PRIORITIZING EDUCATIONAL POLICY FROM NASA RESULTS: META-ANALYSIS OF ALL NASA REPORTS FROM 2011 TO 2022

Shyam Prasad Acharya

Dr. Acharya, Section Officer, Curriculum Development Centre, having experience in Mathematics and NASA. Correspondence regarding this article can be addressed to his address. Email: shyamacharya2024@gmail.com.

Abstract

This article aims to provide insights on how to set priorities in education policies withthe NASA results from the equity perspective. I have tried to extract a few examples of NASA results and analyzed the strengths of different variables to show how those variables have influenced in student learning and why those variables are important from the policy formulation point of view. Meta-analysis does not limit on a single year and subject, rather it is extended over years and subjects in the specific areas. From the analysis, it is revealed that provincial differences and mother education are the upcoming focus in educational priorities. There are other important variables pointed out that has to be snatched from this article.

Keywords: prioritizing education policy, meta-analysis, effect size, mother education, learning assessment results

Background

This article is a meta-analysis of National Assessment of Student Achievement (NASA) datasets and reports from 2011 to 2020 in specific areas. A similar analysis should be done to explore all effective variables to prioritize national interventions.

National Assessment of Student Achievement (NASA) is a national level student learning assessment. It adopts the standards adopted by international assessments like TIMSS and PISA in sampling. As it collects data from a Probability Proportional to Size (PPS) method which is very popular in large scale assessments. While formulating the School Sector Reform Program (2009-2015), NASA has been a key strategy of measuring progress over the years. Education Review Office was established to conduct NASA of grades 3, 5, and 8. But in 2016, School Sector Development Program (2016-2022), grade 10 was in and grade 3 was out from the National Assessment program. In all year NASA, Item Response Theory (IRT) was a tool to standardize the NASA process and compare the results over the years. Therefore, throughout the process of item analysis and result preparation, Item Response Theory (IRT) was used in every assessment from 2011 to the date. This is the main reason why NASA's results are highly reliable and trustable. All the NASA assessments and results of previous assessments are compared with the successive

assessments, which is possible because of the linked items used in all assessments with the previous assessments. Another reason why it is trustworthy is that data collected from the schools are independently collected by a third party. Similarly, a third party is also included while writing a report so that necessary expertise is utilized from the national level and international level. Moreover, there is direct and indirect involvement of development partners as a technical support. At the beginning, technical support was provided through hiring a consultant and in later years, capacity development was supported by the development partners and assessment administration and data analysis was done by the Education Review Office itself. In this scenario, the Education Review Office is able to conduct and report such a big technical work with its own manpower.

National Assessment of Student Achievement is a largescale standardized assessment conducted to measure the academic performance and achievement of students within the country, with seven provinces as explicit stratum. So, the results of these assessments are generalizable throughout the nation and over the provinces. The main objective of this assessment is to provide policymakers, educators, and researchers with valuable insights and policy recommendations by uncovering strengths and weaknesses of the education system that support policy maker's reforms and improve national priorities in the education subsector. The specific details and organization of national assessments may vary from country to country.

Typically, NASA covers core subjects such as mathematics, reading, science, and sometimes additional subjects like writing or Social Studies. The assessments are administered to a representative sample of students from various regions or schools within the country, ensuring that the results reflect the overall performance of the student population. National assessments provide reliable measure of students' knowledge, skills, and competencies, enabling educational stakeholders to monitor trends, identify achievement gaps, and make data-driven decisions to enhance the quality of education. The results often serve as a benchmark against international standards or other countries through linked international test-items and can be compared internationally in a limited strength.

A complete NASA cycle goes over a period of 3 years. In the first year, tasks related to test items (item development, pre-test and analysis) are completed. In the second year, final NASA assessment is administered. At last, in the third year, activities like data analysis, report writing, result dissemination of the assessment and policy feedbacks are performed.

Objectives of National Assessment of Student Achievement (NASA) Studies

There are some common objectives stated in all NASA studies. Those common objectives are listed below:

1. Find out the current status of learning achievement in Mathematics, Nepali language, Science and English language.
2. Identify influencing factors to the learning achievement.
3. Find out trend in student learning
4. Provide policy level recommendations to policy makers, stakeholders and researchers.
5. Provide evidences to reform curriculum, and teachers' capacity development programs.

Methodology

To start this meta-analysis, data were collected from the National Assessment of Student Achievement (NASA) studies and reports published including statistical report (unpublished one) from the Education Review Office (ERO). The ERO database contains detailed information on student performance, including test scores and demographic details, from various NASA studies. The gathered data spanning multiple years to see how things have changed over time. The collected data were smothered to remove any errors and fill in any missing information, ensuring reliable data for this analysis.

The main statistical method used in this analysis was One-Way ANOVA. This technique helps to compare the average scores of different groups of students to see if there are significant differences between them. These groups are based on factors like the year the assessments were taken, and student demographics from the assessments they participated in. For each group comparison, an ANOVA test was used with a significance level of 0.05. This means we are 95% confident that the results are not due to chance. From this analysis, F-statistics and p-values were calculated to compare the significant differences. These values are not stated in this article because only significant differences at 95% confidence level are discussed.

Next, I calculated the effect size to measure how big the differences between groups are. Depending on the data, either Cohen's d or eta-squared (η^2) are used. Cohen's d helps us understand the size of the difference between two group averages, while eta-squared shows the proportion of the total variance that is due to the differences between groups. These calculations help us interpret how meaningful the differences are, classifying them as small, medium, or large. This step is crucial for understanding the practical implications of our findings.

The possible charts are prepared based on the data and parameters calculated. At last, summaries are presented to interpret the results to make them understandable by the readers.

The analysis of the data was done already in concerned NASA cycles. The approach and software used to analyze the data in those NASA cycles is presented in the table 1.

This table is customized from the statistical report 2023 (ERO, 2023)

NASA Year	Technological Shift	Software used	Grade	Subjects	Sample size
2011	Used IRT Percentage of score are reported	OPLM	8	Nepali, Social studies, Mathematics	48682
2012	Used IRT Percentage of score are reported	OPLM	3 and 5	Grade 3: Nepali and Mathematics Grade 5: Nepali, Mathematics and English	80232 Grade 3: 38753 Grade 5: 41479
2013	Used IRT Percentage of score are reported	OPLM	8	Nepali, Mathematics, Science	44067
2015	Used IRT Percentage of score are reported	OPLM	3 and 5	Grade 3: Nepali and Mathematics Grade 5: Nepali, English, Mathematic	73878 Grade 3: 33863 Grade 5: 40015
2017	IRT in full fledged Plausible value-based reporting	Acer conQuest	8	Nepali, Science and Mathematics	46266
2018	IRT in full fledged Plausible value-based reporting	Acer conQuest	5	Nepali and Mathematics	32262
2019	IRT in full fledged Plausible value-based reporting	Acer conQuest	10	Nepali, Mathematics, English and Science	43886
2020	IRT in full fledged Plausible value based reporting	R – open source software	8	Nepali, English, Mathematics and Science	43497

Adapted from NASA Statistical Report 2023 (Unpublished).

The table 1 reveals that all the NASA data analysis process are precious, of high standard and used the standard software that are popular in large scale assessments.

Results

In this section, I have presented a few key variables associated with student learning and their effect on student achievement. Difference in *mean* can give insights to locate the differences. However, mean achievement only does not reveal the extent to which differences are effect on student learning. For this sake, effect size has to be calculated. So, interpretations are presented based on the effect size on student learning with two variables by using Cohen's d and on many variables, partial eta square is used. Presentation of the results are done chronologically from grade 10 to grade 5. Since, there is only one assessment result in grade 10, meta-analysis is carried out over the subjects and variables. While in grade 8 and 5, there are multiple cycles of assessments. So, multiple year data and multiple subjects are analyzed in a limited areas representative manner.

Results of Grade 10 NASA Results

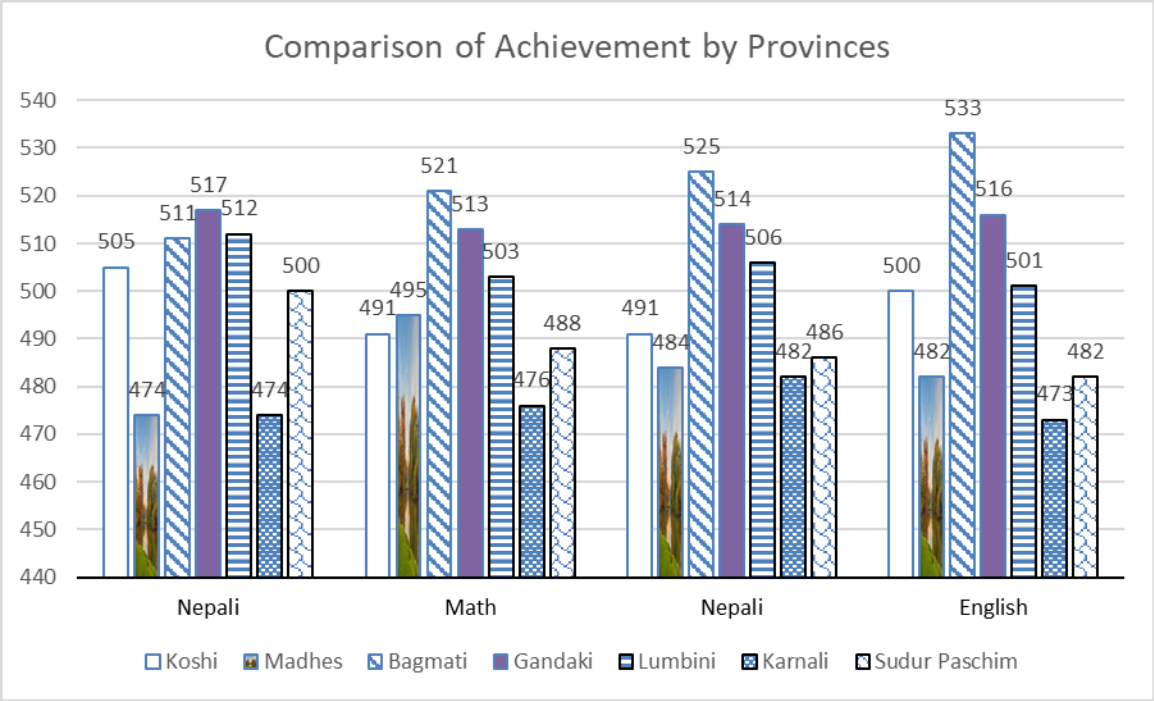
In 2019, grade 10 NASA was conducted in Science, Mathematics, Nepali and English nationwide on about 46000 students. The sample was taken from 7 provinces, provinces as an explicit stratum.

Achievement byProvinces

This section presents the transformed scale score (mean at 500) of grade 10 (NASA 2019) students throughout the seven provinces. While comparing the results, readers are suggested to consider mean at 500 and compare the results accordingly.

Figure 2

Comparison of Learning Achievement of Grade 10 Students in NASA 2020



In Nepali language, except Madhes (474) and Karnali (474), the students of other provinces have outperformed those two provinces. From this results, policy makers and practitioners are suggested to implement some interventions for the Nepali language in those two provinces – Karnali and Madhes to improve learning of the Nepali language as per the results of grade 10. While examining the effect size of the difference in mean score, because of selecting the schools in the province only, student performance can vary by 13% which is shown by the effect size (eta-squared) 0.126.

Similarly, in Mathematics, only Gandaki (521), Lumbini (513), and Lumbini (503) are above average where Gandaki is at the top. In Nepali and English also similar results as Mathematics are observed. This is a big question to the system, why other four provinces are consistently below the average (500). There seems that there is a systemic flaw in the input and process of education programs formulation, implementation, and management to remain in all four provinces Sudur Paschim, Karnali, Madhes, and Koshi.

Evidence shows that there is a high and significant difference by province – 13% in science, 8% in Math, 12% in Nepali, and 18% in English (effect size: 0.126 in Science, 0.084 in Math, 0.12 in Nepali, and 0.18 in English).

In contrast, low gender effect in all subjects while analyzing the data by gender with the help of effect size. The effect size was only 0.03 in science, 0.04 in Mathematics, 0.00 in Nepali, and 0.01 in English. This means that there is a very low observed difference in Mean in Science by 3%, 4% in Mathematics, zero in Nepali and 1% in English compared to high effect size by provinces.

While analyzing the effect of language at home, there is low difference (3% in science, 1% in Mathematics, 3% in Nepali and 5% in English). But there seems a high effect size because of the mother's education (11% in science, 12% in Mathematics, 5% in Nepali and 18% in English) and the father's education also effects in a similarly high level (5% in Nepali to 17% in English). These evidences reveal that parents' education is also another determinant of children's learning.

Policy makers are suggested to consider this variation while formulating policy and programs to fulfil such wide gaps among the provinces in learning achievement. Although, gender is important variable from the equity point of view, it is not much influential in the context of Nepal. In the upcoming grades and subjects, these variables will be dig out further to conform the results.

Results of Grade 8 NASA in Mathematics

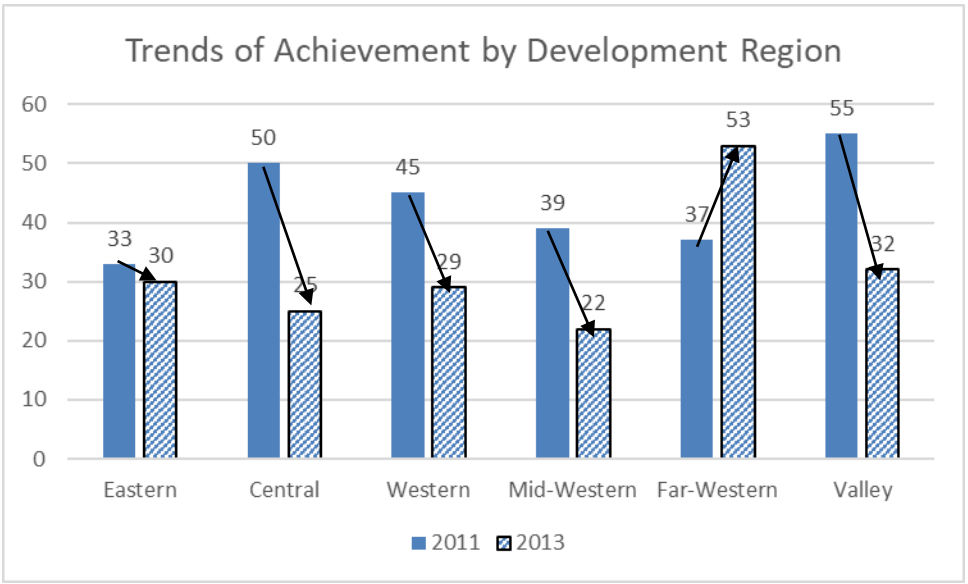
In this section, the effect size of differences of average achievement as well as trends of those achievements are presented. Before 2016, there were only development regions and after 2016, the country was politically divided into 7

provinces by the constitution. So, in this section, both variables are considered to conform to the regional differences by both development regions and provinces.

While analyzing the NASA datasets of Mathematics, as an important subject in grade 8, the first assessment (NASA 2011) and second assessment (NASA 2013) revealed that learning achievement was remarkably decreasing in three years, which is demonstrated in Figure 2.

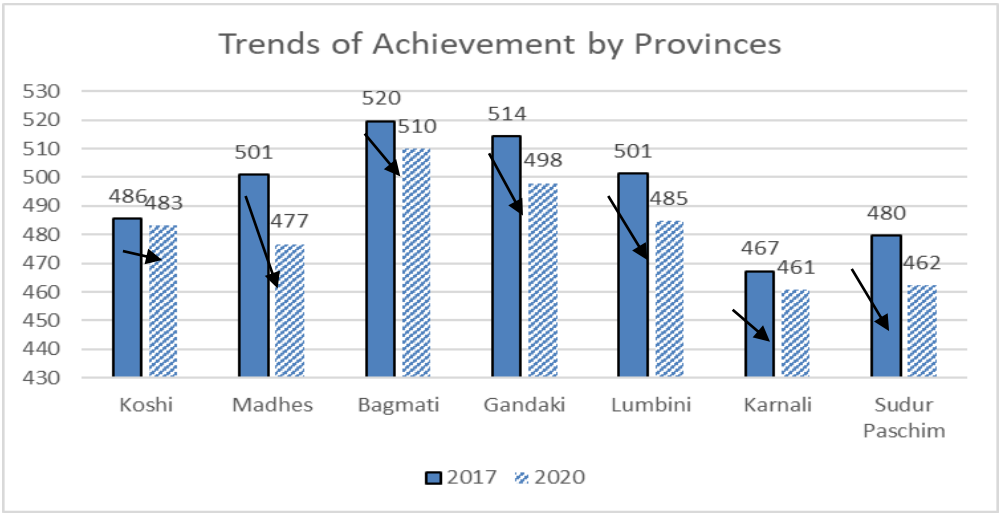
Figure 2

Trends of Learning Achievement by Development Regions



There seemed three types of trends – Central, Western, Mid-western and Kathmandu Valley results were rapidly dropped during those three years, whereas achievement of the Eastern Development region was slightly dropped. In contrast, the achievement of the far-western development region was rapidly increased. Such an increase was observed, not only because of learning was improved but because of the selection of the district. Until 2018, districts were the first level strata of selecting the sample. It does not conform to the increase in learning, but there is a grey area of information.

While exploring the regional differences by provinces, it conforms to the residing dropping trends in learning achievement. Figure 3 depicts how learning dropped over the NASA cycles 2017 and 2020.



The figure 3 reveals that learning achievement has decreased from 2017 to 2020. These results confirm, that there is a continuous slump in learning from 2011 to 2020, no further inquiry is required in this regard. Comparing the achievement scores for the two years, we can observe some changes in the performance of the provinces. In 2017, the highest achievement score was recorded in the Bagmati Province with a score of 519.6, while in 2020, Bagmati Province still had the highest score of 510.0. Bagmati Province appears to maintain its position as the top-performing province in Mathematics. Conversely, in both years, the Karnali Province had the lowest achievement scores among all the provinces, with a score of 467.2 in 2017 and 460.6 in 2020. The other provinces, namelyz Koshi, Madhes, Gandaki, Lumbini, and Sudur Paschim, showed varying levels of achievement scores in both years, but they generally remained relatively stable.

To see, to what extent such declines are affecting learning, the following paragraphs demonstrate the evidences.

Effect size of the difference in average achievement in 2011 was 12% and in 2013 was 25%, higher by development region. After the country was divided into seven provinces. The effect size of the difference of mean from 2017 to 2020 in mathematics by province was increased from 12% in 2017 to 13% in 2020. These data clarify that the difference in learning achievement in grade 8 is high. It is further detailed in the following paragraph.

As the effect size is relatively small to moderate, it suggests that the differences in the achievement scores between the two years are not very substantial.

Results Type of schools

The achievement in mathematics in grade 8 in 2011 was 39% for community schools and 63% in Institutional schools. This achievement in 2013 for community schools was 26% and it was 57 in institutional schools. The effect size of this

difference in student achievement was 18% in 2011 which is widened to 20% in 2020. This result conforms that achievement difference by type of schools is remarkably high,

Achievement difference by gender was high (8%) in 2011 and 2013 but it is dropped to 1% in 2017 and to 2% in 2020. This result conforms that although there seemed the difference in mean achievement score, the effect of gender was dropped every year and reached near zero in 2020. Thus, as an equity indicator, there is no significant effect of gender in learning which is an encouraging result. While looking from the language at home factor, effect size of language was almost zero in 2011 and 2013, which was also near zero (2%) in 2020 in Mathematics. This also confirms that the language is not the factor of affecting learning achievement.

Surprisingly, the most influencing factor in student learning is mother occupation because effect size of mother occupation was 52% in 2011, which was significantly elevated to 82% in 2017 and to 99% in 2020. This result conforms that mother's occupation is the main factor that influences the student learning. Policy makers should identify how mother's occupation is the main factor to influence student learning and prepare to provide opportunity to income generation favorable occupation to mother and provide educational opportunities to female. This trend of very high effect size of mother' education to student learning.

Policy makers are suggested to ensure equal opportunity to quality education for all province children regardless of the type of schools. Furthermore, the difference led by type of schools is serious because of its increasing trend. A societal divide between haves and not haves is a crucial issue that may lead to unstable society causes social conflicts and inequality in long run because type of schools as a single variable is sufficient to explain the differences. A clear long-term policy to minimize the divide of people who admit their children in institutional schools and community schools should be implemented as soon as possible.

Results of Nepali Language in Grade 8

The meta-analysis of grade 8 results from NASA 2011 to 2020 revealed that there is no big difference in student performance by any variable except type of schools and ecological belt. The difference in average achievement by ecological belts – mountain, hill and terai is 11% (partial eta squared = 0.11) in NASA 2011 and 28% (partial eta squared = 0.278) in NASA 2013. It was quite similar to the development regions where effect size due to development region in 2011 was 12% (partial eta square = 0.12) in 2011 and 28% (partial eta square = 0.275) in 2013. This difference also was quite close due to the provinces, 14% (partial eta squared = 0.139). These all facts conform that there is a regional difference in average achievement, and it follows similar results like in grade 10 all subjects and grade 8 Mathematics.

The effect size of difference in average achievement in Nepali is quite low, 10% in NASA 2011, 27% in NASA 2013, and 4% in NASA 2020. These indicators revealed that there is not much variation in learning achievement in Nepali language because of the type of schools as well as because it was 99% in Mathematics, extremely threatening.

While analyzing the results by students' gender, effect size was 9% in NASA 2011 and zero in NASA 2013 and 2020. Thus, gender difference is not an issue in Nepali language, although girls are slightly performing better than girls in this subject.

Regarding mother tongue, there is almost zero effect in learning achievement with this variable. There was 1% effect of home language in 2011 and 6% effect in 2013, This effect was 2% in 2020, showing it ranging from 1% - 6%, which very low compared to other effective variable. Still, home language should not be undermined that language of instruction plays crucial role in students' learning process inside the classroom.

The effect of different variables is acting similar to Mathematics, the effect sizes of those variables are not mentioned in this article at this time.

Results of Grade 5 NASA

In this section, I do not deal much with the above explained variables, however, the significant improvements and differences are observed. In NASA 2012, ecological belt (mountain, hill and terai) was a moderately effective variable in student achievement. Data revealed that ecological belt effect size was 14% (partial eta square = 0.139) in 2012 and it dropped to 5% (partial eta square = 0.045) in NASA 2015. It shows a decreasing trend in effect of ecological belt in Nepali language over three years. It was further low while comparing the effect size of difference in achievement over the seven provinces, only 3% (partial eta square = 0.034), a negligible one. It was even quite moderate effect of type of schools – community and institutional. Effect size of difference in mean achievement of community and institutional schools was 15% in 2012 and 12% in 2015. Similar effect is observed in NASA 2018 as well, 13%. Thus, type of schools is not a very crucial issue in Nepali language where as it was unbelievably high in Mathematics in grade 8 (99%).

There is no gender issue seen in the Nepali language achievement, girls outperformed in this language and effect of difference in mean achievement was only 1% in 2012, zero in 2015, and 2% in 2018. Thus, this result again conforms that gender difference does not exist in Nepali language.

I would like to add one more variable in Language issue, i.e., the caste ethnicity. It can be assumed that there may exist difference in mean achievement by caste/ethnicity. Although there exists mean difference in learning achievement by

caste/ethnicity, the effect size is very low. The effect size was 7% in NASA 2012 (partial eta square = 0.071) and 3% in NASA 2015 (partial eta square = 0.029). It was almost zero in NASA 2018 (partial eta square = 0.004). It concludes that caste/ethnicity is not the main issue of student learning although its effect varies from 0 to 7% by year, compared to other variables, it is very low.

Reflection

NASA datasets are rich for educational policy makers. They should be available to universities and researchers who are willing to analyze and get insights through a certain bond. While prioritizing the educational policies, this rich dataset of NASA should be used wisely. Developing any policy based on a single assessment cycle is a risky game because technological advancement has changed the social scenario and educational practices rapidly. Therefore, meta-analysis of the dataset is essential to inform a firm policy.

The present analysis conforms to some unavoidable issues of provincial differences, ecological belt-wise differences and mother education differences. Upcoming policies will be reliable when they adequately minimize the negative effect of type of schools and increase the positive effect of mother education. Furthermore, policymakers should not be unaware of there is a gender parity in learning to some extent and there is a low effect of caste/ethnicity. However, the difference comes from overall socio-economic variables, leading to selecting type of school, opportunity to get quality by provinces. The funding modalities in the Education Sector should be reviewed from the unconditional grants and equalizer grants to avoid the blanket approach to funding mechanism being more effective. The quality of education is not better than the quality of the infrastructure of schools, learning environments and quality of teachers. My reflection during the working period in ERO for 13 years, resources are the crucial indicators associated to improvement in above mentioned differences.

References

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

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

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

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

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

EDSC. (2001). *National Assessment of Grade 3 Students.* Kathmandu: Educational and Developmental Service Centre.
ERO. (2013). SPSS data of Grade 8 Mathematics in NASA 2011.
ERO. (2013). SPSS data of Grade 8 Nepali in NASA 2011.
ERO. (2013). SPSS data of Grade 8 Social Studies in NASA 2011.
ERO (2014). SPSS data of Grade 5 English in NASA 2012.
ERO. (2014). SPSS data of Grade 5 Mathematics in NASA 2012.
ERO. (2014). SPSS data of Grade 5 Nepali in NASA 2012.
ERO. (2014). SPSS data of Grade 5 Nepali in NASA 2012.
ERO. (2014). SPSS data of Grade 8 Mathematics in NASA 2013.
ERO. (2014). SPSS data of Grade 8 Nepali in NASA 2013.
ERO. (2014). SPSS data of Grade 8 Science in NASA 2013.
ERO. (2015). SPSS data of Grade 5 English in NASA 2015.
ERO. (2015). SPSS data of Grade 5 Mathematics in NASA 2015.
ERO. (2015). SPSS data of Grade 5 Nepali in NASA 2015.
ERO. (2015). SPSS data of Grade 5 Nepali in NASA 2015.
ERO. (2018). SPSS data of Grade 8 Mathematics in NASA 2017.
ERO. (2018). SPSS data of Grade 8 Nepali in NASA 2017.
ERO. (2018). SPSS data of Grade 8 Science in NASA 2017.
ERO. (2019). SPSS data of Grade 5 Mathematics in NASA 2018.
ERO. (2019). SPSS data of Grade 5 Nepali in NASA 2018.
ERO. (2020). SPSS data of Grade 10 English in NASA 2019.
ERO. (2020). SPSS data of Grade 10 Mathematics in NASA 2019.
ERO. (2020). SPSS data of Grade 10 Nepali in NASA 2019.
ERO. (2020). SPSS data of Grade 10 Science in NASA 2019.
ERO. (2022). SPSS data of Grade 8 English in NASA 2020.
ERO. (2022). SPSS data of Grade 8 Mathematics in NASA 2020.
ERO. (2022). SPSS data of Grade 8 Nepali in NASA 2020.
ERO. (2022). SPSS data of Grade 8 Science in NASA 2020.
Greaney, V. & Kellaghan, T. (2008a). *Implementing a national assessment of educational achievement*, Vol 3, the World Bank.
Greaney, V. & Kellaghan, T. (2008b). National assessments of educational achievement: Assessing national achievement levels in education, Vol 1, the World Bank.
Metsämuuronen, J. (2002). *Small book of test theory and test construction*. Helsinki: International Methelp Oy
MOE. (2009). The school sector reform plan, 2009-2015. Kathmandu: The government of Nepal, Ministry of Education MOE.

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

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

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

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

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

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

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