treat them fairly or not; whether school facilities like playing, drinking water and toilets are good or not; and whether students participate in child club and children's programmes or not. The results of their responses are presented in Table 140.

Table 140 Students like or dislike of school behaviour (percentage)

Statement	Totally Agree	Somewhat Agree	Somewhat Disagree	Totally Disagree	Not Stated
I like to come to school.	91.1	4.1	0.9	1.2	2.7
Peers treat me fairly as far as possible.	71.6	19	2.1	1.5	5.8
School: School facilities like playing, drinking water and toilets are good.	81.8	9.3	2.1	1.7	5.1
I participate in child club and children's programmes.	54.5	27.2	4.4	6.6	7.3

Table 140 shows that almost all students totally agreed that they like going to school (91.1%), and the majority of students totally agreed about how they are treated by peers (71.6%) and the good condition of institutional facilities like playing, drinking water and toilets (81.8%). However, only around half (54.5%) of participants totally agreed that they participated in child club and children's programmes.

Bullying in schools

Any type of physical and mental mistreatment of students by other students at the school is known as bullying. Students should feel secure in the school environment. In cases where students are bullied, their achievement and attitudes towards school become negative. Among the many possible bullying activities, stealing, being beaten by friends, misbehaving, teasing, being isolated and being forced to do unwanted things were categorised. Students' responses are presented in Table 141.

Table 141 Types of bullying activities and students' responses of their experience

Bullying	Number of students %	
	No	Yes
1. Some of my belongings were stolen.	80.6	19.4
2. I was beaten by friends.	86.3	13.7
3. Friends have misbehaved towards me.	87.3	12.7
4. Friends have teased me.	79.4	20.6
5. I was left on my own.	89.8	10.2
6. I was forced to do unwanted things.	73.7	26.3

Table 141 shows that most students have not experienced bullying at school (73.7–89.8%). However, around a quarter of students are forced to do unwanted things (26%), around one-fifth of students have had something stolen (19.4%) and a similar number have been teased by their friends (20.6%), which should all be minimised.

The data was also analysed by summing up all six categories of bullying into a variable, and the results obtained are presented in the graph in Figure112.

Figure 112 Status of bullying in schools

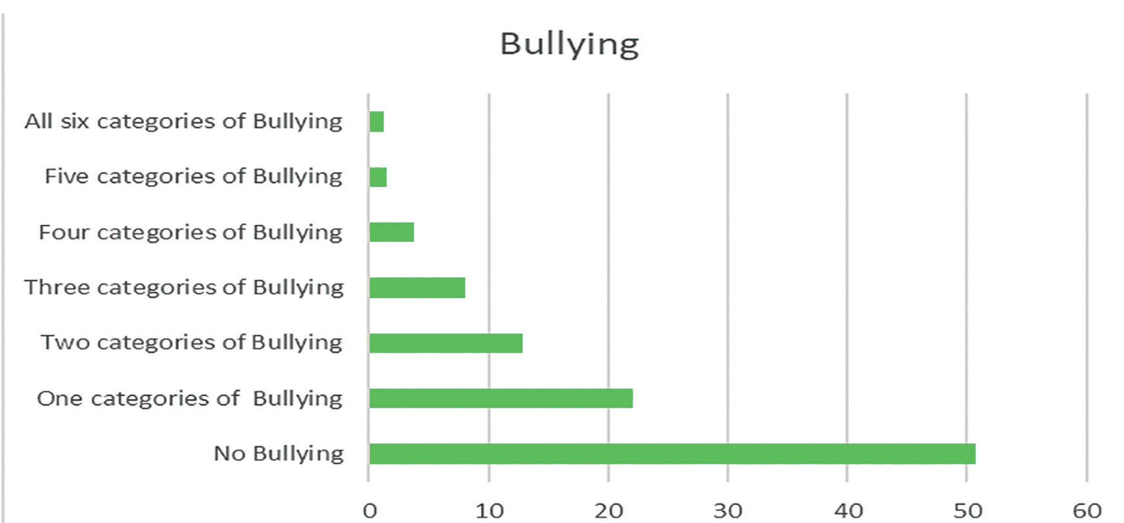


Figure 112 reveals that 51% of students believed that there was no bullying in school. Of the total students, 22% had experienced at least one type of bullying, while 12% had suffered at least two forms of bullying in school. The results show that 2% of students suffer from all five or six types of bullying in school.

Scholarship

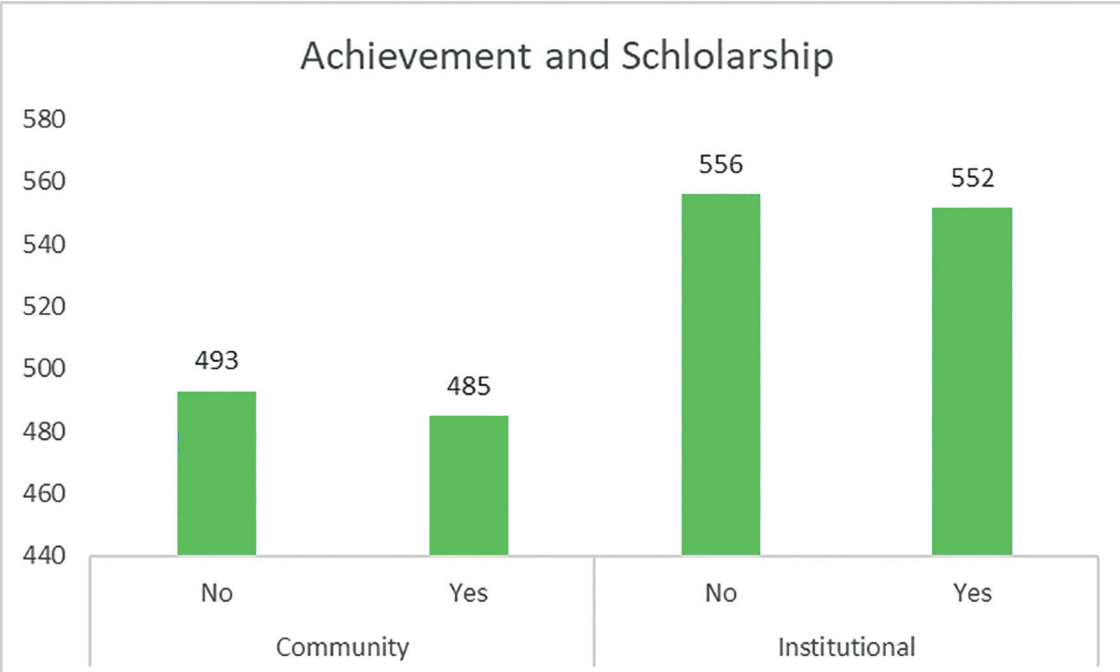
Scholarships are financial assistance to students provided by the school, to promote equal opportunity to poor, intelligent, Dalit and female students. Limited cash amounts are provided to students to buy their books, exercise book and pens. In NASA 2020, students were asked whether they had received a scholarship. The numbers of students who did and did not receive scholarships are given in Table 142.

Table 142 Number of students who received scholarships by type of school

Type of school	Scholarship	N	% of students
Community	Scholarship not received	180708	49.7
	Scholarship received	169231	46.5
	Not stated	13853	3.8
	Total	363791	100
Institutional	Scholarship not received	69375	85.1
	Scholarship received	10763	13.2
	Not stated	1381	1.7
	Total	81519	100

Table 142 shows that around half (46.5%) of students are getting scholarships from their institution. However, that rate is very poor in institutional schools (13.2%), which may be because the Government of Nepal is giving scholarships to all girls and Dalit students at school level, whereas the institutional schools have provision to give scholarships to 10% of enrolled students in the school. The achievement of students who did or did not receive a scholarship is shown in Figure 113.

Figure 113 Achievement scholarship status



Sufficiency of scholarship for exercise books and pen

Students were asked whether the amount provided by scholarships was sufficient or not to buy the necessary books, exercise books and pens. Their responses were categorised as sufficient, not sufficient and not stated. The details of the responses are presented in Table 143.

Table 143 Sufficiency of scholarship for exercise books and pen purchase

School type	Scholarship	N	%
Community	Sufficient	262279	72.1
	Not sufficient	48498	13.3
	Not stated	53013	14.6
	Total	363791	100
Institutional	Sufficient	43507	53.4
	Not sufficient	11323	13.9
	Not stated	26689	32.7
	Total	81519	100

Table 143 shows that the around three-quarters (72.1%) of students from community schools feel that the scholarship is sufficient for purchasing exercise books and pens.

However, that rate is only 53.4% in the case of institutional schools. Furthermore, a similar number of students from community schools (13.3%) and institutional schools (13.9%) feel that the scholarship amount is not sufficient for purchasing pen and exercise books.

Frequency of listening and speaking activities in English

Listening speaking, reading and writing are equally important skills in learning a language. English as a foreign language students need to learn this language through participating in listening and speaking activities frequently. One question students were asked was how often their teacher conducted listening and speaking activities in English. Their responses are presented in Table 144.

Table 144 Frequency of listening and speaking activities in English

School type	Frequency	Per cent
Community	Once a week	52.1
	Once a month	11.2
	3–4 times a year	8
	Never	5.4
	Not stated	23.3
	Total	100
Institutional	Once a week	57.2
	Once a month	17.7
	3–4 times a year	9.7
	Never	5.3
	Not stated	10.1
	Total	100

In Table 144, it can be seen that around half of students from both community schools (52.1%) and institutional schools (57.2%) replied that their English teachers conduct listening and speaking activities in school. However, the percentages of students stating that their teacher never conducts listening and speaking activities are similar and very poor in both types of school (5.4% in community and 5.3% in institutional).

Use of teaching materials

NASA 2020 also enquired about the use of teaching materials, as they make classes lively and facilitate understanding of the subject matter. Students were asked the frequency that materials were used. The responses of the students were categorised as always, 3–4 times a week, 1–2 times a week, never use, not stated. The results are presented in Table 145.

Table 145 Use of teaching materials in the English classroom

School type		N	Per cent
Community	Always	106255	29.1
	3–4 times a week	61462	16.8
	1–2 times a week	63615	17.4
	Never use	44459	12.2
	Not stated	89597	24.5
Institutional	Always	26643	32.4
	3–4 times a week	16438	20
	1–2 times a week	18800	22.9
	Never use	11381	13.9
	Not stated	8845	10.8

Table 145 shows that only around a third of students from community schools (29.1%) and institutional schools (32.4%) replied that their teachers always use teaching materials in their classroom teaching, and 12.2% of community school teachers and 13.9% of institutional school teachers never use such materials in their teaching. Hence, teachers should be encouraged to use appropriate materials in their instructional practice.

Library use in school

Library study also plays a vital role in enhancing the achievement of students. Students from both community and institutional schools were asked how often they used the library. The alternative answers were categorised as always, 3–4 times a week, 1–2 times a month, never uses and not stated. The details are presented in Table 146.

Table 146 Library use in school

School type	Use	Number	Per cent
Community	Always	133590	36.7
	3–4 times a week	103350	28.4
	1–2 times a month	57033	15.7
	Never uses	20998	5.8
	No library	35288	9.7
	Not stated	13532	3.7
	Total	363791	100
Institutional	Always	18138	22.2
	3–4 times a week	18998	23.3
	1–2 times a month	17092	21
	Never uses	8187	10
	No library	17058	20.9
	Not stated	2046	2.5
	Total	81519	100

From Table 146 it can be seen that more students from community schools (36.7%) always use the library than their institutional school counterparts (22.2%). However, 9.7% of community school students and 20.9% of institutional school students stated that they have no library at their school. Furthermore, 5.8% of community school students and 10% of institutional school students never use the library.

Item-wise analysis of reading and writing in English

Reading

Task development

The reading tasks used in the NASA grade 8 test were developed by teachers from community schools in the Kathmandu valley. They were provided with the Grade 8 Language Assessment Framework (LAF) and asked to produce tasks based on it. The 2019 Grade 8 LAF specifies seven types of reading tasks, namely: reading correspondence (personal and formal), reading for orientation, reading for information, reading instructions, reading figures to retrieve information, reading forms, cheques, timetables and CVs, and reading different literary texts. Before the field trials took place, the tasks produced by the teachers were reviewed by an expert team and then approved by the ERO English Subject Committee led by a university professor.

Field trials

Six sets of test papers containing a total of 72 reading and writing tasks were trialled. The reading texts came from a variety of sources, including letters, personal accounts, sports rules, bar diagrams/tables/pie charts, timetables, maps, curriculum vitae, instructions, job advertisements, dialogues, forms, popular scientific texts, newspaper articles, travel diaries, recipes and film reviews. The length of the texts varied from 16 to 220 words.

After piloting, a number of tasks were selected based on the difficulty and discrimination indices resulting from the field trial. These tasks were then submitted to the English Subject Committee for approval before being used in the final test.

Final test administration

The English reading and writing tasks were presented in the same test booklet as a background questionnaire, for which approximately 30 minutes was provided, and test papers measuring the Nepali language.

In total, there were five sets of English tasks each comprising two reading and two writing tasks, which were administered across the seven regions of Nepal. The reading tasks were alternated with the writing tasks (that is, one reading, one writing, one reading, one writing task) and followed an alternating pattern across the test sets so that each reading task appeared in two test booklets as shown in the table below.

Table 147 Reading test booklets

Set	Task	Original task levels*	Text type	Number of words	Totals
Set 1	DP AU	2 4	Letter Bar chart	117 16	**133 words**
Set 2	AU CZ	4 3	Bar chart Information leaflet	16 157	**173 words**
Set 3	CZ AS	3 6	Information leaflet Curriculum vitae	157 220	**377 words**
Set 4	AS HDS	6 5	Curriculum vitae Timetable	220 181	**401 words**
Set 5	HDS DP	5 2	Timetable Letter	181 117	**298 words**

* estimated by the test developer and based on the LAF

Findings

Each test set contained two reading tasks consisting of either six or seven items in total.

Set 1

Table 148 Task 1: Focus – SIID; Test method – SAQ (sentence completion)

Item no	Facility value (%)	Missing cases (%)
Q1a	27.8	14.94
Q1b	35.1	17.12
Q1c	37.6	18.14

The facility values in this task ranged from 27.8% to 37.6%, resulting in an average facility value of 33.5%. The final item (Q1c) had a high Infit Mean Square value (1.22) which suggests that the answers to this item were unpredictable. Such a finding suggests that some of the weaker test takers may have answered the item correctly, and some of the stronger ones incorrectly. The discrimination index was also lower on this item than the others at .51, though still within acceptable parameters.

The percentage of missing cases ranged from approximately 15% to 18%, which is higher than one would expect and should be investigated. By convention, items or tasks with 10% or more missing cases should be examined further, as such a finding would suggest some type of problem. For example, the answer may not be in the text, the item is at an inappropriate level of difficulty, there is an error in the key, an unfamiliar test method has been used, and so on. Another possible reason could be that the task required the test takers to construct their own answers. However, the items in this particular task were relatively straightforward and could be answered using one or two simple words which focused on specific information and important details in the comparatively short text (117 words).

Table 149 Task 3: Focus – SIID; Test method – SAQ

Item no	Facility value (%)	Missing cases (%)
Q3a	44.01	14.7
Q3b	39.98	16.2
Q3c	18.72	20.5
Q3d	25.04	21.8

The facility values in Task 3 ranged from 18.7% to 44%, resulting in an average facility value of 31.9%. Q3c had a higher-than-expected Infit Mean Square value (1.12), suggesting that some of the answers to this item were unpredictable. In addition, the discrimination index was also lower than for the other items at .49, though still acceptable. An analysis of the item revealed that the test takers needed to employ mathematical skills to achieve the correct answer, and this could have led to construct irrelevant variance. In other words, something other than simply English was being tested. In addition, all the items targeted years in their answers, thus resulting in minimal construct and little face validity.

The percentage of missing cases ranged from approximately 14.7% to 21.8%, which is again higher than one would expect. The task required interpretation of a graph rather than comprehension of a text, which is a different skill from reading, and this could well be one of the reasons for this high percentage of no answers. Not all successful readers can interpret graphs. Thought should be given to using such graphs as the basis for writing tasks (with appropriate scaffolding) rather than reading ones.

Set 2

Table 150 Task 1: Focus – SIID; Test method – SAQ

Item no	Facility value (%)	Missing cases (%)
Q1a	45.83	10.92
Q1b	39.26	13.26
Q1c	18.43	15.77
Q1d	24.27	16.8

The facility values in this task ranged from 18.4% to 45.8%, resulting in an average facility value of 32%. The percentage of missing cases ranged from approximately 11% to 17%, which is again higher than one would expect and should be investigated. This task was the same as the one used in Task 3, Set 1, and thus required interpretation of a graph rather than comprehension of a text. This could well be one of the reasons for this high percentage of no answers.

Table 151 Task 3: Focus – SIID; Test method – MCQ

Item no	Facility value (%)	Missing cases (%)
Q3a	34.33	12.99
Q3b	41.05	13.92
Q3c	38.85	18.06
Q3d	33.1	16.63

This task's facility values ranged from 33% to 41%, with an average facility value of 36.8%. Two of the items (Q3a and Q3d) had higher-than-expected Infit Mean Square values (1.17 and 1.13 respectively), suggesting that some random guessing may have taken place. This possibility is supported by weaker discrimination figures (.34 and .37 respectively). In addition, a native speaker would be unable to answer Q3c, as it requires cultural knowledge of Nepal. This type of item should be avoided, as should the wording used in Q1a ('Which of the following … is not available …') as this does not reflect a real-world activity.

Once again, the percentage of missing cases is higher than it should be (ranging from 13% to 18%). The text, however, appears appropriate for grade 8 students. Perhaps this is a case of time management, given that the reading and writing sections of the English paper are combined with the Nepali test paper, and students may give preference to working on the latter and/or simply have insufficient time to complete all the tasks.

Set 3

Table 152 Task 1: Focus – SIID; Test method – MCQ

Item no	Facility value (%)	Missing cases (%)
Q1a	36.98	13.69
Q1b	42.04	12.37
Q1c	38.94	17.83
Q1d	30.87	19.04

The facility values ranged from 30.8% to 42% with an average facility value of 37.2%. One of the items (Q1a) had a higher-than-expected Infit Mean Square value of 1.20, suggesting that random guessing may have led to some of the weaker students answering the item correctly and possibly some stronger ones answering it incorrectly. This is supported by a weaker discrimination figure of .31.

Once again, the percentage of missing cases is higher than it should be (ranging from 12.4% to 19%). The text, which is the same as in Set 2, Task 2, however, appears appropriate for grade 8 students.

Table 153 Task 3: Focus – SIID; Test method – SAQ

Item no	Facility value (%)	Missing cases (%)
Q3a	22.95	16.8
Q3b	42.58	14.91
Q3c	37.43	16.14
Q3d	15.56	20.82

The facility values in this task ranged from 15.6% to 42.6% with an average facility value of 29.6%. The overall findings suggest that this task was rather difficult for many of the test takers, particularly Q3d, and the percentage of missing cases (14.9% to 20.8%) reflects this. An analysis of Q3d reveals that the wording was quite difficult. The fact that this task was placed at Level 6 by the original test developer and contained 220 words may well be an explanation for these findings.

Set 4

Table 154 Task 1: Focus – SIID; Test method – SAQ

Item no	Facility value (%)	Missing cases (%)
Q1a	21.26	18.13
Q1b	41.9	16.4
Q1c	37.51	17.27
Q1d	14.56	22.32

The facility values in this task ranged from 14.6% to 37.5% with an average facility value of 28.8%. As in the previous set (Set 3, Task 3), the overall findings suggest that this task was rather difficult for many of the test population, particularly Q1d. This is once more supported by the percentage of missing cases (16.4% to 22.3%). As above, the wording of Q1d was quite difficult, and the difficulty level (6) and length (220 words) doubtless contributed to the low facility values.

Table 155 Task 3: Focus – SIID; Test method – MCQ

Item no	Facility value (%)	Missing cases (%)
Q3a	47.04	17.06
Q3b	36.8	18.29
Q3c	33.13	23.7
Q3d	34.9	24.7

Task 3's facility values ranged from 33.1% to 47.0% with an average facility value of 38%. The findings reveal that this task was easier than many of the other tasks. Q3b, however, had a higher-than-expected Infit Mean Square (1.17), suggesting some random guessing may have taken place. The percentage of missing cases ranged from 17% to 25%, again higher than expected.

Set 5

Table 156 Task 1: Focus – SIID; Test method – MCQ

Item no	Facility value (%)	Missing cases (%)
Q1a	44.56	10.33
Q1b	41.21	11.19
Q1c	35.87	14.44
Q1d	34.51	16.32

This task's facility values ranged from 34.5% to 44.6% with an average facility value of 39.0%. Q1b had a slightly higher-than-expected Infit Mean Square value (1.11), suggesting some unpredictability in terms of responses. This is supported by a weaker discrimination figure of .33. Once again, the percentage of missing cases is higher than it should be (ranging from 10.3% to 16.3%).

Table 157 Task 3: Focus – SIID; Test method – SAQ

Item no	Facility value (%)	Missing cases (%)
Q3a	24.7	18.12
Q3b	33.09	18.93
Q3c	33.18	19.93

The facility values in this task ranged from 24.7% to 33.2% with an average facility value of 30.3%. The percentages of missing cases (18.1% to 19.9%) are all higher than one would have expected.

In summary, the facility values for the reading test items ranged from 14.6% to 47% with an average facility value of 33.7%, suggesting that the test takers struggled with this test. The percentage of missing cases varied from 10.3% to 25% across the five tasks, all higher than would be expected. Both these sets of findings need to be investigated.

Writing

Task development

As with the reading tasks, the writing tasks used in the NASA Grade 8 test were developed by teachers from community schools in the Kathmandu valley. They were asked to produce tasks based on it. The 2019 LAF specifies nine types of writing tasks, namely: writing paragraphs, form completion, personal letters, dialogues, leaflets, short news reports, short stories, interpreting charts/tables, and essays. Before the field trials took place, the tasks produced by the teachers were reviewed by an expert team and then approved by the ERO English Subject Committee led by a university professor.

Rating scale

Starting with a rating scale which British Council Nepal had used on a previous project, a group of test developers modified the descriptors in order to reflect the six levels of the LAF. It was agreed that, for practical reasons, a holistic as opposed to an analytic scale would be used. However, it was felt that the descriptors in the holistic scale should be based on those criteria of assessment generally found in analytic scales, for example those relating to content, organisation and layout, appropriate use and range of vocabulary, and grammatical range and accuracy. It was also decided that the rating scale should be multifunctional, that is, applicable to a range of different types of writing, such as email, letter, story and essay, rather than employing different scales for different types of output as had been the tradition for NASA Grade 8 tests.

The resulting draft rating scale consisted of seven bands, starting with 0 which equated to a script which was blank or held an insufficient rateable sample, to 6 which included such descriptors as ‘Wide range of vocabulary with appropriate use’, ‘frequent and appropriate use of simple cohesive devices including a few complex ones’ and so on.

Rater training

The raters involved in grading the scripts produced by the grade 8 test takers underwent training, during which a number of sample scripts were scored. Feedback revealed an overall consistency in the scores awarded, though there was some lack of clarity over how such words as ‘few’, ‘some’, ‘majority’, ‘creativity’, and ‘originality’ appearing in the scale should be interpreted.

Other issues which emerged during training included how raters should respond to those test takers who use their own native language rather than English in their answers. The conventional approach to this phenomenon is that such ‘chunks’ must be ignored as they do not provide any empirical evidence of a test taker’s ability to write in English; they simply reveal that they may have (though not necessarily) understood the task. Some raters also faced difficulty in using the new rating scale when attempting to grade the 'dialogue' task. This difficulty is quite understandable, as 'dialogue' tasks are normally associated with speaking rather than writing tasks.

In general, the feedback from those who were involved in using the rating scale revealed that the introduction of such a scale in their localised context was new but an excellent beginning in the process of standardising the grading of the NASA writing test. One rater wrote that although the scale is holistic in nature, it provides a wide range of criteria and guidelines on how to score content, organisation, grammar, language use and the mechanics of written performance. The rater added that from a practical point of view, the descriptors were few in number and consistent across the bands.

Field trials

Six sets of test papers containing a total of 12 reading and writing tasks were trialled. The types of writing tasks trialled included dialogue completion, leaflet, chart interpretation, form completion, news report, essay, story writing (based on textual or visual input) and a letter. Some of the tasks indicated word length, others did not. The range for those which did mention the required length varied from 50 to 200 words. After piloting, a number of tasks were selected based on the difficulty and discrimination indices resulting from the field trial. These tasks were then submitted to the subject committee for approval before being used in the final test.

Final test administration

The final test administration began with all test takers being required to complete a background questionnaire, for which approximately 30 minutes is provided. Five sets of tasks comprising two reading and two writing were administered across the seven regions of Nepal. The writing tasks were alternated with the reading tasks (that is, one reading, one writing, one reading, one writing task) and followed an alternating pattern across the test sets so that each writing task appeared in two test booklets as shown in the table below.

Table 158 Writing test booklets

Set	Task	LAF 2019 S. No.	Text type required	Number of words required
Set 1	Task 2	9	Essay	150
	Task 4	4	Dialogue	6 'exchanges'
Set 2	Task 2	4	Dialogue	6 'exchanges' 120
	Task 4	1	Paragraphs	
Set 3	Task 2	1	Paragraphs	120
	Task 4	3	Letter	100
Set 4	Task 2	3	Letter	100
	Task 4	7	Story	150
Set 5	Task 2	7	Story	150
	Task 4	9	Essay	150

Findings

Based on the six-point rating scale described above, the results were as follows.

Table 159 Set 1, Task 1, Essay (n=3072)

Score	Count	% of total	Cum. total
0	1748	40.6*	40.6
1	744	17.3	57.8
2	535	12.4	70.2
3	516	12.0	82.2
4	440	10.2	92.4
5	270	6.3	98.7
6	57	1.3	100.0

*28.7% = blank paper

Table 160 Set 1, Task 2, Dialogue writing (n=3072)

Score	**Count**	**% of total**	**Cum. total**
0	2707	62.8*	62.8
1	507	11.8	74.6
2	416	9.7	84.2
3	359	8.3	92.6
4	212	4.9	97.5
5	86	2.0	99.5
6	23	0.5	100.0

*53.2% = blank paper

The two tasks in Set 1 required the test takers to write an essay of 150 words and produce a dialogue with six exchanges. The findings show clearly that the test takers found this section of the English test paper difficult. Across the two tasks, more than 51% of the test population scored 0. Indeed, of those who were awarded 0 (blank paper or insufficient rateable sample), 28.7% and 53% of the papers respectively were completely blank. This finding is very worrying.

Table 161 Set 2, Task 1, Dialogue writing (n=4541)

Score	Count	% of total	Cum. total
0	2749	60.5*	60.5
1	658	14.5	75.0
2	472	10.4	85.4
3	352	7.8	93.2
4	199	4.4	97.6
5	98	2.2	99.7
6	13	0.3	100.0

*47.4% = blank paper

Table 162 Set 2, Task 2, Paragraphs x 2 (n=4541)

Score	Count	% of total	Cum. total
0	2734	60.2*	60.2
1	605	13.3	73.5
2	410	9.0	82.6
3	377	8.3	90.9
4	264	5.8	96.7
5	145	3.2	99.9
6	6	0.1	100.0

*44.4% = blank paper

The two tasks in Set 2 required the test takers to write a dialogue with six exchanges and two paragraphs of approximately 120 words. The findings are very similar to those of Set 1, with over 60% of the test population scoring 0. Indeed, of those who were awarded 0 (blank paper or insufficient rateable sample), 47% and 44% of the papers respectively were completely blank.

Table 163 Set 3, Task 1, Paragraphs x 2 (n=4448)

Score	Count	% of total	Cum. total
0	2645	59.5*	59.5
1	606	13.6	73.1
2	442	9.9	83.0
3	338	7.6	90.6
4	262	5.9	96.5
5	142	3.2	99.7
6	13	0.3	100.0

*46% = blank paper

Table 164 Set 3, Task 2, Letter (n=4448)

Score	Count	% of total	Cum. total
0	2696	60.6*	60.6
1	541	12.2	72.8
2	347	7.8	80.6
3	322	7.2	87.8
4	286	6.4	94.2
5	213	4.8	99.0
6	43	1.0	100.0

*48% = blank paper

The two tasks in Set 3 required the test takers to write two paragraphs of approximately 120 words and a letter of approximately 100 words. As with Sets 1 and 2, the results were very disappointing, with approximately 60% of the test takers scoring 0. Indeed, of those who were awarded 0 (blank paper or insufficient rateable sample), 46% and 48% of the papers respectively were completely blank.

Table 165 Set 4, Task 1, Letter (n=4184)

Score	Count	% of total	Cum. total
0	2483	59.3*	59.3
1	558	13.3	72.7
2	345	8.2	80.9
3	289	6.9	87.8
4	286	6.8	94.7
5	183	4.4	99.0
6	40	1.0	100.0

*44.4% = blank paper

Table 166 Set 4, Task 2, Story (n=4184)

Score	Count	% of total	Cum. total
0	2393	57.2*	57.2
1	433	10.3	67.5
2	367	8.8	76.3
3	398	9.5	85.8
4	343	8.2	94.0
5	214	5.1	99.1
6	36	0.9	100.0

*44.6% = blank paper

The two tasks in Set 4 required the test takers to write a letter of approximately 100 words and a story of approximately 150 words. As with the previous sets, the results revealed a high percentage of test takers scoring 0, in this case approximately 58%. Of those who were awarded 0 (blank paper or insufficient rateable sample), 44% of the papers in both tasks were completely blank.

Table 167 Set 5, Task 1, Story (n=4210)

Score	Count	% of total	Cum. total
0	2265	53.8*	53.8
1	516	12.3	66.1
2	382	9.1	75.1
3	368	8.7	83.9
4	362	8.6	92.5
5	262	6.2	98.7
6	55	1.3	100.0

*40.4% = blank paper

Table 168 Set 5, Task 2, Essay (n=4210)

Score	Count	% of total	Cum. total
0	1937	46.0*	46.0
1	665	15.8	61.8
2	445	10.6	72.4
3	399	9.5	81.9
4	419	10.0	91.8
5	275	6.5	98.3
6	70	1.7	46.0

*37% = blank paper

The two tasks in Set 5 required the test takers to write a story of approximately 150 words and an essay of 150 words. Once again, around 50% of the test takers scored 0 across the two tasks. Of those who were awarded 0 (blank paper or insufficient rateable sample), between 40% and 37% respectively submitted blank pages.

To summarise, over half the test population scored 0 regardless of which test set they were given, and in a high percentage of cases (29%–53%) they submitted blank test papers. This finding must be investigated, and solutions found as to why such a large number of test takers are not attempting this part of the English test paper.

Conclusions and suggestion of reading and writing

Reading

1. In English, only five reading tasks were used in the NASA Grade 8 test, each one appearing in two sets. This seems very few, and in most contexts would normally raise a question concerning test security. It is recommended that a larger number of tasks be field trialled in future, thus making it possible to include more in the final test administration.

2. Four out of the five tasks used in the final test had four items, one had three items. This means that a test taker's reading ability was measured on the basis of their performance on seven or eight test items. Given the different types of reading behaviour that a grade 8 student is expected to be able to carry out, this number is entirely insufficient. It is recommended that the testing of reading should emulate that of listening and include a minimum of six tasks and 18 items. In this way, there is a much better chance of obtaining a more valid and reliable picture of the students' reading ability.

3. The 2019 LAF, on which the NASA Grade 8 test was based, categorises reading according to text type rather than reading behaviour, which is the convention in most reading tests. The text is normally seen as a vehicle through which a reader's ability can be measured and should not be the decisive focus of the test specifications. It is recommended that the learning outcomes become the main focus of future reading task development.

4. The teachers who produced these tasks were given no special training in how to develop the reading tasks and this is reflected in their quality. It is strongly recommended that future test developers undergo extensive training in the different types of reading and how these can be measured in a valid and reliable way, and particularly in light of the proposed changes to the Grade 8 LAF.

5. In order to encourage the test takers to engage with the reading tasks, it is important that the texts should come from authentic sources. Such a move would also make it easier to generalise about the test takers' ability outside the local context.

6. All the reading items in the test focused on reading for specific information and important details (SIID); there were no items targeting main ideas comprehension or gist. In light of this, a very limited picture of the test takers' reading ability has been captured. Indeed, in one task (AU), all four items required test takers to give a 'year' (for example, 2013) as the answer. Not only does such a task suffer from face validity, but it also provides very few insights into the test taker's ability. It is important to target as many different types of SIID in one task as possible. The space allowed for SIID answers should also be limited to discourage test takers from answering in sentences, or indeed being put off from attempting the item

on the basis that a sentence might be required. It is recommended that a tabular approach, such as that used in the listening tasks, be adopted.

7. Three of the tasks involved short-answer items and two used multiple-choice items. This is a rather limited range of test methods. It is recommended that others such as multiple-matching type items should be explored in future task development and test administration (see the new LAF for an example of this type). In addition, it is generally acknowledged that including an example in each task helps the test takers to understand more easily what is required of them.

8. The length of the reading texts ranged from 16 words (chart-based task) to 220 words (curriculum vitae). This lack of consistency inevitably led to inequality in terms of the amount of reading each test taker had to process. Those given Set 1 only had to read 133 words, while those test takers given Set 4 had 401 words. Given the fact that all test takers had the same amount of time at their disposal, this is simply not fair. In addition, the assumed levels of difficulty were not equitable across the sets. For example, Set 4 had tasks supposedly targeting Levels 5 and 6, while in Set 1 the tasks were purportedly from Levels 2 and 4. It is strongly recommended that when assembling future test sets, the construct, the difficulty level, the length of the texts, the test methods, the topics and the contexts are all taken into consideration to ensure that one group of test takers is not disadvantaged.

9. The facility values for the reading items ranged from 14.6% to 47%, with an average facility value of 33.7%, indicating that the test takers struggled with this test. This could be due to a number of reasons: firstly, a lack of time. Test takers are only given 30 minutes to complete the two reading tasks, and this is on top of a battery of tests, including those focused on the Nepali language, English writing, as well as a background questionnaire. Such an array of tasks could well have led to test taker fatigue. Another reason could be due to lack of motivation; it is not clear what the test takers themselves gain from such a test administration. A third reason could be the amount of time available for the teaching of reading in the school classroom, the teachers' knowledge of the different types of reading behaviour and the materials at their disposal.

10. It is strongly recommended that all bar charts and graphs are removed from the reading section of the NASA 8 Test. The ability to interpret a graph is a different skill from the comprehension of a text. Instead, thought should be given to using such graphs as the basis for writing tasks (with appropriate scaffolding) rather than reading ones.

11. Those involved in grading the constructed response items (SAQ) should receive regular training in what is, and what is not, acceptable for each answer. There must also be a system in place for deciding how to deal with unexpected but acceptable answers so that the marking is standardised. Where alternative answers are found and

agreed upon, it is important that the markers go through all the previously marked answer sheets to check whether this answer has already been marked as wrong, and where this is found to be the case, it should be recorrected.

12. As with listening, to facilitate and standardise the marking of the reading tasks, thought should be given to adding a column to the short-answer key listing those answers which are not acceptable. This column should be included in the task by the test developer at the task development stage; other unacceptable answers should then be added to it in light of the field trial and final test administration. This will help speed up the marking process as well as contribute to the reliability level of the items concerned. Incorrect spelling and/or punctuation should not be penalised provided meaning has been conveyed.

Writing

1. The findings present a worrying picture of the test takers' writing ability as determined by the tasks used in the final test administration. Over half the test population scored 0 and, of these cases, between 29% and 53% submitted blank test papers. Such an outcome must be investigated and solutions found as to why this is the case.

2. These findings suggest that the way in which writing is currently being taught in the classroom would benefit from a review. For example, it would be useful to document what training the teachers have received in teaching this skill, what materials are available, what types of writing are being taught, how the output is graded and how much time is actually available for teaching writing in the school classroom. The answers to these questions should then be fed back into the task development cycle, thus making the NASA writing test more realistic and relevant to the abilities of the grade 8 students.

3. An analysis of those students scoring 2, 3 or 4 on the tasks revealed that the percentage was higher (30%–35%) on the task which focused on writing about something they had probably written on before, that is, a traditional event. The lowest percentage was on the task which required test takers to write about a specific kind of competition which may not have been familiar to all the test takers. As with speaking, it is vital that the contexts and the topics are equally accessible to all test takers, though the actual tasks must differ from those used in the classroom to avoid measuring memorisation.

4. It is strongly recommended that the current practice of combining the Nepali and English language papers into one test booklet should be reviewed, as this may lead test takers to working on those parts of the test which they perceive to be easier and/or more accessible, regardless of the test booklet order, thus leaving the English writing tasks until the end. This situation may also be exacerbated by the amount of time provided for the English test paper, which is one hour for two reading and two writing tasks. Perhaps a short

break could be provided between the background questionnaire and Nepali test papers, and the English test paper, so as to make the test administration as practical as possible.

5. In order to encourage the test takers to engage with the writing tasks, it is important that they should be seen as being authentic real-life tasks; in other words, they must have face validity. Such tasks as 'write two paragraphs' or 'write a dialogue' are unlikely to be seen as authentic tasks which might be useful for the test takers in their future schoolwork or careers.

6. All writing tasks should contain the kind of information a writer has in real life. To elaborate, when writing an email, one automatically knows who one is writing to and for what purpose. Such details are crucial, as this will impact on the type of language used as well as how much one writes (for example, compare personal and non-personal emails).

7. Tasks should also take into account the test takers' age in order to have face validity. Asking test takers to write a Story starting with Once upon a time is not realistic and is likely to impact on their enthusiasm for, and thus performance on, the task. Test takers need to be engaged with the task to perform at their best.

8. The field trial currently combines approximately eight reading and four reading tasks in each test booklet. According to the information on the outside of the field trial test booklets, test takers are provided with only 2 hours and 22 minutes to do all 12 tasks. It is questionable as to whether this is feasible, and casts doubt on the quality of the responses due to test taker fatigue, particularly with regard to the writing tasks which are placed at the end of each set. While it is acknowledged that field trials are expensive to run, this must be balanced with the need to obtain reliable data. It is recommended that the number of tasks be reduced in future field trials.

9. In the current test, the two English language writing tasks are alternated between two reading tasks. This makes little sense and puts a further cognitive load on the test takers, who must switch from one skill to another rather than focusing on one particular skill.

10. The rationale behind deciding which tasks to combine in each set for the final test administration is not apparent. In Set 2, test takers are required to produce a dialogue and two paragraphs, while for those who receive Set 5, it is a story and an essay. In addition, as the amount of time available for writing two tasks is the same (30 minutes), the number of words the test takers are asked to write must also be equal across the sets. Those test takers who were given Set 5 had to write 300 words, while those who received Set 2 had to produce 120 words and six 'exchanges'. Finally, the amount of scaffolding provided to the test takers, for example bullet points, should be the same. To summarise, the system followed in the final test assembly needs to be

much more transparent and accountable in order to ensure equitability across test sets.

11. A new six-point rating scale applicable for all task types (essay, article, story, etc.) was used for the first time in this administration, and feedback from some of the raters involved indicated it worked well in helping them to standardise the marks they gave. Some improvements were made to the scale in light of using it during the field and final test administrations.
12. It is vital that an annual rater training session is held for all those involved in grading the test takers' scripts. Examples of good, average and poor performances for each task should be identified for the raters to train on, and differences in the ratings awarded should be discussed to ensure that all raters are using the scale in the same way. In addition, this training should include a familiarisation exercise concerning the expected output of each task so that the rater does not impose their own expectations on the task when working individually.
13. Thought should be given to making the new six-point rating scale available to other stakeholders. For example, teachers could be encouraged to use it in the classroom with the tasks they develop. This in turn would help their students become more aware of the criteria of assessment involved in measuring their written performance.

Listening and speaking in English

Listening

Task development

Starting with the level-wise descriptors of the Language Assessment Framework (LAF) for Grade 8 English (2019), a group of test developers were asked to produce a range of listening tasks targeting the six levels described in that document. In developing the tasks, a number of other variables were taken into consideration. Firstly, the sound file, which would be the basis of the task, was discussed in detail, for example the source and type of sound files (conversations, monologues, etc.), the number of speakers, their accents and their speed of delivery (words per minute) when talking. Other issues such as how long the sound file should be, how many times the recording would be heard, and a list of appropriate topics was also examined.

Once decisions about these variables had been made, appropriate sound files were found, and the test developers then used several textmapping procedures (Green, 2017) to determine how best to exploit the content. A number of different test methods, which were felt suitable for grade 8, were identified. These included short-answer items (closed questions / sentence completion), multiple-choice items and a range of matching tasks.

It was agreed that each task would consist of a minimum of three items and one example, and that the test as a whole should have a minimum of 18 items. An example was included in each task for a number of reasons. Firstly, to make it clear what the test takers had to do in order to complete the item (for example, tick a box, write a few words and so on); secondly, to signal the kind of listening behaviour the task was focusing on, for example, listening for specific information and important details. Thirdly, to indicate the level of difficulty being targeted by the task.

Once the tasks had been developed and peer reviewed, the instructions for each task were standardised using short simple sentences. The amount of time which test takers would be allowed for reading the task before the sound file started was agreed upon, as was the time that test takers should have for completing their answers after listening to the second recording.

Five sets of tasks were developed along these lines and made ready for the field trial in early 2020.

Field trials

Prior to the field trials taking place, the procedures involved in administering the listening test was explained to a group of test administrators. This included a document outlining each step which had to be followed, as well as how to use the equipment for playing the sound files and how to help the students use their headphones.

For many test takers, this was the first time they had been asked to attend a listening test and, consequently, they were puzzled. The test administrator had to spend some time explaining what they were expected to do. For example, the test takers were not familiar with using headphones in such a context, nor familiar with the recording devices which operated using Bluetooth. This made them anxious and worried. Substantial numbers of test takers were also unfamiliar with the concept that they could write their answers while listening, and many waited until the end of the recording before they began, even though they were told they could listen and write simultaneously. This explanation was also made in Nepali to facilitate matters. Those test takers from private schools, where most follow English-medium instruction, were less anxious and did well.

In some school classrooms, the Bluetooth facility did not work well, and this necessitated the test takers moving to other rooms where it did work. Some rooms did not have power sockets to enable Bluetooth connectivity. All of the above led to unexpected and unwelcome anxiety and extra effort for those involved.

The marking of the test takers' listening papers was relatively straightforward as the answers were quite objective, and in the majority of cases it was clear whether a particular answer was correct or incorrect. Unfortunately, many answer sheets exhibited illegible writing.

The data resulting from the field trials was then analysed in order to determine which tasks worked the best from both a validity and a reliability point of view. The resulting tasks were then earmarked and assembled into two sets of six tasks for use in the final test administration in 2020.

Final test administration and marking procedure

Two sets containing six listening tasks with a total of 18 items were used in the final test administration. The test administrators, who had also been involved in the field trials, indicated that a large part of the students found the tasks quite difficult as they either differed from those used in their schools or because they were simply not exposed to classes which focused on listening skills. In addition, some of the sound files involved native-speaker voices which many test takers were not familiar with, and this probably added to the difficulty. Other feedback indicated that hearing the sound files twice helped the test takers to feel less stressed. One test administrator felt that the two sets were not equal in terms of the task demands (test methods, item difficulty, speakers' accents) and that fewer short-answer item tasks should have been used.

In order to standardise the correction procedure, the markers attended virtual training sessions run by ERO and the British Council. The markers worked individually from home using an answer key which was based on the original test developers' tasks as well as any alternative acceptable answers resulting from the field trial administration. When a possible new answer was identified, the markers contacted ERO or the British Council for advice, and a consensus was reached on whether it was acceptable or not. In reality, there were few such cases.

Findings

Of the two sets of listening tasks used in the final test administration, half of the students were given Set 1 and the other half Set 2.

Set 1 (n=1155)

Each test taker was required to complete six tasks totalling 18 items. Three of the tasks focused on measuring the ability to listen for specific information and important details (SIID), while the other three targeted main ideas and supporting details (MISD). All six tasks were made up of short-answer questions requiring test takers to provide answers using a maximum of four words.

Table 169 Task 1, Level 1: Focus – SIID; Test method – SAQ

Item no	Facility value (%)
Q1	24.6
Q2	18.6
Q3	30.4

The percentage of test takers answering the items correctly (facility value) in this task ranged from 18.6% to 30.4%, resulting in an average facility value of 24.5%. This suggests that despite this task being classified as Level 1 (in terms of difficulty), most test takers found it demanding. It is generally acknowledged that items with facility values of below 20% are generally considered too difficult for the test population to whom it has been administered (see Green, 2013). In addition, Q3 had a high Infit Mean Square value (1.16), suggesting that the responses to this item were unpredictable, possibly indicating that some of the weaker test takers answered it correctly while some of the stronger ones did not.

Table 170 Task 2, Level 2: Focus – SIID; Test method – SAQ

Item no	Facility value (%)
Q4	39.6
Q5	35.7
Q6	33.8

The facility values in Task 2 ranged from 33.8% to 39.6%, resulting in an average facility value of 36.3%. These findings reveal that the test takers found this task easier than the first one, despite the hypothetical difference in difficulty level.

Table 171 Task 3, Level 3: Focus – SIID; Test method – SAQ

Item no	Facility value (%)
Q7	28.1
Q8	16.4
Q9	27.9

The facility values in Task 3 ranged from 16.4% to 28.1%, resulting in an average facility value of 24.1%. Q7 had a high Infit Mean Square value (1.25). A review of the key used to mark Q3 revealed only one possible answer (one word), though there were two other possibilities. A re-mark of this item might improve its statistics. Q8's low facility value suggests that it might involve a different, or more complex, type of processing than the other two items in this task. Once again, its low facility value suggests it is not appropriate for this test population.

Table 172 Task 4, Level 4: Focus – MISD; Test method – SAQ

Item no	Facility value (%)
Q10	17.7
Q11	6.23
Q12	19.7

Task 4's facility values ranged from 6.23% to 19.7% with an average facility value of 14.6%. Taken as a whole, these findings indicate that this task is too difficult for the test population. Task 4 marks the first MISD task and this could be part of the reason for the depressed results, as could the fact that it was based on a conversation between two native speakers. In addition, one item (Q11) appears to be performing quite differently from the others. An analysis of the task reveals that there is a large gap between the answer to Q10 and that of Q11, and that test takers may have got lost or overwhelmed by the amount of input they had to process. At this targeted level of difficulty (approximately CEFR A2/A+), the tasks should provide adequate scaffolding so that this does not happen.

Table 173 Task 5, Level 5: Focus – MISD; Test method – SAQ

Item no	Facility value (%)
Q13	21.0
Q14	18.6
Q15	20.3

The facility values in Task 5 ranged from 18.6% to 21% with an average facility value of 20%. In addition, Q15 had a high Infit Mean Square (1.21), denoting unpredictable behaviour on this item. The findings as a whole indicate that this task is too difficult for the test population.

Table 174 Task 6, Level 6: Focus – MISD; Test method – SAQ

Item no	Facility value (%)
Q16	9.4
Q17	16.6
Q18	7.27

The facility values in Task 6 ranged from 9.4% to 16.6% with an average facility value of 11%. Q15 had a high Infit Mean Square (1.25), denoting unpredictable behaviour on this item. Also, Q18 contained an error in the wording of the item; ‘mackerel’ was used instead of ‘mako sharks’. Once again, the findings indicate that this task is too difficult for the test population.

To summarise, the average facility values of the tasks in Set 1 range from 11% to 36.3%, confirming that the test takers found this test difficult. If it is accepted that the tasks reflect the learning outcomes for Grade 8, then it follows that the test takers are not at that level. This could well be due to a traditional focus on reading and writing as opposed to listening and speaking in the school classroom. If the goal is that the students should reach the same ability level in all four skills, more listening materials, training in teaching listening, and more class time must be devoted to developing listening skills.

Set 2 (n=1159)

In Set 2, each test taker was required to complete six tasks totalling 18 items. Three of the tasks focused on measuring the ability to listen for specific information and important details (SIID), while the other three targeted main ideas and supporting details (MISD). Five out of the six tasks required test takers to provide short answers in a maximum of four words; the other task employed a multiple-choice test method.

Table 175 Task 1, Level 1: Focus – SIID; Test method – SAQ

Item no	**Facility value (%)**
Q1	40.7
Q2	44.61
Q3	45.64

The facility values in this task ranged from 40.7% to 45.6%, resulting in an average facility value of 43.6%, indicating the test takers performed much better on this task than on any of those in Set 1. An analysis of the task suggests this might have been due to the presence of several words in the items which also appeared in the sound file and which undoubtedly helped lead the test takers to the answers. By convention, unless the items are meant to target recognition, as opposed to comprehension, this approach should be avoided.

Table 176 Task 2, Level 2: Focus – SIID; Test method – SAQ

Item no	Facility value (%)
Q4	39.7
Q5	46.4
Q6	30.97

The facility values in Task 2 ranged from 30.97% to 46.4%, resulting in an average facility value of 39.03%. The findings suggest that this was also an easier task than those in Set 1. An analysis of the task revealed that the answers were indeed quite simple, but that two of the items did not require sentence completion but instead focused on gaps in the item sentence. Completing a gap is much easier than completing a sentence, as test takers can use their grammatical knowledge (not just their reading ability) to complete it. This is why gap-filling exercises are much more prevalent in tests of grammatical knowledge.

Table 177 Task 3, Level 3: Focus – SIID; Test method – SAQ

Item no	Facility value (%)
Q7	31.67
Q8	15.27
Q9	34.17

The facility values in Task 3 ranged from 15.27% to 34.17%, resulting in an average facility value of 27.04%. Q7 had a high Infit Mean Square value (1.25), and Q8's low facility value was also problematic. (See also the findings on Task 3 in Set 1 which is the same task.)

Table 178 Task 4, Level 4: Focus – MISD; Test method – SAQ

Item no	Facility value (%)
Q10	21.23
Q11	5.18
Q12	19.24

Task 4's facility values ranged from 5.18% to 21.23% with an average facility value of 15.22%. These findings confirm that the task is too difficult for the test population. Once again, Task 4 marks the first MISD task, and this could be part of the reason for the depressed results, as could the fact that it was based on a conversation between two native speakers. In addition, one item (Q11) appears to be performing quite differently from the others. (See comments on Task 4 in Set 1 which is the same task.)

Table 179 Task 5, Level 5: Focus – MISD; Test method – MCQ

Item no	Facility value (%)
Q13	31.75
Q14	31.67
Q15	40.72

The facility values in Task 5 ranged from 31.67% to 40.72% with an average facility value of 34.7%. The higher facility values in this task are more than likely due to the fact that it was an MCQ task as opposed to a SAQ one, and thus required no writing.

Table 180 Task 6, Level 6: Focus – MISD; Test method – SAQ

Item no	Facility value (%)
Q16	31.23
Q17	24.68
Q18	4.4

The facility values ranged from 4.4% to 31.23% with an average facility value of 20.1%, once again denoting that this task is too difficult for this test population, particularly Q18. An analysis of the question indicates that the test takers would need to manipulate language in order to answer the item in four words. This must be avoided, as the task is targeting listening, not writing.

To summarise, the average facility values of the tasks in Set 2 ranged from 15.2% to 43.6%, confirming that the test takers found this test difficult. As with the findings in Set 1, if it is accepted that the tasks reflect the learning outcomes for Grade 8, then it follows that the test takers are not at that level, and actions must be taken to improve this situation if it is desired that Grade 8 students perform at the targeted level in listening.

Speaking

Task development

As with listening, the six sets of level-wise descriptors in the LAF for Grade 8 English 2019, were taken as a starting point for task development. The test developers also took into consideration a number of other variables. For example, issues relating to the prompt – whether it should be textual or visual in nature and the source of these input materials – were considered. Secondly, a list of suitable topic areas in line with the Grade 8 curriculum was drawn up and used as the basis for the speaking contexts the test takers would meet in the tasks. Thirdly, the number of tasks needed to obtain a reliable picture of a student's speaking ability was discussed. At first, it was felt that there should be six tasks, one for each of the levels outlined in the LAF. However, during a language testing workshop held in November 2019, it became clear that attempting to develop six distinctly different speaking tasks in terms of focus and difficulty was neither feasible nor practical: six speaking tests would take a long time to deliver and could well lead to test taker fatigue. A decision was therefore taken to collapse the six levels into three, each one of which would cover two levels, i.e. Task 1 would focus on Levels 1 & 2, Task 2 would target Levels 3 and 4, and finally Task 3 would measure performance at Levels 5 and 6.

In an attempt to assuage test taker anxiety at the beginning of the speaking test, it was also agreed that the oral should begin with a warm-up exercise, during which the interlocutor and test taker would introduce themselves and the test taker would answer some simple questions. It was agreed that this warm-up would not be assessed.

One final decision which needed to be made related to the length of time the test taker was expected to talk in each task and how much time they should be given to prepare. This information was added to the task instructions, which were written in simple clear English.

Five sets of tasks were developed along these lines and made ready for the field trial in early 2020.

Rating scale

In order to create an appropriate rating scale for the Grade 8 speaking test, a workshop with the test developers who attended the workshop in November 2019 was carried out. Each of the speaking tasks was analysed in turn against the six levels mentioned in the LAF, which they were designed to represent. Next a series of descriptors was drawn up to reflect each level as well as the learning outcomes. These descriptors were designed to measure the test takers' ability to respond appropriately to the task using a range of linguistic features. Once developed, the descriptors were trialled with different groups of raters, who listened to a range of recordings and then allocated an appropriate level. Findings revealed some differences in the scores, which were assigned so that the descriptors were reviewed to minimise these inconsistencies. Finally, the scale was shared with experts, whose feedback was taken into account by means of some final modifications.

The rating scale which resulted from this procedure consists of six levels: starting with 0 which corresponds to Doesn't respond to any question / cannot produce anything beyond one-word answers which are likely to be incorrect to 6 which equates to Answers all questions with confidence etc. Each level of the rating scale from 1 to 6 has between four and five descriptors detailing the expected test-taker performance at that level. The scale was designed so as to be applicable to all three speaking tasks.

Interlocutor frame development

Once the speaking tasks were finalised, a team of British Council and ERO members discussed the wording of the instructions the interlocutors should use when delivering the tasks, so as to ensure a consistent and reliable procedure. Such aspects as how much time the test takers would need to prepare a given task as well as how much time they would need to talk for in order for the rater to obtain a rateable sample was also confirmed. All of this information was included in the interlocutor frame, which was then trialled on a number of different groups of students.

Feedback from some of the interlocutors involved in the training workshop indicated that they felt the frame to be very comprehensible and that it provided a clear idea of how scores should be allocated, thereby making the system more accountable and reliable. It was thought that the frame provided a stepwise procedure for administering the speaking test.

In light of the workshop, the time allocations and the task instructions underwent a final review.

Interlocutor/rater training

For practical reasons it was agreed that one person would take on both roles of interlocutor and rater during the field and final test administrations of the speaking test. The people chosen to carry out this work were selected from two sources. Firstly, ERO nominated interlocutors from among their pool of test developers. These nominees had a master's degree in English language teaching, teaching experience, and previous experience in testing and rating. The British Council also identified a number of interlocutors who had been involved in similar types of test administration and rating activities and who had proven to perform well.

The interlocutors-raters were trained before the field trial took place. This training consisted of a detailed introduction to the speaking tasks which had been developed, in addition to familiarisation with the interlocutors' frame. This was effected through a series of group and plenary activities. A number of grade 8 students were invited, and the interlocutors were asked to trial the tasks on them individually. Members of the ERO and the British Council observed and provided feedback on the interlocutor performances.

A second workshop focused on how to rate test-taker performance in order to standardise the rating process. This began with a general discussion about the raters' previous experiences of rating, including the challenges they had met in the past. This shared experience led to a general checklist of what should and should not be done when rating. Further input was provided to the raters on how they should record the test takers' scores. Finally, a standardisation procedure was carried out by asking individual raters to grade a series of performances using the rating scale. Their results were then discussed in plenary, and each rater was asked to provide justifications for each of the scores they had allocated.

Feedback from the raters involved in the workshop suggested that in most cases assigning a level was straightforward, but in a few it was difficult to distinguish whether the awarded level should be a 3 or 4. One rater suggested that a two-day training workshop would have been better, so as to allow for more mock practice and thus further familiarisation with the procedures before the field and final test administrations.

Field trials

To ascertain the viability of the speaking tasks which had been developed, the tasks were field trialled on 300 grade 8 students in five districts of Nepal, namely Lalitpur, Kavre, Kaski, Rupandehi and Parsa. The trained interlocutor-raters adhered to the instructions laid down in the interlocutor frame, ensuring that every test taker in each part of the test had sufficient time to do the tasks. All performances were recorded for quality assurance purposes. The data resulting from these field trials was then analysed in order to determine which tasks worked the best from both a validity and a reliability point of view. The resulting tasks were then earmarked for use in the final test administration in 2020.

Final test administration

Two sets of speaking tasks containing three tasks each were used in the final test administration. Each of these sets was administered to approximately 1,150 grade 8 students located in 23 districts across the seven provinces of Nepal, by 75 trained interlocutor-raters. As mentioned, each speaking test started with a warm-up session where the interlocutor asked the test taker a series of simple questions. The function of this part of the oral examination was to try to put the test takers at their ease and to help familiarise them with the interlocutor's voice. Performance on the warm-up task was not graded.

Task 1 aimed at measuring the test takers' ability to describe a picture or to provide a series of instructions; Task 2 focused on asking test takers to explain a personal event or experience; and Task 3 required the test takers to provide their opinion on a particular topic.

Once again, all performances were recorded for quality assurance purposes.

Findings

Interlocutor feedback

Feedback from the interlocutors involved in the field trial indicated that many of the test takers were shy, not used to talking in English and so hesitated to speak. Some of them appeared to be unable to understand the instructions given in the interlocutor frame even though these were worded in simple English. Others seemed worried about how they would be marked and consequently kept quiet. The preparation time was felt to be enough for the test takers who understood the instructions; for those who were unable to understand, even after they had received a repetition of the instructions, the time was neither enough nor necessary for them. Some of the test takers said that they were clear about what they were supposed to do and indicated this by nodding their head or by saying 'yes', but when asked to speak, they did not speak at all. Feedback suggests that this was because they were not used to speaking in natural settings, that is, responding to real-life tasks such as those they encountered during the field trial and final test administration. Those students from a good learning environment spoke well.

Feedback from those interlocutors who were involved in the final test administration recommended that the complexity of the pictures used in Task 1 should be simplified. Test takers were required to talk about just three elements, but the picture used in the task contained a lot more elements and this confused some of them. One interlocutor felt that the two alternative tasks available in Task 1 were not of equal difficulty.

A second recommendation concerned the importance of ensuring that all test takers were equally familiar with the context used in the tasks, as this might impact on their performance. Other suggestions included the availability of supplementary questions for the weaker test takers who had difficulty in answering the task questions, and that where it was obvious a test taker was unable to answer, the interlocutor should move on to the next part of the test so as to minimise any possible embarrassment. Once again, the fact that the test takers are not accustomed to talking in English in the classroom was mentioned as influencing their performance in the test.

In addition, there was one report that some of the recording devices used in the speaking tests did not work well.

Test-taker results

Based on the six-point rating scale described above, the results were as follows:

Table 181 Task 1, Set 1: Focus Levels 1 and 2 – Describing a picture (n=1155)

Score	Count	% of Total	% Cum. Total
0	346	29.96	30
1	199	17.23	47.2
2	74	6.41	53.6
3	130	11.26	64.9
4	134	11.6	76.5
5	171	14.81	91.3
6	101	8.74	100

Table 182 Task 1, Set 2: Focus Levels 1 and 2 – Providing a set of instructions (n=1159)

Score	Count	% of Total	% Cum. Total
0	405	34.94	35
1	178	15.36	50.3

2	87	7.51	57.8
3	148	12.77	70.6
4	165	14.24	84.8
5	142	12.25	97.1
6	34	2.93	100

On average across the two speaking sets, grade 8 students performed as follows: 32.5% were awarded 0 on their speaking test; 16.3% scored 1; 7% scored 2; 12% scored 3; 12.9% scored 4; 13.5% scored 5 and 5.8% scored 6. In other words, approximately 49% scored 1 or less on Task 1 of the speaking test.

Table 183 Task 2, Set 1: Focus Levels 1 and 2 – Explaining a personal event/experience (n=1155)

Score	Count	% of Total	% Cum. Total
0	491	42.51	42.5
1	80	6.93	49.4
2	56	4.85	54.3
3	124	10.74	65.0
4	189	16.36	81.4
5	150	12.99	94.4
6	65	5.63	100

Table 184 Task 2, Set 2: Focus Levels 1 and 2 – Explaining a personal event/experience (n=1159)

Score	Count	% of Total	% Cum. Total
0	443	38.22	38.2
1	66	5.69	43.9
2	55	4.75	48.7
3	147	12.68	61.3
4	229	19.76	81.1
5	161	13.89	95
6	58	5	100

On average across the two sets, grade 8 students performed as follows: 40.4% were awarded 0 on their speaking test; 6.3% scored 1; 4.8% scored 2; 11.71% scored 3; 18.1% scored 4; 13.4% scored 5 and 5.8% scored 6. In other words, approximately 47% scored 1 or less on Task 2 of the speaking test.

Table 185 Task 3, Set 1: Focus Levels 1 and 2 – providing an opinion (n=1155)

Score	Count	% of Total	% Cum. Total
0	533	46.15	46.2
1	74	6.41	52.6
2	68	5.89	58.5
3	113	9.78	68.2
4	137	11.86	80.1
5	137	11.86	92
6	93	8.05	100

Table 186 Task 3, Set 2: Focus Levels 1 and 2 – providing an opinion (n=1159)

Score	Count	% of Total	% Cum. Total
0	490	42.28	42.3
1	61	5.26	47.5
2	67	5.78	53.3
3	120	10.35	63.7
4	145	12.51	76.2
5	184	15.88	92.1
6	92	7.94	100

On average across the two sets, grade 8 students performed as follows: 44.2% were awarded 0 on their speaking test; 5.83% scored 1; 5.83% scored 2; 10.1% scored 3; 12.2% scored 4; 13.8% scored 5 and 8% scored 6. In other words, approximately 50% scored 1 or less on Task 3 of the speaking test.

To summarise, approximately half the test population scored 1 or less on the three parts of the test. Interestingly, across the three speaking tasks and the two sets, the average scores varied very little: Task 1 had an average score of 2.2, Task 2 was also 2.2, while Task 3 averaged 2.15 on the six-point scale.

Conclusions and suggestion of listening and speaking

Listening

1. The first listening test administered by NASA focused on two different types of listening behaviour: listening for specific information and important details, and listening for main ideas and supporting details. It included a variety of different types of input, two different test methods and six tasks, each of which included an example and clear simple instructions. Each test taker was required to complete 18 test items. This was a good beginning.

2. The next test administration should build on this start by including tasks which measure other types of listening, such as listening for gist. It must also include a wider variety of test methods with less emphasis on constructed response tasks (short answer), as these are more demanding, even for more advanced listeners. Future tasks should include more multiple-choice tasks, using both visual as well as textual options (see LAF Grade 8 for an example of an MCQ task using pictures). In addition, multiple-matching tasks, such as matching speakers with comments, matching pictures with options and so on, should be added. Out of six tasks, it is strongly recommended that no more than two tasks require the test takers to produce short answers.

3. The statistical findings confirmed that the test takers found the listening test difficult. If it is accepted that the current tasks reflect the learning outcomes for grade 8, then it follows that the test takers are not at that level. This could well be due to a traditional focus on reading and writing as opposed to listening and speaking in the school classroom. If the goal is that the students should reach the same ability level in all four skills, more listening materials, training in teaching listening, and more class time must be devoted to developing listening skills. This is likely to necessitate practical training for the teachers so that they know how to teach these skills in the classroom.

4. The data also revealed that the test takers did less well on MISD items than SIID. This is not surprising, as MISD items tend to require higher levels of cognitive processing. Some of the sound files used in the test, however, were too difficult for grade 8 in terms of their speed of delivery, and more care must be taken to check this characteristic in future task development. Another reason for the weak MISD performance could be that SIID items are more prevalent in listening lessons and/ or in school textbooks. However, in real life both MISD and SIID, as well as gist, are important listening skills.

5. The use of authentic sound files taken from the internet should be encouraged (and made available) in the classroom, as opposed to the teacher simply using their voice. In addition to thus exposing the students to different accents, the speed of delivery is likely to be more natural than that used by the teacher, who may be overcompensating. Moreover, sound files taken from other sources will naturally include those with more

than one voice, for example conversations, chats, interviews and so on. In addition, thought might be given to training a group of test developers in how to develop sound files based on a series of ‘talking points’.

6. Although the listening test developers made a good beginning with these first listening tasks, the results reveal that more care must be taken in terms of the tasks’ difficulty level. The statistics show that nearly all of the tasks are too difficult for the current grade 8 test population. Some tasks also include items which are worded awkwardly and/or require written manipulation by the test takers in order for them to be answered within four words.
7. To facilitate and standardise marking, thought should be given to adding a column to the short-answer key, listing those answers which are not acceptable. This column should be included in the task by the test developer at the task development stage; other unacceptable answers should then be added to it in light of the field trial and final test administration. This will help speed up the marking process as well as contribute to the reliability level of the items concerned. Incorrect spelling and/or punctuation should not be penalised provided meaning has been conveyed.
8. Those involved in grading the constructed response items (SAQ) should receive regular training on an annual basis in what is, and what is not, acceptable for each answer. There must also be a system in place for deciding how to deal with unexpected but acceptable answers, so that the marking is standardised. Where alternative answers are found and agreed upon, it is important that the markers go through all the previously marked answer sheets to check whether this answer has already been marked as wrong and, if so, recorrect it.
9. It is recommended that more listening tasks be field trialled so that there is a larger bank of items from which to make a selection for the final test administration. In addition, more items should be developed per task (provided the textmapping results allow for this) so as to avoid any last-minute untrialled changes being made to the items prior to the final test. Moving forward, thought should also be given to establishing a bank of items which could be added to year-on-year, with the older tasks being retired for classroom use.
10. Those schools which are used for field trials or final test administrations must be checked prior to the test dates to ensure the availability of Bluetooth connectivity (if required), as well as working power points. Where there are difficulties, CD players or computers with loudspeakers should be made available for use.
11. Schools being used for field trials should receive advance notice of the types of tasks the test takers are to face in the test, so that they can share this information with the test takers prior to the test administration. It is recommended that this be done by way of example tasks.

12. It is important that future data entry makes a clear distinction between missing cases (no answer supplied) and incorrect answers, so that the appropriate conclusions can be drawn. The percentages of missing cases reported for listening were all under 5%, which is extremely unlikely as 11 out of the 12 tasks required short answers from the test takers. This procedure must be improved.

13. In order to encourage the teaching and learning of listening in the school classroom, it is vital that ERO continues to assess this skill in the NASA tests. Thought should also be given to increasing its weighting, which is currently 10%, so that it carries the same weighting as reading and writing. In terms of communication, listening and speaking are equally, if not more, important than reading and writing in today's world.

Speaking

1. The findings clearly indicate that the test takers did not find the speaking test easy, with nearly 50% scoring either 1 or zero on the three tasks. The last task was slightly more difficult (average score of 2.15 vs 2.2 on Tasks 1 and 2). It should be acknowledged that this was probably the first time that many of the test takers had participated in a formal oral examination, so it is not unexpected that the scores were depressed. In addition, many of the test takers were probably not used to talking in English to someone other than their own teacher.

2. Additional scrutiny of the speaking tasks against the grade 8 learning outcomes for that skill suggests that the demands of Task 3 were above the required level, and it is therefore recommended that this task be dropped from future test administrations. It is further recommended that Task 2 take the place of Task 3, Task 1 take that of Task 2, and that a new Task 1 be created which would target the easier learning outcomes, such as ask-and-answer questions (see the new LAF for an example of this).

3. Care must be taken to ensure that the requirements of the tasks are equitable across the sets so that all test takers have an equal chance of performing well. In addition, the situations chosen must be accessible to all, in other words not biased to particular regions or particular educational backgrounds.

4. It is recommended that each of the tasks have supplementary questions available which the interlocutor can use when it is clear that the test taker is struggling. This might not only result in some speech being elicited but might also encourage the test takers to feel better about their own performance. Where it is obvious that a test taker is unable to answer, the interlocutor should move on to the next part of the test so as to minimise any possible embarrassment.

5. An analysis of the test takers' performance by means of the oral recordings made during the test administration also suggests that many were anxious and shy during the

speaking test. The publication of sample tasks demonstrating the focus and demands of each type should help to assuage these levels of anxiety. In addition, access to the interlocutor frame and the rating scale once reviewed would also be beneficial for test takers, teachers and other stakeholders.

6. The use of authentic speaking tasks, such as those used in the NASA administrations, in the school classroom would also help students prepare for the kind of spoken English they are likely to need and meet in real-life contexts.

7. With regard to the performance of the rating scale, the percentage of test takers allocated scores of 2 or 3 were proportionally smaller than scores awarded to the other levels of the rating scale. This could be interpreted to suggest that these particular levels are not being used as frequently by the raters, possibly due to the wording in the rating scale. It is therefore recommended that the wording of the descriptors at levels 2 and 3 be analysed for vagueness and possible overlap with other levels so that rating scale works more efficiently. To this end, it may be useful to interview the raters who were involved in the final test administration to learn whether there were particular reasons as to why those scores were allocated less frequently.

8. The regular training of interlocutor-raters is crucial. This should occur every year before the field trial and final test administration takes place. Feedback suggests that this training should be a two-day event to allow more time for mock practice and further familiarisation with the procedures and the documents involved, such as the frame. It is vital that the interlocutors' level of English and pronunciation be verified as being of an appropriate standard (at least one level, and preferably two levels, above that of the targeted level) for the role. Where this is not the case, there is a danger that it might impede test-taker comprehension in terms of what the task requires and this impact on their performance.

9. For quality control purposes it is essential to check that all recording devices are in good working condition before the test administration begins.

10. In order to encourage the teaching and learning of speaking in the school classroom it is vital that ERO continues to assess this skill in the NASA tests. Thought should also be given to increasing its weighting, which is currently 10%, so that it carries the same weighting as reading and writing. In terms of communication, speaking and listening are equally, if not more, important than reading and writing in today's world.

Chapter 4

Classical Test Theory-Based Analysis

Introduction

This chapter is a classical test theory based reporting in which raw scores are presented. Those raw scores are not comparable to any year result and to any subject. These results are profile of student score and limited presentation of relationship with some demographic information.

Classical analysis in mathematics

Classical Test Theory-based reporting is a reporting in which the raw scores, rather than the true scores, are used to analyse the data. In this chapter, students' raw scores were summed and changed into percentages before reporting the results. In the following results, 'mean' represents the mean of the percentage of the student score. Note that this analysis of classic parameters should not be compared with the previous rounds of NASA results.

Classical results by set of questions in mathematics

The average score of the five databases was separately calculated. Since the average scores were different, they were not parallel. Student scores were transformed into percentages based on the maximum score of the individual sets. The average score of each five sets is presented in comparative form in Table 187.

Table 187 Results by set of questions in mathematics

Sets	Mean (%)	Std. deviation	Std. error of mean	N	% of total N
Set 1	29.6085	17.32682	0.05712	92022	19.90%
Set 2	25.9115	17.10953	0.05665	91229	19.70%
Set 3	24.1956	19.65292	0.06321	96658	20.90%
Set 4	23.083	18.22832	0.06049	90804	19.60%
Set 5	28.9063	21.48943	0.07087	91955	19.90%
Total	26.3284	19.0165	0.02796	462668	100.00%

Table 187 presents a comparison of the five sets of questions asked in NASA 2020. As can be seen, Set 1 was easier than any other set, and Set 4 was most difficult for the students. The national mean of all items used in the test based on student performance is 26 (in round numbers).

Classical results by type of schools in mathematics

In NASA 2020, the classical score in percentage varies by school type as per. The details are presented in Table 188.

Table 188 Classical results by type of school in mathematics

Type of school	Mean (%)	Std. deviation	Std. error of mean	N	% of total N	Significance	Partial correlation
Community	22.4876	15.97392	0.02638	366602	81.60%	Sig. at p <0.001	0.437
Institutional	43.9716	21.57575	0.07493	82911	18.40%		
Total	26.4503	19.0629	0.02843	449513	100.00%		

The classical score in percentage in Table 188 shows that Institutional schools are outperforming community schools by almost double the score. The data indicates that community schools achieved 22, whereas institutional schools achieved 44 scale score. The difference is 22 scale score in round numbers. The difference in mean was significant at $p < 0.001$.

Classical results by province in mathematics

In NASA 2020, a distinct variation was found in the classical score in percentage at province level. The national average classical score in percentage was 26 scale score. The details are presented in Table 189.

Table 189 Classical results by provinces in mathematics

Province	Mean (%)	Std. Deviation	Std. error of mean	N	% of total N	Significance	Partial correlation
Province 1	25.3976	17.64927	0.06846	66464	14.40%	Sig. at p <0.001	0.275
Madhesi	25.6141	19.89468	0.07262	75044	16.20%		
Bagmati	34.9979	21.24696	0.07624	77674	16.80%		
Gandaki	30.5999	19.73862	0.08374	55556	12.00%		
Lumbini	26.2677	18.20688	0.06489	78716	17.00%		
Karnali	18.8855	14.04461	0.06053	53835	11.60%		
Sudur-paschim	19.2902	13.97377	0.05938	55379	12.00%		
Total	26.3284	19.0165	0.02796	462668	100.00%		

The data in Table 189 shows that students' classical score in percentage in Bagmati was highest (34.99), whereas the mean classical score in percentage of Karnali was the lowest (18.88). The difference is around 15% in total. The difference in mean was significant at $p < 0.001$.

Classical results by gender in mathematics

In NASA 2020, variation in score was also analysed based on gender. The national average score based on gender was 26. Other details are presented in Table 190.

Table 190 Classical results by gender in mathematics

bq3sex	Mean (%)	Std. deviation	Std. error of mean	N	% of total N	Significance	Partial correl.
Boy	28.6162	19.77103	0.04323	209209	45.20%	Sig. at p <0.001	0.112
Girl	24.6156	18.21591	0.03753	235593	50.90%		
not specified	22.1253	17.15742	0.12836	17867	3.90%		
Total	26.3284	19.0165	0.02796	462668	100.00%		

In Table 190, the mean classical score in percentage of students by gender reveals that boys achieved 28.61 on average, whereas girls achieved 24.61. The difference is around 6% in total.

Classical Test Theory Results in Science

Classical results of science

In NASA 2020, five sets of questions were asked, and the mean score of the results shows the ease or difficulty level of these sets in science. These are presented in Table 190.

Table 191 Set-wise percentage of correct answers

Set	Mean	N	Std. deviation	Std. error of mean
Set 1	29.9789	89411	17.18055	0.05746
Set 2	30.3883	88602	18.08802	0.06077
Set 3	26.3748	94028	17.99165	0.05867
Set 4	26.8809	87991	17.15555	0.05783
Set 5	24.3698	88811	15.68316	0.05263
Total	27.5875	448844	17.39827	0.02597

The data in Table 191 reveals that the national mean classical score in percentage in science of all five sets is 28. When comparing the five sets, Set 5 remains most difficult and Set 2 seems easiest as the scores of both sets are 24 and 30 respectively.

Classical results by province in science

In NASA 2020, the classical score in percentage by province varies in science too. The responses of students by province reveal variations in learning opportunities in science. The national mean score and province-wise mean are presented in Table 192.

Table 192 Classical results by province in science

Province	Mean	N	Std. deviation	Std. error of mean
Province 1	26.3068	63150	16.44939	0.06546
Madhesh	22.1241	74217	16.73816	0.06144
Bagmati	36.7114	77240	18.9758	0.06828
Gandaki	32.6709	54274	17.56432	0.07539
Lumbini	28.8684	78012	16.20938	0.05803
Karnali	21.9096	51063	13.57025	0.06005
Sudurpaschim	21.6085	50888	13.86644	0.06147
Total	27.5875	448844	17.39827	0.02597

Table 192 shows that the national average mean province-wise classical score in percentage was 28 in science. The highest score (37) was for students in Bagmati, while students of Madhesh, Karnali and Sudurpaschim achieved the lowest mean score (22), which ws below the national average. The difference is 15% in total. The difference in mean was significant at p <0.001. The learning facilities and opportunities for students were not equal by province.

Classical results by type of schools in science

The data revealed that the classical score in percentage by school type also remained consistent (ERO, 2015; ERO, 2017; ERO, 2019) in that institutional schools achieved higher than community schools. The difference is shown in table 193.

Table 193 Classical results by type of school in science

Type of school	**Mean**	**N**	**Std. deviation**	**Std. error of mean**
Community	24.5475	365953	15.18345	0.0251
Institutional	41.005	82808	20.01318	0.06955
Total	27.5843	448761	17.39713	0.02597

Table 193 shows that community schools achieved 25, whereas institutional schools achieved 41, which is significantly different as institutional schools scored 16 scale score higher than community schools.

Classical results by gender in science

Gender-wise variation was also found in NASA 2020 and is shown in Table 194.

Table 194 Classical results by gender in science

Gender of the students	Mean	N	Std. deviation	Std. error of mean
Girls	26.8936	228601	17.17481	0.03592
boys	28.7159	202957	17.63486	0.03914
Not-stated	23.5164	17285	16.42246	0.12491
Total	27.5875	448844	17.39827	0.02597

Table 194 indicates that boys achieved a score of 29, whereas girls achieved 27, which is slightly lower than boys, while the gender-wise national average in science was 27.

Classical Results for Nepali

Classical results for Nepali

In NASA 2020, five sets of questions were asked, and the mean score of the results shows the ease or difficulty level of these sets, shown in Table 195.

Table 195 Classical results for Nepali

Table 186 Task 3, Set 2: Focus Levels 1 and 2 – providing an opinion (n=1159)	Mean	N	Std. deviation	Std. error of mean
Set 1	39.2872	90526	21.14018	0.070
Set 2	44.5048	95223	23.11715	0.075
Set 3	38.5556	93281	21.01927	0.069
Set 4	37.9884	87588	20.68745	0.070
Set 5	34.4317	88269	19.27555	0.065
Total	39.0371	454887	21.36703	0.032

Table 195 shows that the national mean classical score in percentage of all five sets is 39. When comparing the five sets, Set 5 remains most difficult and Set 2 seems easiest, as the scores are 34.43 and 44.50 respectively.

Classical results by provinces in Nepali

Province-wise mean scores vary, revealing variations in the learning opportunities in Nepali too. The national mean score and province-wise mean are presented in Table 196.

Table 196 Classical results by province in Nepali

Province	Mean	N	Std. deviation	Std. error of mean	% of student N	Sig.
Province 1	40.65	65567	21.56752	0.084	14	Sig. at p <0.001
Madhesh	30.35	74431	21.90568	0.080	16	
Bagmati	45.96	77323	19.93809	0.072	17	
Gandaki	43.78	54754	20.44475	0.087	12	
Lumbini	40.53	72307	20.40463	0.076	16	
Karnali	35.04	54444	20.51654	0.088	12	
Sudurpaschim	36.46	56060	20.12154	0.085	12	
Total	39.04	454887	21.36703	0.032	100	

Table 196 illustrates that the national average mean province-wise classical score in percentage is 39.04. Bagmati (46) and Gandaki (44) achieved higher mean scores than the national average, whereas Madhesh achieved the lowest mean score (30).

Classical results by type of school in Nepali

In NASA 2020, the classical score in percentage by school type also remained consistent with ERO, 2015; ERO, 2017; ERO, 2019, in that institutional schools achieved higher than community schools. The difference is shown in Table 197.

Table 197 Classical results by type of school in Nepali

School type	N	Mean	Std. deviation	Std. error of mean	Sig.
Community	363791	37.2493	21.00831	0.03483	Sig. at p <0.001
Institutional	81519	46.9205	21.12212	0.07398	

Table 197displays the fact that community schools achieved 37, whereas institutional schools achieved 47, which is significantly different as institutional schools scored 10 scale score higher than community schools.

Classical results by gender of the student in Nepali

Gender-wise variation was also found in NASA 2020. The mean classical scores in percentage of boys and girls are presented in Table 198.

Table 198 Classical results by gender of the student in Nepali

Gender of the student	N	Mean	Std. deviation	Std. error of mean	Sig.
Boys	205461	38.664	21.01021	0.04635	Sig. at p <0.001
Girls	229798	40.1289	21.48299	0.04481	

Table 198 shows that boys achieved a score of 39, whereas girls achieved 40, which is very slightly higher than boys.

Classical Test Theory for English

Classical results for English

The average score of the five databases was separately calculated. Since the average scores were different, they were not parallel. Student scores were transformed into percentages based on the maximum score of the individual sets. The average score of each of the five sets is presented in the comparative form in Table 199.

Table 199 Classical results for English

Set	N	Minimum	Maximum	Mean	Std. error	Std. deviation
1	90767	0	100	24.55	0.083	25.026
2	95541	0	100	24.22	0.075	23.265
3	93727	0	100	23.45	0.077	23.557
4	88283	0	100	24.23	0.083	24.736
5	88815	0	100	27.61	0.086	25.496
Total	457133	0	100	25	0.036	24.449

The data in Table 199 reveals that the national classical score in percentage in English of all five sets is 25. When comparing the five sets, Set 3 remains most difficult and Set 5 seems easiest, as the scores of both sets are 23.45 and 27.61 respectively.

Classical results by province in English

Province-wise mean scores vary, which shows variations in learning opportunities in English too. The national mean score and province-wise mean are presented in Table 200.

Table 200 Classical results by province in English

Provinces	N	Mean	Std. deviation	Std. error of mean
Province 1	66066	24.958	24.247	0.094
Madhesh	74431	19.634	21.864	0.080
Bagmati	77372	38.588	26.547	0.095
Gandaki	55189	34.239	26.493	0.113
Lumbini	72552	21.610	21.984	0.082
Karnali	54900	14.867	17.031	0.073
Sudurpaschim	56624	16.981	20.312	0.085
Total	457133	25	24.449	0.036

Table 200 shows that the average classical score in percentage was found to be high with respect to Bagmati (mean = 38.58, SD = 26.54) and Gandaki (mean = 34.23, SD = 26.49) in comparison with the other provinces. However, the scores were found to be poorest in

Karnali (mean = 14.86, SD = 17.03) and Sudurpaschim (mean = 16.98, SD = 20.44).

Classical results by type of school

In NASA 2020, the classical scores in percentage varied based on the school type. Schools funded by the government, known as community/public schools, achieved lower than the institutional schools funded by the private sector. The details are presented in Table 201.

Table 201 Classical results by type of school

School type	N	Mean	Std. deviation	Std. error of mean
Community	365388	17.938	19.231	0.032
Institutional	82106	55.790	21.287	0.074
Total	447494	25	24.490	0.037

From Table 201 it can be seen that the average classical score in percentage of students in English was found to be more than three times higher in institutional schools (mean = 55.79, SD = 21.29) than in community schools (mean = 17.94, SD = 19.23).

Classical results by gender of students

In NASA 2020, the classical scores in percentage varied based on the student's gender. The difference in student learning achievement is presented in table202.

Table 202 Classical results by student's gender in English

Gender	Mean	N	Std. Deviation
Boys	26.7	206568	24.512
Girls	23.7	230547	24.456
Total	25.12	437116	24.53

Table 202 reveals that male students scored 26.7% where girls scored 23.7%, the difference is 3% in raw score.

Chapter 5

Findings, Conclusions and Recommendations

5.1 Findings

The main finding of NASA 2020 is that the national average of students' achievement in to be improved subjects has lowered compared to NASA 2017. Students were struggling to acquire even minimum learning (NASA, 2017; NASA, 2019). The situation is worsened because the majority of students are unable to learn what is taught in all subjects in NASA 2020. The majority of students have achieved or mastered less than 50% of the curriculum in all subjects**.** Most of the students could not solve higher-order thinking items. Basically, there are problems in the teaching–learning strategy, remedial actions and the role of head teachers. On average, students in institutional schools have massively outperformed students in community schools. However, it is worth noting that the average scores of students in some community schools were the highest among all schools in all subjects. Deeper analysis of the reasons behind their success should be considered, as they can provide valuable lessons for other community schools and policy makers alike.

Province level

The comparative study of province-wise achievement in mathematics shows variations in the achievement level of the students. Overall, only 32.1% of students have shown their performance to be above the minimum level. The national average of NASA 2017 was 500, whereas the national average of NASA 2020 is 483, a decrease of 17 scale score. The achievement of students in Bagmati (509), Gandaki (497) and Lumbini (485) was, on average, better than in other provinces and was above the national average (483). Similarly, the national mean for science was 470, which was 30 scale score below the national mean in NASA 2017. Bagmati (492), Gandaki (485), Lumbini (475) and Province 1 (472) were high-performing provinces, whereas other provinces scored below the national mean in science. The achievements in Nepali of Province 1 (505), Bagmati (511), Gandaki (521) and Lumbini (498) students were distinctly above the national average. The disparity in achievement by province was much wider in English, though. The national mean was 500, and the achievement of Bagmati (529), Gandaki (522) and (Province 1 (501) students was above the national average. The performance of Madhesh, Karnali and Sudur Paschim was lower in all four subjects than the national average.

Gender

Learning disparity between boys and girls was one of the major findings in the study. There was a statistically significant difference between the achievement of boys (490) and girls (478) in mathematics. The difference in the achievement of boys and girls in science and English was also significant, but there was no noticeable difference in the achievement in Nepali, as boys scored 501 and girls scored 500. The achievement of boys was above the national average in science, mathematics and English, whereas girls performed below the national average in all these subjects except Nepali.

Age

A distinct variation in achievement was seen by age group as well. Most students who participated in the assessment were aged between 12 and 16 years participated in the assessments. Among them, students aged 13 years were the highest scorers in all four subjects assessed. Achievement scores for students aged 15 years or more was lower, on average. This result was consistent in all four subjects.

Home language

There was a significant difference in the achievement of students who use Nepali as their home language compared to the achievement of students who use other languages at home. The gaps between the achievement of students who used Nepali as a home language and those who used other languages at home were 15 scale score in Nepali, and 13 in English.

School type

The comparative study of achievement showed a vast gap between community schools and institutional schools. The institutional schools topped the community schools in students' achievement in all subjects. For instance, the achievement of students in institutional schools in science was 500, whereas community school students scored 463. This was a significant difference, with 37 scale score between them. Overall, the achievement of community schools was below the national average, whereas the achievement of institutional schools was distinctly above the national average.

Achievement by students' career aspirations

Based on their future goals, the study showed that students wishing to be civil servants or to work abroad, where in-depth learning is required, had higher achievement in subjects like mathematics and science than students hoping to be farmers, teachers or employees in private sectors jobs.

Parental education

Parents' educational level has a direct positive association with children's achievement in all subjects assessed. Based on students' achievement, it can be said confidently that on average the higher the educational qualifications of the father or mother, the greater the scores of the children. Educated fathers and mothers contributed significantly to their children's learning achievement, whereas children whose father or mother was illiterate performed lower. The achievement differs significantly from illiterate to literate parents, and from lower qualification to higher qualification of the parents. This result is consistent with the study carried out by Kainuwa & Yusuf (2013), who stated that children of fathers or mothers with university degrees perform particularly well and get the highest level of education.

Parental occupation

Analysis of the relationship between parental occupation and student learning showed that students' performance was highest for those whose parents were teachers. Students whose parents were involved in government jobs, business or handling only household work also had higher scores. Children whose father and mother were involved in agriculture and households, working in others' homes or handling only households had, on average, lower scores.

Family size

Family size was also seen to be an important predictor in the learning achievement of students. Students residing in households where the family size was four or five members had higher achievement scores. Beyond that, achievement decreased with additional family members.

Teacher's reliability

The regular presence of a teacher in the classroom indicates both dedication and awareness of the importance of delivering quality education to shape a bright future for students. Reliable attendance in the classroom means that teachers can give in-depth knowledge in the subject, and it is easier for them to complete the curriculum on time. Therefore, it is an important predictor in students' achievement. Thus, considering the findings above, teachers who were dedicating all their time in the classroom were successful in improving students' achievement. Meanwhile, the students of teachers who arrived late and went early, or who did not come to class at all, had poor performances.

Interest in subjects assessed

Developing a strong interest in a subject encourages the student to work harder in that subject, which helps boost their achievement. The findings show that the majority of students who enjoyed the different subjects investigated here wanted to learn and excel in those subjects.

Homework and feedback

Based on the analysis of data, any feedback after homework has boosted students' performance. In addition, feedback given on a regular basis was found to be more helpful. The performance of students who received regular feedback on their homework was higher than those who never received feedback. There is a significant difference in achievement between students who received regular homework and feedback and those who got neither homework nor regular feedback. This difference was 15 scale score in science and 34 scale score in Nepali. The mean score differences in science and Nepali were statistically significant, which indicates the importance of receiving homework and feedback regularly.

Home possessions

Variation was seen in achievement based on home possession of material goods such as a permanent house, car, motorcycle, TV and computer. For instance, out of 22,385 students in mathematics, 51% have a TV at home and only 43% of students have permanent houses, while 57% did not have computers, 52% did not have motorcycles, and 72% of students did not have cars at home. Similar findings were observed in other subjects as well.

Findings for Mathematics and Science

Students in institutional schools perform, on average, much better than community school students in mathematics. Though this is not a causal relationship, there are many who believe that institutional schools are more effective than community schools in improving student learning. Similarly, the relationship between socio-economic status and mathematics scores are positively correlated, but the magnitude is much smaller than that for institutional schools.

Female students are, on average, faring worse in mathematics than boys, and the difference is both substantial in magnitude and statistically significant. Similarly, student age and mathematics score are negatively correlated. Compared to Brahman and Chhetri students, Dalit students are doing significantly worse in mathematics. There is an expected positive relationship between father’s education level and the child’s achievement in mathematics.

There are some school-level variables that are also important. For example, students in schools where the head teacher is permanent have higher scores, on average, in mathematics. Similarly, students in schools where the mathematics teachers are permanent are also doing, on average, better than students where mathematics teachers are not permanent. This is perhaps an indication that these teachers and head teachers can focus more on teaching or administrative duties and not worry about other aspects related to their tenure status.

Findings for Nepali and English

Regarding community schools, English-medium community schools are performing better that Nepali-medium community schools. This has opened a grey area for further research, even though English medium does not mean a quality education.

Students in schools where the head teacher is a secondary -level appointee are performing better than others, and the difference is statistically significant. Similarly, students in schools that have introduced initiatives to reward teachers have also performed better in Nepali.

With regard to child-level characteristics, female students are doing worse than male students in Nepali, but the magnitude of the difference is substantially lower than in mathematics. Similarly, the age of the student and Nepali scores are negatively correlated, a finding consistent with mathematics. There is a positive relationship between having a dictionary and other educational reference books at home and achievement. The positive coefficient for dictionary and other educational reference books may be a proxy for these households prioritising education.

5.2 Conclusion

An educational system covers the input, process and output of education. Curriculum, pedagogy, teaching and learning practices and assessment are at centre-stage of attention for the formation, implementation, and monitoring and evaluation of educational policies. Rigorous research and evidence-based findings are the pillars for assessing the overall system of education. NASA has been making endeavours for assessment of the educational output of schools to be one of its core activities since its establishment in Nepal.

The main objective of this assessment was to prepare the baseline data for the School Sector Development Plan (SSDP), as well as compare the learning achievement of 2020 with the previous cycle of NASA (2017), to analyse how quality education in the school system has evolved over time. The study, as before, shows variation in the performance of province-level achievement in mathematics, science, Nepali and English. Bagmati, Gandaki and Lumbini are high-performing provinces, whereas Province 1, Madhesh, Karnali and Sudur Paschim are low-performing ones. The disparity seems greater in gender-based achievement, as boys have performed higher than girls.

The most appropriate age for learning grade 8 seems to be 12 or 13 years (starting grade 1 at age 5 or 6), as students in this age group, on average, achieved higher scores than other age groups. Students older than 14 years score lower, perhaps a reflection that these children are repeating grades or that children, presumably with less conducive learning environments at home, are starting school later.

A substantial difference in achievement has been observed based on the home language. Children whose home language is Nepali scored higher than those whose home languages were other than Nepali. This important finding has a notable influence on the use of classroom pedagogy and achievement of students, even in earlier grades.

The achievement of institutional schools is comparatively far better than community schools. Despite the investment of huge resources by the government, the achievement of community school students remained below the average level. Raising the quality of community schools has been one of the greatest challenges.

There is remarkable difference in the achievement of children from illiterate and literate parents, with a positive relationship between student achievement and parents with at least grade 8-level education. Similarly, parental profession also has a positive influence on the achievement of students. Scores were lower for students whose parents were involved in agriculture, household work and working for other households.

Children from a nuclear family have achieved higher scores, on average, than those from a blended family. Data shows that the greater the number of family members, the lower the achievement of students. Similarly, students with positive attitudes have succeeded in excelling academically by scoring good grades in various subjects. Likewise, teachers who dedicated all their time in the classroom were successful in improving the students' achievement.

Likewise, providing feedback on homework is leading to improvement in students' achievement. The availability of a table for study, a separate study room, a computer for school work, access to the internet, children's magazines, stories/poetry and pictures, a dictionary and reference books, and so on, at home contributes to boosting their learning performance. Lastly, permanent head teachers and teachers are associated with higher achievement scores. In the same way, permanent school buildings and infrastructures also positively influence learning as shown by the data.

5.3 Recommendations

1. **A large number of students are below grade level and an alarming gap exists between the intended and achieved curriculum.**

   When considering the proficiency levels of students in achievement, the results show their low level of ability with 32.1% in mathematics, 37.7% in science, 58.8% in Nepali and 51.5% in English of students having passed the basic proficiency levels, while 67% in mathematics, 62.3% in science, 41.2% in Nepali and 48.5% in English have achieved below the basic proficiency level. These achievement levels indicate students' poor competence levels, and that only a small number of students have the higher level of proficiency. The majority of students have achieved or mastered less than 50% of the curriculum in Math and Science. This evidence indicates an alarming gap between intended and achieved curriculum.

   Recommendation: The overall gaps of intended and achieved curriculum demand a radical change in the policy, resource management, curricular design and implementation process, and monitoring and evaluation strategies. Policy reformation, allocation of the required budget, activity-based curricula, emphasis on pedagogical delivery, and resource management are some of the strategies the government should implement instantly for removing the gaps between intended and achieved curriculum. Moreover, given that below-grade-level learning is already pronounced by grade 5, as previous administration of NASA at grade 5 has amply demonstrated, remedial education should be seriously considered in earlier grades. Furthermore, training curricula for teacher professional development (TPD) should be reoriented to better equip teachers to identify, and provide tailored instruction to, students entering a particular grade with knowledge below that grade level (Schaffner, Glewwe and Sharma, 2020). More specifically, a campaign of ***No child is left below the minimum level of learning*** is highly recommended at school level. In this campaign, the Curriculum Development Centre is advised to define the minimum level of learning (learning standards) with the technical coordination of ERO; CEHRD is advised to prepare teacher training guidelines in line with this campaign; and NEB to prepare a guideline to evaluate such learning.

2. Wide gaps in achievement between provinces

The study shows variation in the performance of province-level achievement in mathematics, science, Nepali and English. A huge gap between the high-performing and low-performing provinces in achievement has a scale of 49 in mathematics, 48 in science, 55 in Nepali and 28 in English. Bagmati, Gandaki and Lumbini are high-performing provinces, whereas Province 1 Madhesh, Karnali and Sudur Paschim are low-performing ones

Recommendation: To address the wide gap between high-performing and low-performing provinces, justified distribution of resources is a necessity. In Province 1, Madhesh, Karnali and Sudur Paschim, policy reformation, special emphasis on budget allocation, development of human resources, contextualisation of curriculum and close monitoring and evaluation of educational programmes are suggested areas of primary intervention by the government. A minimum standard of infrastructure, learning opportunities, resources, incentives and retention of good teachers and identification of learning difficulties along with remedial teachings are supportive activities to enhance learning and increase students' achievement. Specific curricula and instruction methods that can be embodied in daily teaching guides and related instructional materials can be developed, and distribution of these guides and materials and the teacher training can be packaged together to improve student learning (Schaffner, Glewwe and Sharma, 2020). In addition, small-scale policy experiments should be designed and analysed to help improve the implementation aspects so that programmes have a high success probability.

3. Huge disparity in achievement by type of school

A huge disparity in achievement between community and institutional schools may create a two-tiered society in future. A huge gap is seen in achievement between institutional and community schools, with a range of scale score of 37 in science, as an example. The range of differences are similar in other subjects too.

Recommendation: The gap should be filled by upgrading community schools through strategic interventions in school education. It is imperative to identify malfunctions in input, process and output of the community school mechanism and reform policy for improvement to the existing conditions. A comprehensive analysis of better-performing institutional and community schools is sorely needed to explore how poor-performing community schools can be improved. Local governments also have an important role to play in improving the quality of public education.

4. Home language also brought a remarkable gap in achievement

A remarkable gap in achievement has been revealed by the language used at home, which ranges in scale score of 15 in Nepali and 22 in English between the high achievers and low achievers.

Recommendation: This gap can be narrowed by teachers using the home language of children in the classroom, even in the earlier grades. Teachers need at least a basic-level language learning package for their students or language of the community surrounding the school. Teachers have to be able to communicate in the community language, and they have to teach translating, changing codes, using trans-language strateies, and empowering those children who use languages other than Nepali at home. A comprehensive language learning package for teachers for their professional development deserves incorporation in TPD.

5. There is a visible gap in the learning achievement between boys and girls

The results revealed that boys scored higher than girls. The study shows a noticeable disparity between boys and girls in their achievement. The gap ranges in scale scores of 12 in mathematics, 5 in science and 10 in English, though generally there is no gap in Nepali. This indicates the need to work on gender equity in learning achievement.

Recommendation: The reasons behind such disparity in learning between boys and girls are worth exploring further so that effective interventions to reduce gender differences in learning can be devised. Suggested interventions include teachers paying attention to student-friendly (more focused on girls) behaviour and teaching and learning activities in the classroom, including remedial education. Affirmative action such as scholarships and additional incentives to girls may reduce gender disparity in achievement. Regular interactions with female role models may also help. Apart from these, teachers should create a suitable learning environment for girls by being sensitive in terms of their needs, interest, voices and providing equal opportunity for classroom participation. Parents are to be encouraged in their roles in supporting equality in their children's education.

6. Students at the appropriate age performed better

Students studying in grade 8 at the ages of 12 and 13 scored higher than underage and overage students studying at the same level. The similarity in age group among students may have encouraged them to share and discuss their education-related problems, thereby enabling them to excel academically. The gap in the achievement of students aged 12 or 13 compared to other age groups has been in scale scores of 25 in mathematics, 19 in science and 24 in Nepali.

Recommendation: If the student is below age 14 while in grade 8, the child was in grade 1 at or before age 4. Similarly, if the child is aged 15 or above in grade 8, it is most likely an indication that they have repeated grades or started grade 1 in a less conducive environment. In addition to encouraging children to enrol on time, teachers should be trained in formative assessments in earlier grades and remedial education so that students do not fall behind in their studies and repeat grades.

7. The relationship between students' academic performance and socio-economic status is substantial, but its magnitude varies by subjects

The socio-economic status of a student's family has varying effects on their achievement. Many students have performed better in Nepali language, with satisfactory performance in mathematics and science, despite their low socio-economic status. This situation was reversed in English. This depicts that the socio-economic background of the students does not entirely decide their academic performance.

Recommendation: Although the socio-economic status of students has varying effects on their achievement, it is not the only major deciding factor. Students can excel and achieve better if they focus more on their studies and practise hard, despite the minimum resources available to them. Despite the different levels of socio-economic status of students, if the schools, for example, provide sufficient learning materials, library facilities and manage student clubs and study programmes, the students can perform well irrespective of their socio-economic status.

8. The achievement of Janajati and Dalit children is lower than other ethnicities

Ethnicity has influenced the achievement of students in Nepali, English and science. Generally. Brahman/Chhetri scored higher than Janajati and Dalit students. Students from Brahman/Chhetri communities are, on average, high achievers, whereas students from Dalit communities are achieving lower. There is a significant difference in achievement between Hill Brahman and Madhesi Dalit of scale score 26 in science, 26 in English and 30 in Nepali.

Recommendation: The achievement scores of students from Janajati communities and Dalit communities are below the national average compared to students from Brahman and Chhetris communities. The differences may have been caused by medium of instruction, language background, curriculum content, teachers or cultural background. To reduce these gaps, an inclusive curriculum, remedial teaching, incorporation of local ideologies in the curriculum, inclusiveness in teaching profession, change of learning culture in Dalit –

and more importantly Madhesi Dalit – students need to be seriously considered.

9. **Teacher regularity and availability of study resources have positive relationships with learning achievement**

   Teachers who were dedicating all their time in the classroom were successful in improving students' achievement. Meanwhile, students of teachers who arrived late to class and left early, or who did not come to class at all, had negative performances. Similarly, the availability of study resources, such as textbooks, question banks, guides, reference materials and other support, has a positive influence on learning achievement.

   Recommendation: School administration should maintain a strict code of conduct for teachers to be in the school regularly and it should be made as one of the criteria for their performance evaluation. Regular teachers should be rewarded with incentives. Similarly, government or non-government agencies, supporting students through scholarships or any other incentives, should consider the availability of basic study resources to students. Parents should also consider making these essential resources available to meet the primary needs of their children.

10. **Decreasing patterns of achievement and consistency of NASA results**

    One-third of students in mathematics and science and nearly half of the students in English scored below the national average. The consistently weak performance of students in NASA 2012, 2015, and 2018, 2019 and 2020 indicates a low return on the investment made by the government in education. The recurring trend underscores the need for ensuring sufficient government intervention to enhance quality education.

    Recommendation: The time has already come to carry out a diagnostic study to identify the challenges in the educational system, with a focus on the teaching–learning process. The critical factors that hinder achievement and quality education should be investigated and immediate steps taken to recover the educational loss. Pedagogical intervention in the delivery system deserves exploration, and the adoption of activity-based, learner-centred and research-based learning approaches in teaching – along with problem solving, critical thinking and other 21st-century skills – with close monitoring and evaluation has now become a necessity. The involvement of parents and community members to make schools accountable for their students' low achievement should be ensured.

11. **There is a positive relationship between students' academic performance and the use of their leisure time in school**

    The performance of students who were engaged in classwork or homework during their leisure time was higher than those who spent their time in playing games or returning home. There was a significant difference in achievement of scale score 32 in mathematics, 42 in science, and 54 in Nepali.

Recommendation: School administration should manage the gap periods in school properly. Meanwhile identification of weak students and remedial classes for them during leisure time remains a milestone for their recovery of learning lost due to various factors. School administration should manage the students who have leisure time, engaging them in doing either classwork or homework under the guidance of teachers. The head teacher needs to pay due consideration when arranging teacher training or any other work, so as not create gaps in teaching hours. The relevant educational authority can also manage the learning environment and remedial classes with a budget to recover lost learning.

12. Access to social media also has a positive effect on student achievement

NASA 2020 showed the importance of access to social media. Particularly during the pandemic, having access worked well. Students who had access to social media had a higher achievement score than those who lacked it. There is a significant difference between the achievement of students who had access to social media such as mobile phone, internet and TV and other means of social media, with 21 scale score in mathematics and 10 scale score in science.

Recommendation: Students deserve access to social media for information, and communication devices that could facilitate their learning and keep them updated with current information which ultimately enhances their learning. However, these devices should be for the purpose of learning. Unauthorised programs and uncensored programs unsuitable for their age level could hamper them psychologically. Any social media and devices made available to students must be censored and loaded with educational programs that enhance learning opportunities for better achievement.

13. Medium of instruction has impacted on achievement in community schools

Medium of instruction plays an important role in education and the pedagogical process. Basically, institutional schools follow English-medium instruction, whereas community schools use Nepali as the medium of instruction. In NASA 2020, the achievement level of English-medium community school students was higher than for those studying in Nepali. There is a significant difference between community schools where English is the medium of instruction and those applying Nepali as the medium of instruction. For instance, the difference in achievement of students from English-medium and Nepali-medium schools was 17 scale score in science.

Recommendation: Medium of instruction is a fundamental process for communication and comprehension of the content as well as pedagogical process in schools. The language that students feel easiest with must be used as the medium of instruction. In NASA 2020, language and the achievement of students showed a direct relation, but a major factor is the comprehension of content delivered in the classroom pedagogy. This is the grey area for further research, rather than just prescribing a particular medium of instruction, as the world is growing ever more multilingual and our languages are vast resources of knowledge.

14. Noticeable gaps in school governance

Some identified factors of school governance, like teacher unreliability, bullying in school, a lack of improvement in classroom practices, unavailability of school facilities and conducive learning environment, unavailability of timely textbook, students' perception towards school/ teachers, the large numbers of students lacking appropriate learning opportunities, and a lack of remedial action to improve learning are some of the factors causing the detrimental situation in students' achievement outcomes. The difference in achievement between community schools and institutional schools is also one of the types of evidence of difference in school governance.

Recommendation: These issues are basically the concern of the school governance. These issues need to be solved for achievement in school to improve. Monitoring and evaluation of school governance, empowering head teachers, and reward and punishment systems are necessary steps to improve the situation. Local government should play a crucial role in the improvement of governance in the school.

Note: Sepecfic finding and suggestions of reading, writing, listening and speaking are mentioned in the section of result of English (See pp. 196-200, 214-217).

References

ACARA (2017). NAP Sample Assessment Science Literacy Technical Report: Sydney: Australian Curriculum Assessment and Reporting Authority (ACARA).

Akyuz, V. (2004). The Effects of Textbook Style and Reading Strategy on Students' Achievement and Attitude Towards Heat and Temperature. Ankara: Middle East Technical University.

An, Xinming and Yung, Yiu-Fai (2014). Paper SAS364-2014 Item Response Theory: What it is and how you can use the IRT procedure to apply it. Retrieved from https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf.

Baker, F.B. (2001). The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation. Retrieved from http://echo.edres.org:8080/irt/baker/final.pdf.

Bhatta, Saurav D. and Uttam, Sharma (2019). Whither Quality? What Do Recent National Assessments of Student Learning Outcomes in Bangladesh Tell Us? Washington D.C.: World Bank Group.

CDC (2013). Lower Secondary Curriculum. Sanothimi: CDC (Curriculum Development Centre).

CDC (2014). National Curriculum Framework. Sanothimi: Author.

Chalmers, P.P. (2012). A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software, Volume 48, Issue 6. http://www.jstatsoft.org/ York University.

CMEC (2014). PCAP 2013: Report on the Pan-Canadian Assessment of Science, Reading, and Mathematics. Toronto: CMEC.

Cohen, L. and Holliday, M. (1996) Practical Statistics for Students. London: Paul Chapman.

Cohen, L., Manion, L. and Morrison, K. (2007). Research Methods in Education. Sixth edition. Routledge, Taylor & Francis Group, London and New York.

Cooper, H. and Valentine, J.C. (2001). Using Research to Answer Practical Questions about Homework. Educational Psychologists, 36, 143–153.

ERO (2013). Report of National Assessment of Student Achievement 2011, Grade 8. Sanothimi: Education Review Office.

ERO (2015). Report of National Assessment of Student Achievement 2012, Grade 3 and 5. Sanothimi: Education Review Office.

ERO (2015a). Report of National Assessment of Student Achievement 2013, Grade 8. Sanothimi: Education Review Office.

ERO (2016). Report of National Assessment of Student Achievement 2015, Grade 3 and 5. Sanothimi: Education Review Office.

ERO (2017). NASA 2017: Assessment Framework for Grade 5 in Mathematics and Nepali. Sanothimi: Education Review Office.

Fennema, E. and Sherman, J.A. (1976). Fennema-Sherman Mathematics Attitudes Scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. Journal for Research in Mathematics Education, 7, 324–326. doi: 10.2307/748467.

Greaney, V. and Kellaghan, T. (2007). Assessing National Achievement Levels in Education [edited]. Washington DC: The World Bank.

IEA (2016). Methods and Procedures in TIMSS 2015. Martin, M.O., Mullis, V. S. and Hooper, M. (eds). Boston: International Association for the Evaluation of Educational (IEA).

Kainuwa, A. and Yusuf, N.B.M. (2013). Influence of Socio-Economic and Educational Background of Parents on their Children's education in Nigeria. International Journal of Scientific and Research Publications, 3 (10), 1–8.

Krejcie, R.V. and Morgan, D.W. (1970). Determining Sample Size for Research Activities. Educational and Psychological Measurement, 30, 607–610.

Laukaitytė, Inga (2016). Statistical Modeling in International Large-scale Assessments. Department of Statistics, Umeå School of Business and Economics.

Le, Dai-Trang (2013). Applying Item Response Theory Modeling in Educational Research. Owa State University.

Lewis, T. (2015). Replication Techniques for Variance Approximation. Paper 2601-2015. University of Maryland.

Masters, G.M. and Wright, B.D. (1997). The Partial Credit Model. In W.J. Van der Linden and R.K. Hambleton (eds), Handbook of Modern Item Response Theory, pp. 110–122. New York, Springer.

Meinck, S. (2015). Computing Sampling Weights in Large-scale Assessments in Education. Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach. Retrieved from http://surveyinsights.org/?p=5353.

Murphy, M. and Schulz, W. (2006). Sampling for National Surveys in Education. Victoria, Australia: Australian Council for Educational Research (ACER).

Murphy, P., Greaney, V., Lockheed, M.E. and Rojas, C. [eds] (1996). National assessments: testing the system (English). EDI learning resources series. Washington, DC: The World Bank. http://documents.worldbank.org/curated/en/920211468743345819/

National-assessments-testing-the-system

OECD (2009). PISA Data Analysis Manual, SPSS, Second Edition. Retrieved from https://www.ipi.lu.lv/fileadmin/_migrated/content_uploads/Datu_analizes_rk.pdf.

OECD (2016). PISA 2015: Technical Report. Retrieved from http://www.oecd.org/pisa/data/2015-technical-report/.

OECD (2017). PISA 2015 Assessment and Analytical Framework (Science, Reading, Mathematics, Financial Literacy and Collaborative Problem Solving). Retrieved from www.oecd-ilibrary.org/education/pisa-2015-assessment-and....

OECD 2019. OECD Future of Education 2030. Making Physical Education Dynamic and Inclusive for 2020. International Curriculum Analysis. https://www.oecd.org/education/2030-project/contact/OECD_FUTURE_OF _EDUCATION_2030_MAKING_PHYSICAL_DYNAMIC_AND_INCLUSIVE_FOR_2030.pdf)

O'Grady, K. and Houme, K. (2015). PCAP 2013: Technical Report. Toronto: Council of Ministers of Education, Canada.

Pant L.B., Singh, G.B. and Poudel, L.N. (2016). NASA 2017: Assessment Framework for Grade 8 in Mathematics, Nepali and Science. Sanothimi: Education Review Office.

Rahi, S. (2017). Research design and methods: A systematic review of research paradigms, sampling issues and instruments development. International Journal of Economics & Management Sciences, 6(2), 1-5. DOI: 10.4172/2162-6359.1000403

Schaffner, Julie, Glewwe, Paul and Sharma, Uttam (2020). Evaluation of the Design, Implementation and Impacts of SSDP Trainings for Teachers of Secondary Math and Science in Nepal. Final Report submitted to International Initiative for Impact Evaluation (3ie).

Sudal, L.A. and Sudol, C. (2010). Analysing Test Items: Using Item Response Theory to Validate Assessments. Proceedings of the 41st ACM technical symposium on computer science education, pp 436–440. Retrieved from http://ims.mii.lt/ims/konferenciju_medziaga/SIGCSE%2710/docs/p436.pdf.

UNESCO (n.d.). Extra-curricular activities. Retrieved 14 June 2020, from International Bureau of Education: http://www.ibe.unesco.org/en/glossary-curriculum-terminology/e/extra-curricular-activities

Verhelst, N.D. and Verstralen, H.H.F.M. (2008). Some Considerations on the Partial Credit Model. *Psicologica (2008),* 29, 229–254.

Von Davier, M., Gonzalez, E. and Mislevy, R.J. (2009). What are plausible values and why are they useful? In M. Von Davier and D. Hastedt (eds), *IERI Monograph Series: Issues* *and methodologies in large scale assessments.* Vol. 2, pp. 9–36. Hamburg, Germany: IERI Institute.

Xie, Qingshu, Zhang, MacroSys Ting (2017). *Analyses Using Achievement Levels Based on Plausible Values AIR* - NAEP Working Paper #2017-02 NCES Data R Project Series #01 American Institutes for Research.

Yamamoto, K. and Kulick, E. (2000). Scaling Methodology and Procedures for the TIMSS Mathematics and Science Scales. In M.O. Martin, K.D. Gregory and S.E. Stemler (eds), *TIMSS 1999 technical report.* Chestnut Hill, MA: Boston College.

Yansaneh, I.S. (2003). *Construction and use of sample weights.* United Nations Secretariat.

Zieky, M. and Perie, M. (2006). *A Primer on Setting Cut Scores on Tests of Educational Achievement.* Educational Testing Service (ETS).

Appendices

Appendix 1

Mathematics classical results

a. Classical results by caste/ethnicity

Bq8ethnicgroup	Mean	Std. Deviation	Std. Error of Mean	N	% of Total N	Significance	Partial Correl
Madhesi Brahman / Chhetri	29	20.96	0.11	39777.00	0.09	Sig	0.17
Madhesi Janjati	27	18.39	0.08	59351.00	0.13		
Madhesi Dalit	23	16.57	0.13	17178.00	0.04		
Madhesi other	28	19.54	0.11	33844.00	0.07		
Pahadi Brahman/ Chhetri	30	20.97	0.06	119612.00	0.26		
Pahadi Janjati	26	17.37	0.06	90671.00	0.20		
Pahadi Dalit	22	14.98	0.09	29997.00	0.07		
Pahadi Other	25	17.31	0.13	16844.00	0.04		
Himali Brahman/ Chhetri	20	16.55	0.21	6422.00	0.01		
Himali Janjati	24	17.68	0.28	4057.00	0.01		
Himali Dalit	18	13.52	0.41	1082.00	0.00		
Himali other	21	17.25	0.52	1101.00	0.00		
Not mentioned	20	16.59	0.08	42734.00	0.09		
Total	26	19.02	0.03	462668.00	1.00		

b. Classical results by medium of instruction

BQ9b Medium of instruction	Mean (%)	Std. Deviation	Std. Error of Mean	N	% of Total N	Sig	Part. correl
Nepali	21.8438	15.21304	0.03108	239642	51.80%	Sig at $p<0.001$	0.345
English	36.7231	21.93984	0.06059	131122	28.30%		
Other and not specified	23.1917	17.71587	0.05844	91905	19.90%		
Total	26.3284	19.0165	0.02796	462668	100.00%		

c. Classical results by mother's education

Mother's education	Mean	Std. Deviation	Std. Error of Mean	N	% of Total N	Significance	Partial correl
Illiterate	21.8076	15.46983	0.04134	140040	30.30%	Sig. at p <0.001	0.335
Just literate	24.4357	16.49071	0.04838	116207	25.10%		
Grade 8 pass	25.8249	18.02011	0.06265	82736	17.90%		
Grade 10 pass	31.9532	20.74335	0.08378	61299	13.20%		
Grade 12 pass	38.2611	23.12615	0.13058	31363	6.80%		
Bachelor's	44.0142	25.18762	0.2284	12161	2.60%		
Master's or above	52.6313	24.05961	0.34794	4781	1.00%		
Not specified	14.596	14.46658	0.12192	14080	3.00%		
Total	26.3284	19.0165	0.02796	462668	100.00%		

d. Classical results by father's education

Father's education	Mean	Std. Deviation	Std. Error of Mean	N	% of Total N	Significance	Partial correl
Illiterate	20.0441	14.17702	0.05719	61461	13.30%	Sig. at p <0.001	0.339
Just literate	23.1364	15.45437	0.05147	90163	19.50%		
Grade 8 pass	23.5944	16.14534	0.04947	106494	23.00%		
Grade 10 pass	27.3389	18.83491	0.05915	101394	21.90%		
Grade 12 pass	33.6135	21.80766	0.09324	54699	11.80%		
Bachelor's	40.6812	24.20527	0.16624	21200	4.60%		
Master's or above	49.0868	24.62782	0.21943	12597	2.70%		
Not specified	17.6848	17.49778	0.14452	14660	3.20%		
Total	26.3284	19.0165	0.02796	462668	100.00%		

e. Classical results by mother's profession

Mother's profession	Mean	Std. Deviation	Std. Error of Mean	N	% of Total N	Significance	Partial correl
Agriculture and household work	23.1068	16.27893	0.03134	269843	58.30%	Sig.	0.256
Household work only	31.2959	21.04171	0.06677	99321	21.50%		
Work in other's house	22.4536	18.24025	0.28462	4107	0.90%		
Work for wage	25.3951	18.20339	0.23298	6105	1.30%		
Work abroad	26.9851	17.49228	0.19006	8471	1.80%		
Teaching	38.2199	23.33894	0.21778	11485	2.50%		
Business	31.6102	21.06153	0.12022	30694	6.60%		
Government job	37.222	23.05172	0.23659	9493	2.10%		
Other	38.0865	23.08542	0.24948	8562	1.90%		
Not specified	18.7333	19.27354	0.15958	14586	3.20%		
Total	26.3284	19.0165	0.02796	462668	100.00%		

f. Classical results by father's profession

Father’s profession	Mean	Std. Deviation	Std. Error of Mean	N	% of Total N	Significance	Partial correl
Agriculture and household work	21.6359	15.44964	0.04213	134483	29.10%	Sig.	0.337
Household work only	19.0225	14.20137	0.14239	9948	2.20%		
Work in other's house	19.4071	13.74578	0.14917	8491	1.80%		
Work for wage	24.8491	15.95603	0.08561	34734	7.50%		
Work abroad	25.658	17.82753	0.05436	107541	23.20%		
Teaching	35.2244	22.62773	0.18981	14212	3.10%		
Business	31.9569	21.56803	0.08306	67429	14.60%		
Government job	32.4986	21.65473	0.11971	32725	7.10%		
Other	35.6368	21.86395	0.12136	32456	7.00%		
Not specified	20.3247	18.40325	0.12806	20651	4.50%		
Total	26.3284	19.0165	0.02796	462668	100.00%		

g. Classical results by home facilities

Home facilities	Mean	Std. Deviation	Std. Error of Mean	N	% of Total N	Significance	Partial correl
5 or less	26.1127	18.61488	0.02844	428297	95.10%	Sig.	0.082
5–10	37.3194	24.0047	0.16664	20752	4.60%		
11 or more	30.87	24.38277	0.70006	1213	0.30%		
Total	26.6421	19.06136	0.02841	450262	100.00%		

h. Classical item parameters of mathematics sets
(i) Set 1 item parameters of Mathematics

SN	Item Number	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of correct answer (P)
1	Math1Q1_scored	188157	0	1	0.8	0.396	80%
2	Math1Q2_scored	96131	0	1	0.54	0.499	54%
3	Math1Q3_scored	94021	0	1	0.51	0.5	51%
4	Math1Q4_scored	92022	0	1	0.44	0.497	44%
5	Math1Q5_scored	92022	0	1	0.25	0.434	25%
6	Math1Q6_scored	92022	0	1	0.48	0.499	48%
7	Math1Q7_scored	92022	0	1	0.47	0.499	47%
8	Math1Q8_scored	92022	0	1	0.22	0.416	22%
9	Math1Q9_scored	183251	0	1	0.27	0.444	27%
10	Math1Q10_scored	92022	0	1	0.7	0.46	70%
11	Math1Q11_scored	183251	0	1	0.48	0.499	48%

SN	Item Number	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of correct answer (P)
12	Math1Q12_scored	92022	0	1	0.29	0.454	29%
13	Math1Q13_scored	92022	0	1	0.39	0.488	39%
14	Math1Q14_scored	183251	0	1	0.47	0.499	47%
15	Math1Q15_scored	183251	0	1	0.17	0.377	17%
16	Math1Q16_scored	92022	0	2	0.16	0.485	8%
17	Math1Q17_scored	92022	0	2	0.19	0.392	10%
18	Math1Q18_scored	92022	0	2	0.88	0.976	44%
19	Math1Q19_scored	183251	0	2	0.49	0.823	25%
20	Math1Q20_scored	92022	0	2	0.11	0.416	6%
21	Math1Q21_scored	92022	0	2	0.13	0.464	7%
22	Math1Q22_scored	92022	0	2	0.12	0.436	6%
23	Math1Q23_scored	92022	0	2	0.21	0.523	11%
24	Math1Q24_scored	92022	0	2	0.28	0.638	14%

(ii) Set 2 item parameters of mathematics

SN	Item Number\	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of correct answer (P)
2	Math2Q2_scored	91229	0	1	0.2338	0.42323	23%
3	Math2Q3_scored	91229	0	1	0.2907	0.45411	29%
4	Math2Q4_scored	91229	0	1	0.6751	0.46834	68%
5	Math2Q5_scored	91229	0	1	0.4661	0.49885	47%
6	Math2Q6_scored	91229	0	1	0.4336	0.49558	43%
8	Math2Q8_scored	91229	0	1	0.2722	0.44507	27%
10	Math2Q10_scored	91229	0	1	0.3578	0.47936	36%
11	Math2Q11_scored	182033	0	1	0.2128	0.40927	21%
12	Math2Q12_scored	187886	0	1	0.3512	0.47733	35%
13	Math2Q13_scored	187886	0	1	0.4273	0.49468	43%
16	Math2Q16_scored	91229	0	1	0.1009	0.30116	10%
17	Math2Q17_scored	91229	0	2	0.1336	0.441	7%
18	Math2Q18_scored	91229	0	2	0.2656	0.66002	13%
21	Math2Q21_scored	91229	0	2	0.0703	0.31995	4%
22	Math2Q22_scored	91229	0	2	0.215	0.55266	11%
23	Math2Q23_scored	91229	0	2	0.1865	0.4962	9%
24	Math2Q24_scored	91229	0	1	0.064	0.24484	6%

(iii) Set 3 item parameters of mathematics

SN	Item Number\	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of correct answer (P)
1	Math3Q1_scored	96658	0	1	0.4231	0.49406	42%
2	Math3Q2_scored	96658	0	1	0.2953	0.4562	30%
3	Math3Q3_scored	96658	0	1	0.1074	0.30959	11%
6	Math3Q6_scored	188613	0	1	0.5179	0.49968	52%
7	Math3Q7_scored	96658	0	1	0.4685	0.49901	47%
8	Math3Q8_scored	96658	0	1	0.2543	0.43544	25%
9	Math3Q9_scored	96658	0	1	0.2334	0.423	23%
10	Math3Q10_scored	96658	0	1	0.2833	0.4506	28%
11	Math3Q11_scored	187462	0	1	0.3689	0.4825	37%
12	Math3Q12_scored	96658	0	1	0.2455	0.43041	25%
13	Math3Q13_scored	96658	0	1	0.6219	0.48492	62%
14	Math3Q14_scored	187462	0	1	0.4539	0.49787	45%
15	Math3Q15_scored	187462	0	1	0.3469	0.47599	35%
16	Math3Q16_scored	96658	0	2	0.4699	0.81344	23%
17	Math3Q17_scored	96658	0	2	0.3703	0.70761	19%
18	Math3Q18_scored	96658	0	2	0.3174	0.68285	16%
19	Math3Q19_scored	96658	0	2	0.3019	0.68746	15%
20	Math3Q20_scored	96658	0	2	0.0359	0.24117	2%
21	Math3Q21_scored	96658	0	2	0.3054	0.69922	15%
22	Math3Q22_scored	96658	0	2	0.1856	0.54501	9%
23	Math3Q23_scored	187462	0	2	0.1606	0.49456	8%
24	Math3Q24_scored	96658	0	2	0.2743	0.66344	14%
25	Math3Q25_scored	96658	0	2	0.1375	0.4828	7%

(iv) Set 4 item parameters of mathematics

SN	Item Number\	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of correct answer (P)
1	Math4Q1_scored	90804	0	1	0.3969	0.48925	40%
3	Math4Q3_scored	90804	0	1	0.3427	0.47462	34%
5	Math4Q5_scored	90804	0	1	0.4	0.48991	40%
6	Math4Q6_scored	90804	0	1	0.4663	0.49887	47%
7	Math4Q7_scored	90804	0	1	0.4287	0.49489	43%
8	Math4Q8_scored	90804	0	1	0.2622	0.43986	26%
12	Math4Q12_scored	90804	0	1	0.2258	0.41809	23%
13	Math4Q13_scored	90804	0	1	0.3622	0.48063	36%

SN	Item Number\	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of correct answer (P)
14	Math4Q14_scored	90804	0	1	0.2446	0.42985	24%
15	Math4Q15_scored	182760	0	1	0.4729	0.49927	47%
16	Math4Q16_scored	90804	0	2	0.1903	0.5541	10%
17	Math4Q17_scored	90804	0	2	0.2038	0.58646	10%
18	Math4Q18_scored	90804	0	2	0.2394	0.61913	12%
19	Math4Q19_scored	90804	0	2	0.1943	0.57436	10%
20	Math4Q20_scored	90804	0	2	0.3512	0.69659	18%
21	Math4Q21_scored	90804	0	2	0.0995	0.41815	5%
23	Math4Q23_scored	182760	0	2	0.2405	0.62765	12%
24a	Math4Q24a_scored	90804	0	1	0.1519	0.3589	15%
24b	Math4Q24b_scored	90804	0	1	0.051	0.21996	5%

(v) Set 5 item parameters of mathematics

SN	Item Number\	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of correct answer (P)
1	Math5Q1_scored	91955	0	1	0.6741	0.46873	67%
2	Math5Q2_scored	91955	0	1	0.4272	0.49467	43%
4	Math5Q4_scored	91955	0	1	0.6201	0.48537	62%
5	Math5Q5_scored	91955	0	1	0.4593	0.49835	46%
6	Math5Q6_scored	91955	0	1	0.1377	0.34463	14%
8	Math5Q8_scored	91955	0	1	0.1058	0.30762	11%
9	Math5Q9_scored	91955	0	1	0.0531	0.22414	5%
10	Math5Q10_scored	91955	0	1	0.2991	0.45786	30%
11	Math5Q11_scored	91955	0	1	0.3464	0.47583	35%
12	Math5Q12_scored	91955	0	1	0.2431	0.42896	24%
13	Math5Q13_scored	91955	0	1	0.4345	0.4957	43%
14	Math5Q14_scored	91955	0	1	0.4151	0.49274	42%
16	Math5Q16_scored	91955	0	2	0.2939	0.68767	15%
17	Math5Q17_scored	91955	0	2	0.5616	0.87478	28%
18	Math5Q18_scored	91955	0	2	0.617	0.85297	31%
19	Math5Q19_scored	91955	0	2	0.8277	0.95339	41%
20	Math5Q20_scored	91955	0	2	0.4932	0.83938	25%
22	Math5Q22_scored	91955	0	2	0.1797	0.55333	9%
23	Math5Q23_scored	91955	0	2	0.2075	0.57385	10%
24	Math5Q24_scored	91955	0	2	0.2046	0.5466	10%
25	Math5Q25_scored	91955	0	2	0.5875	0.86594	29%
26a	Math5Q26a_scored	91955	0	2	0.0452	0.20873	2%
26b	Math5Q26b_scored	91955	0	1	0.0541	0.2263	5%

i. IRT parameters of mathematics items

SN	item	alpha	Beta	tau.Cat1	tau.Cat2	tau.Cat3
1	G8M17A01	1.233	-1.432	NA	NA	NA
2	G8M17A2C2	1.211	-0.83	NA	NA	NA
3	G8M17A3C	1.022	-0.63	NA	NA	NA
4	G8M17A4B3	1.526	-0.265	NA	NA	NA
5	G8M17A5	1.152	-0.452	NA	NA	NA
6	G8M17A6	1.127	-0.913	NA	NA	NA
7	G8M17A7B8C9	0.782	-1.051	NA	NA	NA
8	G8M205Q13S	1.242	0.236	NA	NA	NA
9	G8M17A9	0.492	-0.567	NA	NA	NA
10	G8M17A10	1.256	0.139	NA	NA	NA
11	G8M17A11	1.528	0.248	NA	NA	NA
12	G8M17A12	1.072	0.207	NA	NA	NA
13	G8M17A13	0.854	0.326	NA	NA	NA
14	G8M17A14	0.291	2.168	NA	NA	NA
15	G8M17A15	0.742	1.389	NA	NA	NA
16	G8M17A16	0.581	0.363	NA	NA	NA
17	G8M17A17B17C15	1.102	0.066	NA	NA	NA
18	G8M201Q4S	0.805	0.383	NA	NA	NA
19	G8M17A19B16	0.619	1.658	NA	NA	NA
20	G8M201Q7S	1.161	0.175	NA	NA	NA
21	G8M17A21	1.279	-0.296	NA	NA	NA
22	G8M17A22B19	0.823	1.052	NA	NA	NA
23	G8M17A23	0.371	1.826	NA	NA	NA
24	G8M17A24	0.379	1.709	NA	NA	NA
25	G8M17A25	0.606	2.314	NA	NA	NA
26	G8M17A26	0.399	2.907	NA	NA	NA
27	G8M17A27	0.345	1.599	NA	NA	NA
28	G8M201Q8S	0.304	4.26	NA	NA	NA
29	G8M17A29	0.116	20.159	NA	NA	NA
30	G8M17A30C34	1.101	0.24	0.151	-0.151	NA
31	G8M17A31	1.296	0.261	0.508	-0.508	NA
32	G8M17A32	0.437	0.666	2.777	-2.777	NA
33	G8M17A33	0.8	0.416	1.475	-1.475	NA
34	G8M17A34	1.096	1.217	0.913	-0.913	NA

SN	item	alpha	Beta	tau.Cat1	tau.Cat2	tau.Cat3
35	G8M17A35	1.121	0.596	1.451	-1.451	NA
36	G8M17A36B38	1.248	0.654	0.459	-0.459	NA
37	G8M17A37	1.305	1.046	0.723	-0.723	NA
38	G8M17A38B44	1.139	1.237	0.861	-0.861	NA
39	G8M17A39	0.954	1.233	0.741	-0.162	-0.579
40	G8M17A40	1.15	1.043	1.254	-1.254	NA
41	G8M17A41	1.157	1.714	1.002	-1.002	NA
42	G8M17A42	0.937	1.619	1.165	-0.386	-0.779
43	G8M17A43	1.113	2.146	0.404	0.055	-0.46
44	G8M17A44	1.571	1.433	0.13	-0.13	NA
45	G8M17A45	1.828	1.712	NA	NA	NA
46	G8M17A46	1.102	1.131	0.732	-0.732	NA
47	G8M17A47	1.145	2.268	0.25	-0.25	NA
48	G8M17A48a	2.245	1.171	NA	NA	NA
49	G8M17A48b	2.082	1.336	NA	NA	NA
50	G8M17A48c	1.772	2.523	NA	NA	NA
51	G8M201Q1S	1.032	-1.659	NA	NA	NA
52	G8M17B2	1.035	-1.477	NA	NA	NA
53	G8M17B4	0.81	-1.713	NA	NA	NA
54	G8M17B5	0.901	-0.941	NA	NA	NA
55	G8M17B6C6	0.936	-0.08	NA	NA	NA
56	G8M17B7C7	0.992	-0.441	NA	NA	NA
57	G8M17B9	1.568	0.061	NA	NA	NA
58	G8M17B10	1.283	0.134	NA	NA	NA
59	G8M17B11	0.843	0.513	NA	NA	NA
60	G8M17B12	0.821	0.063	NA	NA	NA
61	G8M17B13	0.879	0.065	NA	NA	NA
62	G8M17B14	0.93	0.354	NA	NA	NA
63	G8M17B15	0.681	0.48	NA	NA	NA
64	G8M17B18	0.486	1.803	NA	NA	NA
65	G8M17B20	0.686	-0.5	NA	NA	NA
66	G8M17B21	0.691	1.343	NA	NA	NA
67	G8M17B22	0.968	0.023	NA	NA	NA
68	G8M17B25	0.153	10.573	NA	NA	NA
69	G8M205Q8S	0.381	5.717	NA	NA	NA

SN	item	alpha	Beta	tau.Cat1	tau.Cat2	tau.Cat3
70	G8M17B27	0.675	1.658	NA	NA	NA
71	G8M17B28C29	1.014	1.613	NA	NA	NA
72	G8M17B29C28	1.615	0.083	NA	NA	NA
73	G8M17B30	1.052	1.921	NA	NA	NA
74	G8M17B31	1.279	1.506	1.34	-1.34	NA
75	G8M17B32	1.293	0.646	0.294	-0.294	NA
76	G8M17B33	1.253	0.816	0.836	-0.836	NA
77	G8M17B34	2.373	0.745	NA	NA	NA
78	G8M17B35C32	0.938	0.129	1.726	-1.726	NA
79	G8M17B36	0.658	1.023	1.194	-0.54	-0.654
80	G8M17B37	1.095	0.961	1.378	-1.378	NA
81	G8M17B39	1.457	1.356	1.157	-1.157	NA
82	G8M17B40	1.138	1.362	1.513	-1.513	NA
83	G8M17B41	0.874	1.934	1.551	-0.362	-1.188
84	G8M17B42	1.382	1.27	1.022	-1.022	NA
85	G8M17B43	1.462	1.285	0.22	-0.22	NA
86	G8M17B45	1.465	0.782	0.707	-0.707	NA
87	G8M17B46	1.101	1.565	1.442	-1.442	NA
88	G8M17B47	1.209	1.708	0.338	-0.338	NA
89	G8M17B48	2.065	1.92	NA	NA	NA
90	G8M17B49C46	0.942	2.542	1.918	-1.918	NA
91	G8M17B50C49	1.518	2.437	NA	NA	NA
92	G8M17B51C47	2.143	0.788	NA	NA	NA

Appendix 2

Classical results for science

a. Classical results in science by ethnic group

Ethnic group	Mean	N	Std. Deviation	Std. Error of Mean
Hill Brahmin	32.2006	114787	19.02071	0.05614
Hill Janjati	28.9673	87628	16.19507	0.05471
Hill Others	28.1329	16161	16.06324	0.12636
Madhesi Brahmin	27.7668	38680	18.43736	0.09375
Hill Dalit	26.6244	29190	14.5569	0.0852
Madhesi Janjati	25.7414	58441	16.15724	0.06684
Madhesi Others	24.7376	33430	16.67926	0.09122
Madhesi Dalit	23.1937	16744	15.23697	0.11775
Mountain Brahmin	23.0628	6236	16.20416	0.20519
Not-stated	20.5814	47546	15.80166	0.07247
Total	27.5875	448844	17.39827	0.02597

b. Classical results by the type of local level

Local level	Mean	N	Std. Deviation	Std. Error of Mean
Rural municipality	24.3899	147190	14.51382	0.03783
Urban municipality	29.5605	278724	18.58006	0.03519
Not stated	24.1304	22930	15.86991	0.1048
Total	27.5875	448844	17.39827	0.02597

c. Classical results by mother's education

Mother's education	Mean	N	Std. Deviation	Std. Error of Mean
Illiterate	23.5924	136067	14.50463	0.03932
Literate	26.9423	111883	15.28292	0.04569
Grade 8	26.6887	80357	16.28089	0.05743
Grade 10	31.6353	59619	18.93095	0.07753
Grade 12	38.0738	30722	20.85762	0.119
Bachelor's	42.5265	11934	22.86493	0.20931
Master's or above	49.3174	4740	23.0209	0.33437
Not stated	15.9951	13522	14.81797	0.12743
Total	27.5875	448844	17.39827	0.02597

d. Classical results in science by father's education

Father's education status	Mean	N	Std. Deviation	Std. Error of Mean
Illiterate	21.8454	59226	13.44517	0.05525
Literate	26.1325	87309	14.49555	0.04906
Grade 8	25.0668	103060	15.23402	0.04745
Grade 10	28.1489	98575	17.03945	0.05427
Grade 12	33.5232	53520	19.77808	0.08549
Bachelor's	39.3467	20600	21.29924	0.1484
Master's or above	47.0124	12431	22.35969	0.20054
Not stated	18.3932	14122	17.52799	0.1475
Total	27.5875	448844	17.39827	0.02597

e. Classical results by father's occupation

Father's occupation	Mean	N	Std. Deviation	Std. Error of Mean
Agriculture + household	23.5023	130051	14.35749	0.03981
Household	21.6563	9575	14.56034	0.1488
Work in other home	21.3151	8243	13.48444	0.14852
Labour	28.5708	34072	15.16429	0.08215
Foreign	26.9933	104023	16.40262	0.05086
Teaching	34.6553	13653	20.5188	0.1756
Business	31.9614	65836	19.28296	0.07515
Government job	32.7158	32063	19.26859	0.10761
Others	35.4008	31546	19.99905	0.1126
Not stated	21.1537	19782	18.1662	0.12916
Total	27.5875	448844	17.39827	0.02597

f. Classical item parameters in science

Item	N	Maximum	Mean		Std. Deviation	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Percentage of correct answer
G8S20Q2	89411	1	0.72	0.001	0.447	72
G8S20Q53	89411	1	0.28	0.001	0.448	28
G8S20Q59	89411	1	0.26	0.001	0.44	26
G8S20Q93	89411	1	0.32	0.002	0.468	32
G8S20Q49	89391	1	0.32	0.002	0.466	32
G8S20Q120	89391	1	0.33	0.002	0.47	33
G8S20Q121	89391	1	0.5	0.002	0.5	50
G8S20Q122	89391	1	0.44	0.002	0.497	44
G8S20Q123	89391	1	0.22	0.001	0.415	22
G8S20Q124	89391	1	0.34	0.002	0.474	34
G8S20Q125	177848	1	0.49	0.001	0.5	49

Item	N	Maximum	Mean		Std. Deviation	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Percentage of correct answer
G8S20Q126	177848	1	0.41	0.001	0.492	41
G8S20Q127	177732	1	0.44	0.001	0.497	44
G8S20Q128	177629	1	0.43	0.001	0.496	43
G8S20Q129	177318	1	0.37	0.001	0.484	37
G8S20Q140	181008	1	0.35	0.001	0.476	35
G8S20Q141	180738	1	0.48	0.001	0.5	48
G8S20Q142	180572	1	0.23	0.001	0.419	23
G8S20Q143	180467	1	0.54	0.001	0.498	54
G8S20Q144	180153	1	0.56	0.001	0.496	56
G8S20Q145	88602	1	0.52	0.002	0.499	52
G8S20Q146	88581	1	0.45	0.002	0.497	45
G8S20Q147	88581	1	0.63	0.002	0.482	63
G8S20Q148	88581	1	0.36	0.002	0.48	36
G8S20Q149	88581	1	0.34	0.002	0.474	34
G8S20Q160	181978	1	0.37	0.001	0.484	37
G8S20Q161	181689	1	0.42	0.001	0.493	42
G8S20Q162	93715	1	0.39	0.002	0.488	39
G8S20Q163	93674	1	0.13	0.001	0.331	13
G8S20Q164	181497	1	0.26	0.001	0.439	26
G8S20Q165	181373	1	0.41	0.001	0.492	41
G8S20Q166	180968	1	0.67	0.001	0.47	67
G8S20Q167	93529	1	0.38	0.002	0.486	38
G8S20Q168	93509	1	0.37	0.002	0.484	37
G8S20Q169	93509	1	0.24	0.001	0.429	24
G8S20Q170	93488	1	0.32	0.002	0.466	32
G8S20Q180	176678	1	0.34	0.001	0.475	34
G8S20Q181	176678	1	0.41	0.001	0.492	41
G8S20Q182	176630	1	0.28	0.001	0.447	28
G8S20Q183	176630	1	0.23	0.001	0.422	23
G8S20Q184	87991	1	0.5	0.002	0.5	50
G8S20Q185	176568	1	0.57	0.001	0.495	57
G8S20Q186	87971	1	0.17	0.001	0.375	17
G8S20Q187	87971	1	0.42	0.002	0.494	42
G8S20Q188	87885	1	0.47	0.002	0.499	47
G8S20Q189	87865	1	0.13	0.001	0.341	13
G8S20Q199	88811	1	0.7	0.002	0.458	70

Item	N	Maximum	Mean		Std. Deviation	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Percentage of correct answer
G8S20Q200	88811	1	0.42	0.002	0.493	42
G8S20Q201	88790	1	0.17	0.001	0.377	17
G8S20Q202	88770	1	0.28	0.002	0.448	28
G8S20Q203	88749	1	0.52	0.002	0.499	52
G8S20Q204	88708	1	0.37	0.002	0.483	37
G8S20Q205	88687	1	0.43	0.002	0.495	43
G8S20Q206	88687	1	0.2	0.001	0.401	20
G8S20Q207	88062	1	0.24	0.001	0.429	24
G8S20Q130	86176	2	0.19	0.001	0.416	9.5
G8S20Q131	86011	1	0.07	0.001	0.249	7
G8S20Q132	85928	2	0.65	0.003	0.747	32.5
G8S20Q133	85346	1	0.46	0.002	0.498	46
G8S20Q103	84622	1	0.15	0.001	0.352	15
G8S20Q106	84498	1	0.07	0.001	0.25	7
G8S20Q69	84167	2	0.18	0.002	0.538	9
G8S20Q77	83981	2	0.16	0.002	0.436	8
G8S20Q115	83774	1	0.14	0.001	0.345	14
G8S20Q134	169957	2	0.26	0.001	0.442	13
G8S20Q135	167546	2	0.39	0.001	0.585	19.5
G8S20Q136	248493	2	0.37	0.001	0.677	18.5
G8S20Q137	162640	2	0.16	0.001	0.457	8
G8S20Q138	161686	2	0.39	0.002	0.671	19.5
G8S20Q139	74801	1	0.3	0.002	0.457	30
G8S20Q150	84541	2	0.16	0.002	0.442	8
G8S20Q151	84231	2	0.33	0.002	0.638	16.5
G8S20Q152	83915	2	0.18	0.002	0.457	9
G8S20Q153	83825	2	0.86	0.003	0.868	43
G8S20Q154	171936	2	0.42	0.002	0.691	21
G8S20Q155	171006	2	0.56	0.002	0.741	28
G8S20Q156	167115	1	0.13	0.001	0.332	13
G8S20Q157	242484	1	0.13	0.001	0.34	13
G8S20Q158	162727	2	0.28	0.001	0.458	14
G8S20Q159	71542	2	0.66	0.003	0.702	33
G8S20Q171	87241	2	0.29	0.002	0.602	14.5
G8S20Q172	85979	2	0.52	0.003	0.752	26
G8S20Q173	253405	2	0.08	0.001	0.308	4

Item	N	Maximum	Mean		Std. Deviation	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Percentage of correct answer
G8S20Q174	168158	1	0.24	0.001	0.429	24
G8S20Q175	166159	2	0.19	0.001	0.484	9.5
G8S20Q176	164916	1	0.07	0.001	0.258	7
G8S20Q177	164394	2	0.36	0.002	0.691	18
G8S20Q178	79038	2	0.96	0.003	0.87	48
G8S20Q179	67937	2	0.43	0.003	0.698	21.5
G8S20Q190	84572	2	0.62	0.003	0.811	31
G8S20Q191	84047	1	0.15	0.001	0.356	15
G8S20Q192	83425	1	0.18	0.001	0.381	18
G8S20Q193	168901	2	0.75	0.002	0.789	37.5
G8S20Q194	166328	2	0.29	0.001	0.589	14.5
G8S20Q195	165326	1	0.13	0.001	0.336	13
G8S20Q196	164789	2	0.23	0.001	0.557	11.5
G8S20Q197	157329	2	0.25	0.001	0.545	12.5
G8S20Q198	76220	2	0.85	0.003	0.856	42.5
G8S20Q208	85973	1	0.2	0.001	0.4	20
G8S20Q209	84596	2	0.04	0.001	0.259	2
G8S20Q210	84033	1	0.25	0.001	0.432	25
G8S20Q211	83194	2	0.19	0.002	0.494	9.5
G8S20Q212	82653	2	0.14	0.001	0.421	7
G8S20Q213	82421	1	0.4	0.002	0.489	40
G8S20Q214	81328	2	0.58	0.003	0.822	29
G8S20Q215	79194	1	0.08	0.001	0.275	8
G8S20Q216	75000	2	0.57	0.003	0.717	28.5

g. Item order-wise parameters

Items	N	Max	Mean			
G8S20Q2	89411	1	0.72	0.001	0.447	72
G8S20Q199	88811	1	0.7	0.002	0.458	70
G8S20Q166	180968	1	0.67	0.001	0.47	67
G8S20Q147	88581	1	0.63	0.002	0.482	63
G8S20Q185	176568	1	0.57	0.001	0.495	57
G8S20Q144	180153	1	0.56	0.001	0.496	56
G8S20Q143	180467	1	0.54	0.001	0.498	54
G8S20Q145	88602	1	0.52	0.002	0.499	52
G8S20Q203	88749	1	0.52	0.002	0.499	52
G8S20Q121	89391	1	0.5	0.002	0.5	50

Items	N	Max	Mean			
G8S20Q184	87991	1	0.5	0.002	0.5	50
G8S20Q125	177848	1	0.49	0.001	0.5	49
G8S20Q141	180738	1	0.48	0.001	0.5	48
G8S20Q178	79038	2	0.96	0.003	0.87	48
G8S20Q188	87885	1	0.47	0.002	0.499	47
G8S20Q133	85346	1	0.46	0.002	0.498	46
G8S20Q146	88581	1	0.45	0.002	0.497	45
G8S20Q122	89391	1	0.44	0.002	0.497	44
G8S20Q127	177732	1	0.44	0.001	0.497	44
G8S20Q128	177629	1	0.43	0.001	0.496	43
G8S20Q205	88687	1	0.43	0.002	0.495	43
G8S20Q153	83825	2	0.86	0.003	0.868	43
G8S20Q198	76220	2	0.85	0.003	0.856	42.5
G8S20Q161	181689	1	0.42	0.001	0.493	42
G8S20Q187	87971	1	0.42	0.002	0.494	42
G8S20Q200	88811	1	0.42	0.002	0.493	42
G8S20Q126	177848	1	0.41	0.001	0.492	41
G8S20Q165	181373	1	0.41	0.001	0.492	41
G8S20Q181	176678	1	0.41	0.001	0.492	41
G8S20Q213	82421	1	0.4	0.002	0.489	40
G8S20Q162	93715	1	0.39	0.002	0.488	39
G8S20Q167	93529	1	0.38	0.002	0.486	38
G8S20Q193	168901	2	0.75	0.002	0.789	37.5
G8S20Q129	177318	1	0.37	0.001	0.484	37
G8S20Q160	181978	1	0.37	0.001	0.484	37
G8S20Q168	93509	1	0.37	0.002	0.484	37
G8S20Q204	88708	1	0.37	0.002	0.483	37
G8S20Q148	88581	1	0.36	0.002	0.48	36
G8S20Q140	181008	1	0.35	0.001	0.476	35
G8S20Q124	89391	1	0.34	0.002	0.474	34
G8S20Q149	88581	1	0.34	0.002	0.474	34
G8S20Q180	176678	1	0.34	0.001	0.475	34
G8S20Q120	89391	1	0.33	0.002	0.47	33
G8S20Q159	71542	2	0.66	0.003	0.702	33
G8S20Q132	85928	2	0.65	0.003	0.747	32.5
G8S20Q93	89411	1	0.32	0.002	0.468	32
G8S20Q49	89391	1	0.32	0.002	0.466	32
G8S20Q170	93488	1	0.32	0.002	0.466	32
G8S20Q190	84572	2	0.62	0.003	0.811	31
G8S20Q139	74801	1	0.3	0.002	0.457	30

Items	N	Max	Mean			
G8S20Q214	81328	2	0.58	0.003	0.822	29
G8S20Q216	75000	2	0.57	0.003	0.717	28.5
G8S20Q53	89411	1	0.28	0.001	0.448	28
G8S20Q182	176630	1	0.28	0.001	0.447	28
G8S20Q202	88770	1	0.28	0.002	0.448	28
G8S20Q155	171006	2	0.56	0.002	0.741	28
G8S20Q59	89411	1	0.26	0.001	0.44	26
G8S20Q164	181497	1	0.26	0.001	0.439	26
G8S20Q172	85979	2	0.52	0.003	0.752	26
G8S20Q210	84033	1	0.25	0.001	0.432	25
G8S20Q169	93509	1	0.24	0.001	0.429	24
G8S20Q207	88062	1	0.24	0.001	0.429	24
G8S20Q174	168158	1	0.24	0.001	0.429	24
G8S20Q142	180572	1	0.23	0.001	0.419	23
G8S20Q183	176630	1	0.23	0.001	0.422	23
G8S20Q123	89391	1	0.22	0.001	0.415	22
G8S20Q179	67937	2	0.43	0.003	0.698	21.5
G8S20Q154	171936	2	0.42	0.002	0.691	21
G8S20Q206	88687	1	0.2	0.001	0.401	20
G8S20Q208	85973	1	0.2	0.001	0.4	20
G8S20Q135	167546	2	0.39	0.001	0.585	19.5
G8S20Q138	161686	2	0.39	0.002	0.671	19.5
G8S20Q136	248493	2	0.37	0.001	0.677	18.5
G8S20Q177	164394	2	0.36	0.002	0.691	18
G8S20Q192	83425	1	0.18	0.001	0.381	18
G8S20Q186	87971	1	0.17	0.001	0.375	17
G8S20Q201	88790	1	0.17	0.001	0.377	17
G8S20Q151	84231	2	0.33	0.002	0.638	16.5
G8S20Q103	84622	1	0.15	0.001	0.352	15
G8S20Q191	84047	1	0.15	0.001	0.356	15
G8S20Q171	87241	2	0.29	0.002	0.602	14.5
G8S20Q194	166328	2	0.29	0.001	0.589	14.5
G8S20Q115	83774	1	0.14	0.001	0.345	14
G8S20Q158	162727	2	0.28	0.001	0.458	14
G8S20Q163	93674	1	0.13	0.001	0.331	13
G8S20Q189	87865	1	0.13	0.001	0.341	13
G8S20Q134	169957	2	0.26	0.001	0.442	13
G8S20Q156	167115	1	0.13	0.001	0.332	13
G8S20Q157	242484	1	0.13	0.001	0.34	13
G8S20Q195	165326	1	0.13	0.001	0.336	13

Items	N	Max	Mean			
G8S20Q197	157329	2	0.25	0.001	0.545	12.5
G8S20Q196	164789	2	0.23	0.001	0.557	11.5
G8S20Q130	86176	2	0.19	0.001	0.416	9.5
G8S20Q175	166159	2	0.19	0.001	0.484	9.5
G8S20Q211	83194	2	0.19	0.002	0.494	9.5
G8S20Q69	84167	2	0.18	0.002	0.538	9
G8S20Q152	83915	2	0.18	0.002	0.457	9
G8S20Q77	83981	2	0.16	0.002	0.436	8
G8S20Q137	162640	2	0.16	0.001	0.457	8
G8S20Q150	84541	2	0.16	0.002	0.442	8
G8S20Q215	79194	1	0.08	0.001	0.275	8
G8S20Q131	86011	1	0.07	0.001	0.249	7
G8S20Q106	84498	1	0.07	0.001	0.25	7
G8S20Q176	164916	1	0.07	0.001	0.258	7
G8S20Q212	82653	2	0.14	0.001	0.421	7
G8S20Q173	253405	2	0.08	0.001	0.308	4
G8S20Q209	84596	2	0.04	0.001	0.259	2

h. Classical item parameters in science

Items	N	Maximum	Mean	SE	Std. Deviation
G8S20Q2	89411	1	0.72	0.001	0.447
G8S20Q53	89411	1	0.28	0.001	0.448
G8S20Q59	89411	1	0.26	0.001	0.44
G8S20Q93	89411	1	0.32	0.002	0.468
G8S20Q49	89391	1	0.32	0.002	0.466
G8S20Q120	89391	1	0.33	0.002	0.47
G8S20Q121	89391	1	0.5	0.002	0.5
G8S20Q122	89391	1	0.44	0.002	0.497
G8S20Q123	89391	1	0.22	0.001	0.415
G8S20Q124	89391	1	0.34	0.002	0.474
G8S20Q125	177848	1	0.49	0.001	0.5
G8S20Q126	177848	1	0.41	0.001	0.492
G8S20Q127	177732	1	0.44	0.001	0.497
G8S20Q128	177629	1	0.43	0.001	0.496
G8S20Q129	177318	1	0.37	0.001	0.484
G8S20Q140	181008	1	0.35	0.001	0.476
G8S20Q141	180738	1	0.48	0.001	0.5

Items	N	Maximum	Mean	SE	Std. Deviation
G8S20Q142	180572	1	0.23	0.001	0.419
G8S20Q143	180467	1	0.54	0.001	0.498
G8S20Q144	180153	1	0.56	0.001	0.496
G8S20Q145	88602	1	0.52	0.002	0.499
G8S20Q146	88581	1	0.45	0.002	0.497
G8S20Q147	88581	1	0.63	0.002	0.482
G8S20Q148	88581	1	0.36	0.002	0.48
G8S20Q149	88581	1	0.34	0.002	0.474
G8S20Q160	181978	1	0.37	0.001	0.484
G8S20Q161	181689	1	0.42	0.001	0.493
G8S20Q162	93715	1	0.39	0.002	0.488
G8S20Q163	93674	1	0.13	0.001	0.331
G8S20Q164	181497	1	0.26	0.001	0.439
G8S20Q165	181373	1	0.41	0.001	0.492
G8S20Q166	180968	1	0.67	0.001	0.47
G8S20Q167	93529	1	0.38	0.002	0.486
G8S20Q168	93509	1	0.37	0.002	0.484
G8S20Q169	93509	1	0.24	0.001	0.429
G8S20Q170	93488	1	0.32	0.002	0.466
G8S20Q180	176678	1	0.34	0.001	0.475
G8S20Q181	176678	1	0.41	0.001	0.492
G8S20Q182	176630	1	0.28	0.001	0.447
G8S20Q183	176630	1	0.23	0.001	0.422
G8S20Q184	87991	1	0.5	0.002	0.5
G8S20Q185	176568	1	0.57	0.001	0.495
G8S20Q186	87971	1	0.17	0.001	0.375
G8S20Q187	87971	1	0.42	0.002	0.494
G8S20Q188	87885	1	0.47	0.002	0.499
G8S20Q189	87865	1	0.13	0.001	0.341
G8S20Q199	88811	1	0.7	0.002	0.458
G8S20Q200	88811	1	0.42	0.002	0.493
G8S20Q201	88790	1	0.17	0.001	0.377
G8S20Q202	88770	1	0.28	0.002	0.448
G8S20Q203	88749	1	0.52	0.002	0.499
G8S20Q204	88708	1	0.37	0.002	0.483

Items	N	Maximum	Mean	SE	Std. Deviation
G8S20Q205	88687	1	0.43	0.002	0.495
G8S20Q206	88687	1	0.2	0.001	0.401
G8S20Q207	88062	1	0.24	0.001	0.429
G8S20Q130	86176	2	0.19	0.001	0.416
G8S20Q131	86011	1	0.07	0.001	0.249
G8S20Q132	85928	2	0.65	0.003	0.747
G8S20Q133	85346	1	0.46	0.002	0.498
G8S20Q103	84622	1	0.15	0.001	0.352
G8S20Q106	84498	1	0.07	0.001	0.25
G8S20Q69	84167	2	0.18	0.002	0.538
G8S20Q77	83981	2	0.16	0.002	0.436
G8S20Q115	83774	1	0.14	0.001	0.345
G8S20Q134	169957	2	0.26	0.001	0.442
G8S20Q135	167546	2	0.39	0.001	0.585
G8S20Q136	248493	2	0.37	0.001	0.677
G8S20Q137	162640	2	0.16	0.001	0.457
G8S20Q138	161686	2	0.39	0.002	0.671
G8S20Q139	74801	1	0.3	0.002	0.457
G8S20Q150	84541	2	0.16	0.002	0.442
G8S20Q151	84231	2	0.33	0.002	0.638
G8S20Q152	83915	2	0.18	0.002	0.457
G8S20Q153	83825	2	0.86	0.003	0.868
G8S20Q154	171936	2	0.42	0.002	0.691
G8S20Q155	171006	2	0.56	0.002	0.741
G8S20Q156	167115	1	0.13	0.001	0.332
G8S20Q157	242484	1	0.13	0.001	0.34
G8S20Q158	162727	2	0.28	0.001	0.458
G8S20Q159	71542	2	0.66	0.003	0.702
G8S20Q171	87241	2	0.29	0.002	0.602
G8S20Q172	85979	2	0.52	0.003	0.752
G8S20Q173	253405	2	0.08	0.001	0.308
G8S20Q174	168158	1	0.24	0.001	0.429
G8S20Q175	166159	2	0.19	0.001	0.484
G8S20Q176	164916	1	0.07	0.001	0.258
G8S20Q177	164394	2	0.36	0.002	0.691

Items	N	Maximum	Mean	SE	Std. Deviation
G8S20Q178	79038	2	0.96	0.003	0.87
G8S20Q179	67937	2	0.43	0.003	0.698
G8S20Q190	84572	2	0.62	0.003	0.811
G8S20Q191	84047	1	0.15	0.001	0.356
G8S20Q192	83425	1	0.18	0.001	0.381
G8S20Q193	168901	2	0.75	0.002	0.789
G8S20Q194	166328	2	0.29	0.001	0.589
G8S20Q195	165326	1	0.13	0.001	0.336
G8S20Q196	164789	2	0.23	0.001	0.557
G8S20Q197	157329	2	0.25	0.001	0.545
G8S20Q198	76220	2	0.85	0.003	0.856
G8S20Q208	85973	1	0.2	0.001	0.4
G8S20Q209	84596	2	0.04	0.001	0.259
G8S20Q210	84033	1	0.25	0.001	0.432
G8S20Q211	83194	2	0.19	0.002	0.494
G8S20Q212	82653	2	0.14	0.001	0.421
G8S20Q213	82421	1	0.4	0.002	0.489
G8S20Q214	81328	2	0.58	0.003	0.822
G8S20Q215	79194	1	0.08	0.001	0.275
G8S20Q216	75000	2	0.57	0.003	0.717

Appendix 3

Classical results for Nepali

a. Classical results in Nepali by caste/ethnicity

Ethnic group	N	Mean	Std. Deviation	Std. Error of Mean
Madhesi Brahman	37167	37.84	22.51	0.12
Madhesi Janjati	55579	36.71	20.94	0.09
Madhesi Dalit	14481	33.70	20.97	0.17
Madhesi Others	31714	35.92	21.52	0.12
Hill Brahman	121238	43.69	20.86	0.06
Hill Janjati	82348	42.71	19.68	0.07
Hill Dalit	30295	40.09	20.18	0.12
Hill Others	15884	40.76	19.45	0.15
Mountain Brahman	6535	37.65	23.25	0.29
Mountain Janjati	4040	39.76	19.68	0.31
Mountain Dalit	1292	36.97	24.39	0.68
Mountain Others	1337	30.52	19.77	0.54
Total	401910	40.42	21.04	0.03

b. Classical results in Nepali by medium of instruction

	N	Mean	Std. Deviation	Std. Error of Mean
Nepali	278851	37.58	20.41	0.04
English	95464	45.86	21.68	0.07
Other/Not Specified	80571	36.00	22.51	0.08
Total	454887	39.04	21.37	0.03

c. Classical results in Nepali by mother's education

Mother's education level	N	Mean	Std. Deviation	Std. Error of Mean
Illiterate	142241	36.57	20.22	0.05
Literate	112945	40.88	20.21	0.06
Grade 8	83128	38.67	21.02	0.07
Grade 10	53105	42.38	22.39	0.10
Grade 12	30853	44.00	22.53	0.13
Bachelor's	10610	49.10	22.38	0.22
Master's or above	5538	44.70	23.63	0.32
Total	438420	39.71	21.11	0.03

d. Classical results in Nepali by father's education

Father's education level	N	Mean	Std. Deviation	Std. Error of Mean
Illiterate	61954	33.88	20.48	0.08
Literate	87184	40.64	20.00	0.07
Grade 8	108180	37.55	20.61	0.06
Grade 10	98550	40.01	21.27	0.07
Grade 12	50807	42.83	21.53	0.10
Bachelor	21753	47.09	22.04	0.15
Master's or above	11251	50.16	21.82	0.21
Total	439680	39.60	21.14	0.03

e. Classical results in Nepali by mother's profession

Mother's occupation	N	Mean	Std. Deviation	Std. Error of Mean
Agriculture + household	262975	38.08	20.61	0.04
Household	96559	40.71	21.61	0.07
Work in other home	5071	35.74	22.07	0.31
Labour	5505	40.05	20.61	0.28
Foreign	7990	40.30	21.59	0.24
Teaching	12059	48.11	22.32	0.20
Business	31043	43.41	20.66	0.12
Government job	8812	45.76	21.55	0.23
Others	9050	44.41	21.62	0.23
Total	439064	39.63	21.10	0.03

f. Classical results in Nepali by father's profession

Father's occupation	N	Mean	Std. Deviation	Std. Error of Mean
Agriculture + household	130001	36.50	20.76	0.06
Household	10962	31.13	20.37	0.19
Work in other home	9063	33.60	19.55	0.21
Labour	33620	41.91	20.26	0.11
Foreign	104663	39.65	20.80	0.06
Teaching	14296	46.07	22.70	0.19
Business	65455	41.77	21.04	0.08
Government job	32528	44.16	20.93	0.12
Others	32238	43.18	21.38	0.12
Total	432827	39.67	21.12	0.03

Classical score in percentage by type of local level (Rural and Urban municipality) in English

In NASA 2020 study, classical score in percentage of English by the type of local levels vary as per the facilities and opportunities provided to students. The details of the mean classical score in percentage have been provided in table 167.

Table 203 Classical score in percentage by type of local level (Rural and Urban municipality)

Local Level	N	Mean	Std. Deviation	Std. Error of Mean
Rural Municipality	149300	16.955	18.546	0.048
Municipality	276501	29.678	26.085	0.050
Total	425802	25	24.480	0.038

Table 167 shows that the classical score in percentage score of students in English found to be around double with respect to municipality (Mean=29.68, SD=24.48) than their rural municipality (Mean=16.96, SD=18.55) counterpart. This indicates division of two classes of citizen by school education in Nepal.

g. Item-wise percentage of correct answers in Nepali

S.No.	Item ID	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of Correct Answer
1	G8N20q1_scored	169190	0	1	0.47	0.50	47
2	G8N20q2_scored	169219	0	1	0.65	0.48	65
3	G8N20q3_scored	168815	0	1	0.42	0.49	42
4	G8N20q4_scored	163488	0	1	0.46	0.50	46
5	G8N20q5_scored	160122	0	1	0.43	0.50	43
6	G8N20q6_scored	163075	0	2	0.64	0.70	32
7	G8N20q7_scored	153745	0	1	0.32	0.47	32
8	G8N20q8_scored	157030	0	2	0.61	0.66	31
9	G8N20q9_scored	156598	0	3	0.82	0.80	27
10	G8N20q10_scored	178270	0	1	0.60	0.49	60
11	G8N20q11_scored	178210	0	1	0.64	0.48	64
12	G8N20q12_scored	176162	0	1	0.62	0.49	62
13	G8N20q13_scored	174557	0	1	0.74	0.44	74
14	G8N20q14_scored	168682	0	1	0.52	0.50	52
15	G8N20q15_scored	168651	0	2	0.65	0.73	33
16	G8N20q16_scored	164349	0	1	0.34	0.47	34
17	G8N20q17_scored	164936	0	3	0.85	0.86	28
18	G8N20q18_scored	179435	0	1	0.84	0.37	84
19	G8N20q19_scored	176721	0	1	0.76	0.43	76
20	G8N20q20_scored	166420	0	1	0.42	0.49	42

S.No.	Item ID	N	Minimum	Maximum	Mean	Std. Deviation	Percentage of Correct Answer
21	G8N20q21_scored	170957	0	2	0.81	0.75	41
22	G8N20q22_scored	161817	0	1	0.53	0.50	53
23	G8N20q23_scored	86113	0	2	0.86	0.76	43
24	G8N20q24_scored	166357	0	4	1.20	1.06	30
25	G8N20q25_scored	169060	0	1	0.67	0.47	67
26	G8N20q26_scored	170596	0	1	0.71	0.45	71
27	G8N20q27_scored	168448	0	1	0.30	0.46	30
28	G8N20q28_scored	162191	0	2	0.37	0.62	19
29	G8N20q29_scored	160705	0	1	0.44	0.50	44
30	G8N20q30_scored	162055	0	2	0.67	0.77	34
31	G8N20q31_scored	167467	0	4	1.35	0.95	34
32	G8N20q32_scored	162890	0	1	0.54	0.50	54
33	G8N20q33_scored	154893	0	1	0.33	0.47	33
34	G8N20q34_scored	159258	0	1	0.67	0.47	67
35	G8N20q35_scored	147709	0	1	0.33	0.47	33
36	G8N20q36_scored	145105	0	2	0.45	0.74	23
37	G8N20q37_scored	151144	0	3	0.93	0.93	31

Appendix 3

IRT parameters of Nepali items

SN	item	alpha	beta	tau.Cat1	tau.Cat2	tau.Cat3
1	G8N20q1S	0.743	0.178	NA	NA	NA
2	G8N20q2S	0.596	-1.099	NA	NA	NA
3	G8N20q3S	0.607	0.601	NA	NA	NA
4	G8N20q4S	0.415	0.415	NA	NA	NA
5	G8N20q5S	1.248	0.26	NA	NA	NA
6	G8N20q6S	0.78	1.013	-0.816	0.816	NA
7	G8N20q7S	1.282	0.731	NA	NA	NA
8	G8N20q8S	0.939	1.091	-1.046	1.046	NA
9	G8N20q9S	1.038	0.402	-0.531	0.531	NA
10	G8N20q10S	1.332	-0.399	NA	NA	NA
11	G8N20q11S	1.267	-0.589	NA	NA	NA
12	G8N20q12S	1.27	-0.483	NA	NA	NA
13	G8N20q13S	0.844	-1.411	NA	NA	NA
14	G8N20q14S	1.253	-0.097	NA	NA	NA
15	G8N20q15S	0.843	0.907	-0.614	0.614	NA
16	G8N20q16S	1.016	0.754	NA	NA	NA
17	G8N20q17S	0.991	0.336	-0.061	0.061	NA
18	G8N20q18S	2.005	-1.299	NA	NA	NA
19	G8N20q19S	1.976	-0.909	NA	NA	NA
20	G8N20q20S	1.417	0.354	NA	NA	NA
21	G8N20q21S	1.185	0.455	-0.671	0.671	NA
22	G8N20q22S	1.477	-0.088	NA	NA	NA
23	G8N20q23S	1.052	0.398	-0.664	0.664	NA
24	G8N20q24S	1.045	0.545	-0.546	-0.515	1.061
25	G8N20q25S	0.95	-0.805	NA	NA	NA
26	G8N20q26S	1.177	-0.885	NA	NA	NA
27	G8N20q27S	0.223	3.851	NA	NA	NA
28	G8N20q28S	0.876	1.737	-0.337	0.337	NA
29	G8N20q29S	1.198	0.274	NA	NA	NA
30	G8N20q30S	1.029	0.766	-0.294	0.294	NA
31	G8N20q31S	0.989	0.463	-1.104	-0.67	1.774
32	G8N20q32S	0.896	-0.15	NA	NA	NA
33	G8N20q33S	-0.358	-1.915	NA	NA	NA

SN	item	alpha	beta	tau.Cat1	tau.Cat2	tau.Cat3
34	G8N20q34S	1.871	-0.557	NA	NA	NA
35	G8N20q35S	1.257	0.765	NA	NA	NA
36	G8N20q36S	0.819	1.301	0.719	-0.719	NA
37	G8N20q37S	1.003	0.221	0.298	-0.298	NA

Appendix 3. IRT parameters of science items

SN	item	alpha	beta	tau.Cat1	tau.Cat2
1	G8S20Q2	0.807	-1.305	NA	NA
2	G8S20Q53	0.15	6.427	NA	NA
3	G8S20Q59	0.2	5.245	NA	NA
4	G8S20Q93	0.346	2.23	NA	NA
5	G8S20Q49	0.178	4.393	NA	NA
6	G8S20Q120	0.547	1.442	NA	NA
7	G8S20Q121	0.282	0.033	NA	NA
8	G8S20Q122	0.735	0.419	NA	NA
9	G8S20Q123	0.594	2.329	NA	NA
10	G8S20Q124	0.398	1.788	NA	NA
11	G8S20Q125	0.864	0.09	NA	NA
12	G8S20Q126	0.765	0.582	NA	NA
13	G8S20Q127	0.7	0.402	NA	NA
14	G8S20Q128	0.724	0.462	NA	NA
15	G8S20Q129	0.7	0.862	NA	NA
16	G8S20Q140	0.79	0.896	NA	NA
17	G8S20Q141	1.259	0.074	NA	NA
18	G8S20Q142	0.611	2.149	NA	NA
19	G8S20Q143	0.992	-0.197	NA	NA
20	G8S20Q144	0.93	-0.31	NA	NA
21	G8S20Q145	0.826	-0.083	NA	NA
22	G8S20Q146	0.947	0.309	NA	NA
23	G8S20Q147	1.227	-0.516	NA	NA
24	G8S20Q148	0.522	1.218	NA	NA
25	G8S20Q149	0.742	1.031	NA	NA
26	G8S20Q160	0.707	0.797	NA	NA
27	G8S20Q161	1.031	0.371	NA	NA
28	G8S20Q162	0.939	0.505	NA	NA
29	G8S20Q163	0.376	5.243	NA	NA
30	G8S20Q164	0.464	2.333	NA	NA
31	G8S20Q165	0.829	0.472	NA	NA

SN	item	alpha	beta	tau.Cat1	tau.Cat2
32	G8S20Q166	0.437	-1.717	NA	NA
33	G8S20Q167	0.837	0.593	NA	NA
34	G8S20Q168	0.648	0.818	NA	NA
35	G8S20Q169	0.365	3.147	NA	NA
36	G8S20Q170	0.657	1.212	NA	NA
37	G8S20Q180	0.634	1.105	NA	NA
38	G8S20Q181	0.295	1.267	NA	NA
39	G8S20Q182	0.509	1.988	NA	NA
40	G8S20Q183	0.236	5.105	NA	NA
41	G8S20Q184	1.082	0.036	NA	NA
42	G8S20Q185	0.62	-0.531	NA	NA
43	G8S20Q186	0.627	2.759	NA	NA
44	G8S20Q187	0.917	0.435	NA	NA
45	G8S20Q188	0.282	0.496	NA	NA
46	G8S20Q189	0.415	4.648	NA	NA
47	G8S20Q199	0.681	-1.417	NA	NA
48	G8S20Q200	0.512	0.664	NA	NA
49	G8S20Q201	0.43	3.74	NA	NA
50	G8S20Q202	0.547	1.813	NA	NA
51	G8S20Q203	0.437	-0.264	NA	NA
52	G8S20Q204	0.631	0.853	NA	NA
53	G8S20Q205	0.648	0.433	NA	NA
54	G8S20Q206	0.637	2.288	NA	NA
55	G8S20Q207	0.585	2.026	NA	NA
56	G8S20Q130	0.89	3.102	-0.957	0.957
57	G8S20Q131	1.722	2.219	NA	NA
58	G8S20Q132	1.094	0.855	-0.458	0.458
59	G8S20Q133	1.249	0.255	NA	NA
60	G8S20Q103	1.512	1.682	NA	NA
61	G8S20Q106	1.865	2.135	NA	NA
62	G8S20Q69	0.806	2.192	2.141	-2.141
63	G8S20Q77	1.254	2.215	-0.054	0.054
64	G8S20Q115	1.332	1.86	NA	NA
65	G8S20Q134	1.025	3.702	-2.404	2.404
66	G8S20Q135	0.923	1.878	-0.807	0.807
67	G8S20Q136	1.297	1.32	0.179	-0.179
68	G8S20Q137	1.196	2.119	0.312	-0.312
69	G8S20Q138	1.09	1.412	0.041	-0.041
70	G8S20Q139	1.513	0.891	NA	NA

SN	item	alpha	beta	tau.Cat1	tau.Cat2
71	G8S20Q150	0.695	2.973	0.363	-0.363
72	G8S20Q151	1.231	1.451	0.167	-0.167
73	G8S20Q152	0.389	4.581	0.82	-0.820
74	G8S20Q153	1.213	0.314	0.081	-0.081
75	G8S20Q154	1.127	1.264	0.031	-0.031
76	G8S20Q155	0.842	1.111	-0.137	0.137
77	G8S20Q156	1.738	1.642	NA	NA
78	G8S20Q157	1.42	1.753	NA	NA
79	G8S20Q158	1.698	2.428	-1.572	1.572
80	G8S20Q159	0.552	1.38	-1.05	1.050
81	G8S20Q171	0.924	1.763	0.326	-0.326
82	G8S20Q172	0.814	1.163	0.231	-0.231
83	G8S20Q173	1.173	2.751	0.128	-0.128
84	G8S20Q174	0.809	1.625	NA	NA
85	G8S20Q175	1.225	2.013	0.011	-0.011
86	G8S20Q176	1.337	2.459	NA	NA
87	G8S20Q177	1.321	1.279	0.403	-0.403
88	G8S20Q178	0.869	0.132	0.199	-0.199
89	G8S20Q179	1.74	1.154	-0.201	0.201
90	G8S20Q190	1.471	0.717	0.015	-0.015
91	G8S20Q191	1.615	1.559	NA	NA
92	G8S20Q192	1.191	1.666	NA	NA
93	G8S20Q193	1.156	0.518	-0.353	0.353
94	G8S20Q194	1.617	1.433	-0.13	0.130
95	G8S20Q195	2.267	1.426	NA	NA
96	G8S20Q196	1.755	1.508	0.085	-0.085
97	G8S20Q197	1.512	1.624	-0.135	0.135
98	G8S20Q198	0.762	0.419	0.256	-0.256
99	G8S20Q208	1.243	1.376	NA	NA
100	G8S20Q209	0.898	3.098	1.746	-1.746
101	G8S20Q210	1.448	1.012	NA	NA
102	G8S20Q211	1.539	1.717	-0.004	0.004
103	G8S20Q212	1.102	2.253	0.198	-0.198
104	G8S20Q213	1.545	0.365	NA	NA
105	G8S20Q214	0.949	0.821	0.635	-0.635
106	G8S20Q215	1.749	1.927	NA	NA
107	G8S20Q216	0.998	0.995	-0.444	0.444

Appendix 4

Classical results for English

a. Classical results in English by caste and ethnicity

Ethnic group	N	Mean	Std. Deviation	Std. Error of Mean
Madhesi Brahmin	37300	27.636	25.591	0.133
Madhesi Janjati	55662	22.470	22.549	0.096
Madhesi Dalit	14564	18.610	21.019	0.174
Madhesi Others	31776	22.288	22.909	0.129
Hill Brahmin	121950	30.331	26.671	0.076
Hill Janjati	82727	26.680	24.407	0.085
Hill Dalit	30468	19.048	19.728	0.113
Hill Others	16013	25.470	25.013	0.198
Mountain Brahmin	6576	20.728	22.134	0.273
Mountain Janjati	4060	24.480	22.617	0.355
Mountain Dalit	1292	22.493	21.672	0.603
Mountain Others	1337	19.194	23.531	0.644
Total	403725	25	24.709	0.039

b. Classical results in English by medium of instruction

Medium of instruction	N	Mean	Std. Deviation	Std. Error of Mean
Nepali	279931	16.383	17.216	0.033
English	96113	50.493	24.851	0.080
Not Stated	81090	23.330	24.199	0.085
Total	457133	25	24.449	0.036

c. Classical results in English by mother's education

Mother's education level	N	Mean	Std. Deviation	Std. Error of Mean
Illiterate	142957	17.057	18.639	0.049
Literate	113357	21.568	21.460	0.064
Grade 8	83566	25.612	23.887	0.083
Grade 10	53395	36.683	26.941	0.117
Grade 12	30915	42.475	28.846	0.164
Bachelor's	10692	50.975	27.285	0.264
Master's or above	5580	49.267	29.053	0.389

d. Classical results by father's education

Father's education level	N	Mean	Std. Deviation	Std. Error of Mean
Illiterate	62413	15.006	17.454	0.070
Literate	87457	19.588	20.123	0.068
Grade 8	108584	20.037	20.644	0.063
Grade 10	99042	28.614	25.138	0.080
Grade 12	51055	36.658	27.431	0.121
Bachelor's	21857	43.879	28.348	0.192
Master's or above	11251	54.203	26.731	0.252

e. Classical results in English by mother's profession

Mother's occupation	N	Mean	Std. Deviation	Std. Error of Mean
Agriculture + household	264344	18.753	20.139	0.039
Household	96854	33.167	26.685	0.086
Work in other home	5113	24.111	24.056	0.336
Labour	5505	25.840	23.328	0.314
Foreign	8073	28.639	24.677	0.275
Teaching	12121	43.785	28.978	0.263
Business	31188	36.582	26.989	0.153
Government job	8853	39.133	27.398	0.291
Others	9099	44.752	26.760	0.281

f. Classical results in English by father's profession

Father's occupation	N	Mean	Std. Deviation	Std. Error of Mean
Agriculture + household	130654	16.110	18.019	0.050
Household	11066	16.657	18.396	0.175
Work in other home	9126	14.920	17.074	0.179
Labour	33641	21.527	21.213	0.116
Foreign	105171	25.046	23.863	0.074
Teaching	14317	37.068	28.356	0.237
Business	65749	33.990	27.213	0.106
Government job	32652	36.346	27.252	0.151
Others	32452	38.125	26.786	0.149

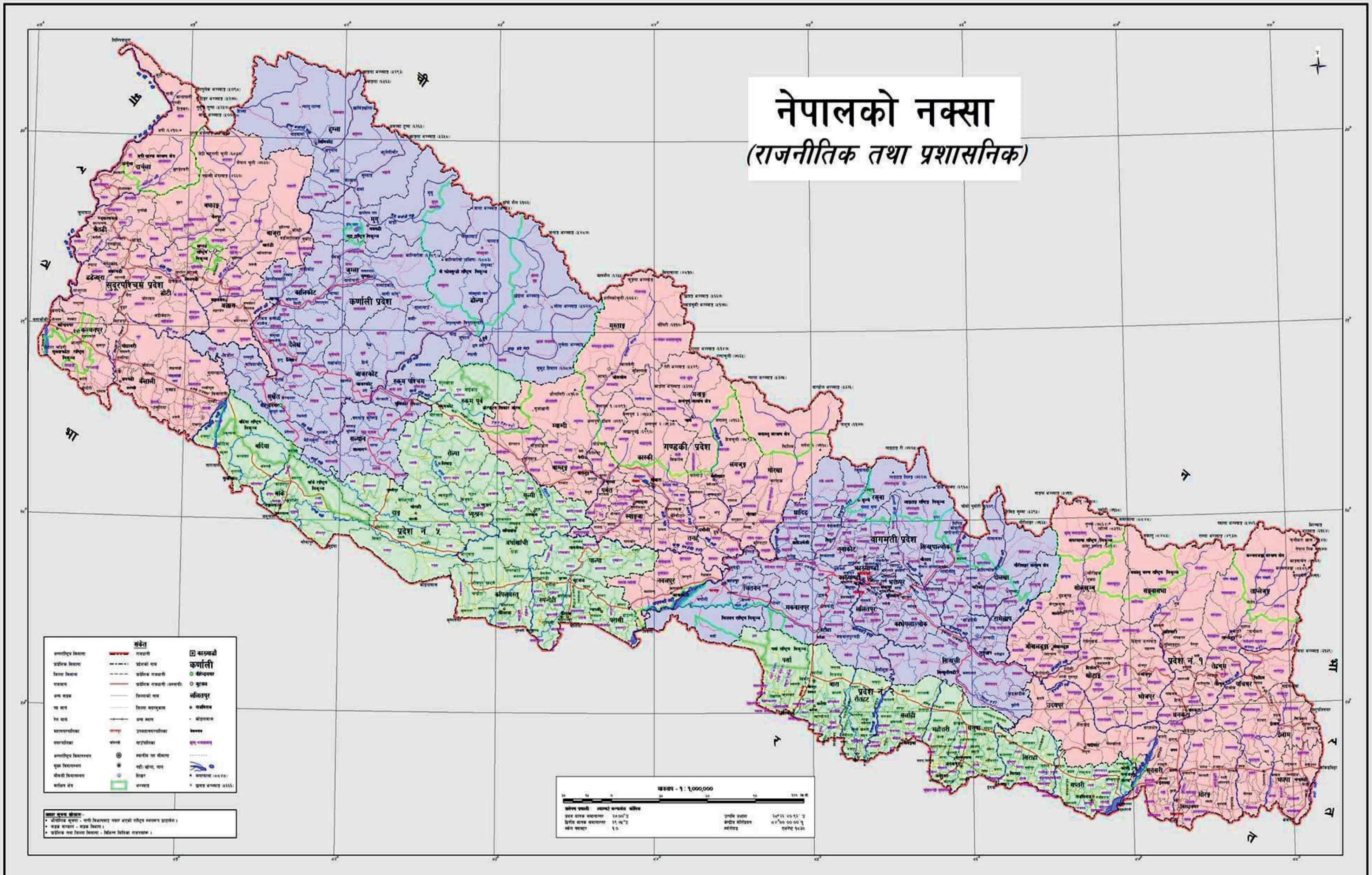
g. Classical item parameters of English

	N	Minimum	Maximum	Mean	Std. Deviation	% of correct answer
G8E20q1s	179582	0	1	0.26	0.44	26
G8E20q2s	179582	0	1	0.34	0.47	34
G8E20q3s	179582	0	1	0.35	0.48	35
G8E20q4s	179582	0	6	1.53	1.74	26
G8E20q5s	186308	0	1	0.44	0.50	44
G8E20q6s	186308	0	1	0.39	0.49	39
G8E20q7s	186308	0	1	0.18	0.39	18
G8E20q8s	186308	0	1	0.24	0.43	24
G8E20q9s	186308	0	6	0.88	1.37	15
G8E20q10s	189268	0	1	0.36	0.48	36
G8E20q11s	189268	0	1	0.42	0.49	42
G8E20q12s	189268	0	1	0.39	0.49	39
G8E20q13s	189268	0	1	0.32	0.47	32
G8E20q14s	189268	0	6	0.97	1.45	16
G8E20q15s	182010	0	1	0.22	0.41	22
G8E20q16s	182010	0	1	0.41	0.49	41
G8E20q17s	182010	0	1	0.37	0.48	37
G8E20q18s	182010	0	1	0.15	0.36	15
G8E20q19s	182010	0	6	1.05	1.59	18
G8E20q20s	177099	0	1	0.45	0.50	45
G8E20q21s	177099	0	1	0.39	0.49	39
G8E20q22s	177099	0	1	0.35	0.48	35
G8E20q23s	177099	0	1	0.34	0.48	34
G8E20q24s	177099	0	6	1.24	1.70	21

Appendix 4. IRT parameters of English items

S.No	item	alpha	beta	tau. Cat1	tau. Cat2	tau. Cat3	tau. Cat4	tau. Cat5	tau. Cat6
1	G8E20q1s	2.613	0.792	NA	NA	NA	NA	NA	NA
2	G8E20q2s	2.687	0.516	NA	NA	NA	NA	NA	NA
3	G8E20q3s	1.468	0.59	NA	NA	NA	NA	NA	NA
4	G8E20q4s	1.293	1.098	-0.744	-0.707	-0.547	-0.196	0.499	1.695
5	G8E20q5s	2.512	0.215	NA	NA	NA	NA	NA	NA
6	G8E20q6s	2.337	0.381	NA	NA	NA	NA	NA	NA
7	G8E20q7s	1.682	1.314	NA	NA	NA	NA	NA	NA
8	G8E20q8s	2.451	0.9	NA	NA	NA	NA	NA	NA
9	G8E20q9s	1.529	1.606	-0.655	-0.906	-0.521	0.022	0.545	1.515
10	G8E20q10s	0.713	0.936	NA	NA	NA	NA	NA	NA

S.No	item	alpha	beta	tau. Cat1	tau. Cat2	tau. Cat3	tau. Cat4	tau. Cat5	tau. Cat6
11	G8E20q11s	1.44	0.345	NA	NA	NA	NA	NA	NA
12	G8E20q12s	1.18	0.51	NA	NA	NA	NA	NA	NA
13	G8E20q13s	0.939	0.969	NA	NA	NA	NA	NA	NA
14	G8E20q14s	1.424	1.653	-0.713	-0.949	-0.685	-0.246	0.341	2.252
15	G8E20q15s	1.862	1.06	NA	NA	NA	NA	NA	NA
16	G8E20q16s	3.478	0.231	NA	NA	NA	NA	NA	NA
17	G8E20q17s	2.385	0.403	NA	NA	NA	NA	NA	NA
18	G8E20q18s	2.45	1.277	NA	NA	NA	NA	NA	NA
19	G8E20q19s	1.47	1.366	-0.474	-0.662	-0.512	-0.228	0.322	1.554
20	G8E20q20s	1.134	0.193	NA	NA	NA	NA	NA	NA
21	G8E20q21s	0.881	0.584	NA	NA	NA	NA	NA	NA
22	G8E20q22s	1.086	0.704	NA	NA	NA	NA	NA	NA
23	G8E20q23s	1.172	0.677	NA	NA	NA	NA	NA	NA
24	G8E20q24s	1.339	1.259	-0.338	-0.856	-0.673	-0.241	0.376	1.732


नेपालको नक्सा
(राजनीतिक तथा प्रशासनिक)
सुदूरपश्चिम प्रदेश
कर्णाली प्रदेश
गण्डकी प्रदेश
बागमती प्रदेश
प्रदेश नं. १
प्रदेश नं. २
प्रदेश नं. ५
संकेत