NEJEA
ERO

Volume 1 • Issue 1
2016 January

Nepalese Journal of EDUCATIONAL ASSESSMENT

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

ISSN: 2467-9488 (Print)

Nepalese Journal of Educational Assessment

A Biannual Publication of the Education Review Office, Sanothimi, Bhaktapur

Volume 1 • Issue 1
2016 January

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

Year of Publication:

January 2016 (Poush 2072 BS)

Nepalese Journal of Educational Assessment, 2016, Volume 1, Issue 1

Printed at: Dharma Printing Press, Kathmandu

ISSN: 2467-9488 (Print)

Nepalese Journal of Educational Assessment
Education Review Office
Sanothimi, Bhaktapur, Nepal
Phone: +977-1-6633556, +977-1-6633557, Fax: +977-1-6633556
Email: nejea.ero@gmail.com; Website: www.ero.gov.np

NEJEA
ERO

Nepalese Journal of Educational Assessment
Volume 1, Issue 1, January 2016

EDITORIAL

Rise of Educational Assessments

Various forms of assessment have evolved since the institutionalized beginning of formal schooling. Examinations and tests, usually coming at certain interval or end of the course, are the dominant means for the assessment of student learning with their inclination more towards selection, promotion and certification purposes. In addition to the examination centred assessment of learning, with the concept of assessment for and as learning, classroom based assessment evolved as an approach to continuous assessment for motivating students in learning and to provide feedback to the teacher for improving the instructional process. Though both forms have their own significance in informing individual students, teachers and parents on what one has acquired and in serving the purpose of selection, promotion and certification; they are considered inadequate to inform the policy mechanics and national system on how well students have learnt, what they can do and what the gaps are there within the system.

To satisfy these concerns, educational and psychological sciences together with the policy makers have opted for large scale assessment called National Assessment of Student Achievement by means of standardized tests, which has been adopted in various countries around the world considering it a reliable means to provide systemic policy feedback for its improvement.

Along with the rise of assessment and testing procedures for student learning achievement, assessment of organizational performance, called performance audit, also has evolved to assess the organizational performance of educational institutions against their given mandates, goals to achieve and target to meet with a view to hold them accountable to student achievement. Being cognizant of these assessments in the endeavors of quality improvement, Nepal also has been adopting both of them to hold the system accountable towards improving the quality and service delivery in education.

With a view to disseminate and extend the knowledge and technology developed in the field of educational assessment, especially student assessment and performance audit of educational institutions, this journal is conceived to create a forum for building knowledge, spreading wisdom and bringing forth the different practices in this field. It also aims to create an avenue for sharing innovative practices and technologies in the field of educational assessment by means of scholarly discourses.

The present issue revolves around the theme of educational assessment– student assessment and performance audit. The contents on it delve into the concepts, practices, issues and challenges regarding these assessments illuminating relevant theoretical perspectives. They also look into the results of these assessments in Nepalese context along with the main concerns behind them.

Readers in the field of education in general and educational assessment in particular will be benefitted from this issue. Constructive suggestions and feedbacks from readers are expected to further improve its forthcoming issues. We also invite scholarly articles from all authors to support our endeavor to build and share knowledge, wisdom and innovative ideas in the field of educational assessment.

January 2016

Nepalese Journal of Educational Assessment

• Volume 1 • Issue 1 • 2016 January

Table of Contents

Reviewing the Practice of National Assessment of Student Achievement in Nepal

Lekha Nath Poudel, PhD,

Joint Secretary, Education Review Office, Ministry of Education

Abstract

Along with classroom based assessment and public examination, Nepal has realised the need for large-scale national level assessment of student achievement to monitor properly the quality of education. Government instituted Education Review Office to conduct large-scale student assessment of school level students and to carry out performance audit of schools and other institutions under the Ministry of Education on a regular basis. The aim of the large-scale assessment conducted by Education Review Office called National Assessment of student Assessment (NASA), is to provide policy feedback at system level to enhance the quality of and equity in school education in Nepal. Drawing upon the literature as well as reviewing the practices of Nepal, this article claims that the enabling environment for the national assessment is still weak and there is a need for improving the overall system and practice of NASA in order to improve the quality of national assessment.

Keywords: classroom based assessment, public examination, large-scale assessment, national assessment

Introduction

Student assessment is an integral part of an education system. The objectives of student assessment are: (1) to provide feedback for classroom teaching and improve student learning; (2) to certify the grade and qualification of students; and (3) to monitor and evaluate the quality of education system and improve it. In order to achieve these objectives, three forms of assessment have widely been practiced. We use Classroom Based Assessment (CBA) for improving student learning by providing feedback to classroom teaching learning whereas public examinations certify grades and qualifications of students. Similarly, a large-scale assessment, the National Assessment of Student Achievement (NASA) for instance, aims at improving education system by providing feedback for the purpose of policy formulation as well as implementation. Nepal has been conducting these three types of assessments at school level. Among these three types of assessments, this article focuses mainly on large-scale national assessment of student achievement.

The main objective of this article is to review the Nepalese practice on national assessment of student achievement. The analysis of the practice of NASA in Nepal shows that there is a need

of further strengthening it in order to improve its quality.

This article begins with the introduction. The second section of this article presents conceptual discussions about the large-scale assessment of student achievement. The third section includes a brief review of the practice of national assessment in Nepal. The fourth section identifies the strengths and weaknesses of the practice of national assessment of student achievement in Nepal by reviewing the present practices. Before concluding the article, it suggests some ideas for improving its quality.

Method and Framework for Analysis

I have developed this article by reviewing the relevant literatures, studies/reports and international practices together with the study of previous practice of NASA in Nepal. Comparing the Nepalese practice of NASA with the standard international practices has led towards suggesting future directions for NASA. During the analysis, I have also included some primary data obtained from teachers, experts and technical personnel from ERO and DEOs.

The strengths and weaknesses of the present Nepalese practice of NASA have been identified by reviewing national assessment using the framework suggested by Clarke (2012). The framework for reviewing the overall quality of assessment includes enabling environment, system alignment, and assessment quality including the use of assessment results. The quality of NASA has also been analysed with the help of the five criteria for high quality assessment suggested by Darling-Hammond et al. (2013). These five criteria are: (i) assessment of high-order cognitive skills; (ii) high-fidelity assessment of critical abilities; (iii) standards that are internationally benchmarked; (iv) assessments that are internationally sensitive and educationally valuable; and (v) assessments that are valid, reliable and fair.

Concept, Origin and Purpose of National Assessment

An education system takes assessment as a means to check whether the students have achieved the expected competencies set in the curriculum. Student assessment, by identifying strengths and weaknesses of educational delivery, provides an opportunity for reshaping the delivery system and improving student learning. Assessment includes "the processes of gathering and evaluating information on what students know, understand, and can do in order to make an informed decision about next steps in the educational process" (Clarke, 2012, p 1). If we consider assessment as a system, we define assessment "is a group of policies, structures, practices, and tools for generating and using information on student learning and achievement" (ibid.). In this way, student assessment is a process of obtaining and analysing information about students' learning and providing feedback to the students as well as other stakeholders including teachers, parents, educational managers, program developers, and policy makers for further improvement in students' learning.

The overall objective of student assessment is to enhance student learning by improving learning conditions and processes. To achieve the overall objective of assessment, student assessment focuses particularly on three purposes: (i) to provide feedback to classroom teaching learning; (ii) to certify grades and qualifications of students; and (iii) to improve education system. Several education systems in the world have been adopting these three types of assessments in order to achieve the above-mentioned three objectives. These three types of assessments are: (i) classroom based assessment; (ii) public examinations; and (iii) national assessment of learning achievement.

In order to provide feedback to policies and programs, periodic assessment of student learning achievement, generally known as large-scale assessment or system level assessment, is conducted at national/sub-national as well as regional/international level, which is carried out by taking a sample of students or by conducting census of the students of certain grade or age. It generally assesses the quality of and equity in education system and provides bases for the reform in policies and practices. The purpose of large-scale assessment of student achievement is not only to inform about the quality of and equity in education, but also to improve accountability of the system. There are different practices across the globe regarding the agency responsible for conducting large-scale assessments of student achievement. For example, in some countries, an independent or semi-autonomous agency has been established for this purpose, while in some countries government agency has been working for the same, and there are even the practices of assigning the task of assessment to an external agency.

There are close relations between assessment and student learning. We can interpret such relations in three different ways: "Assessment of, for, and as student learning, rather than as a separate disjointed element" of the educational practice (Clarke, 2012. p 5). When we see the characteristics of various types of assessment, public examinations mostly serve as an assessment of student learning. On the other hand, both the classroom based assessment and large-scale system level assessment are not limited to the assessment of student learning only; they further provide feedback to learning process and practices at different levels and work as an assessment for learning. Besides, classroom based assessments also work as a part of student learning process, that is, assessment as student learning. The relation between assessment and student learning justifies a need for effective system of assessment for a successful education system. In this connection, a national assessment program differs from the existing tests and examinations conducted for promotion, selection, and certification. It is not merely a testing program. Rather, it is an information gathering system to facilitate the making of cost-effective decisions (Nwana, 1996, p. 24).

The focus of national assessment of student achievement is to assess and monitor the quality of and equity in education system rather than to assess or monitor the learning of an individual student. National assessments are "systematic, regular measures of learning achievement in a country that are designed to assist policymaking" (Murphy, Greaney, Marlaine, & Rojas, 1996,

p. 2). This definition clarifies that national assessment is not the assessment of an individual student; it is the assessment of education system aiming at providing feedback for policymaking and programme implementation. A national assessment is "a survey of schools and students that is designed to provide evidence about the levels of student achievement in identified curriculum areas for a whole education system or for a clearly defined part of an education system" (IIEP, 2008, p. 1). Instead of assessing the quality of education from input perspective – e.g. considering the factors such as student participation rates, physical facilities, curricular materials, and teacher training – it assesses from outcome perspectives, such as the knowledge and skills acquired from schooling, which is 'a shift of the perspective' of assessment (Greaney & Kellaghan, 2008). However, along with providing information on outcomes, it also gives some information on inputs as well as delivery process. The information provided to policy makers and education managers about the outcomes through national assessment demonstrates the evidence about strength/ success or weakness/failure of education system, which indeed works as a basis for corrective measures. It also provides information to stakeholders including teachers, parents and the public about the student achievement and factors affecting achievement of students from which they could redefine their roles.

A large-scale national assessment of student achievement assesses the health of the entire education system and provides policy feedback to the government as well as teachers, parents and students. This kind of assessment does not consider individual student and teacher as the unit of analysis, but it compares various sub-groups based on their achievement scores. We can interpret test results using certain norms as well as some pre-determined criteria. If we want to interpret test results with certain criteria, we should define clearly the criteria while designing the test. Overall quality of assessment depends not only on the use of standardized tools for testing, but also on the administration of test, generation of data and analysis and interpretation of test results. Moreover, the usefulness of assessment also depends upon the mechanism set for the continued improvement based on the feedback received from the assessment.

A national assessment of student achievement is neither directly related to the improvement of teaching-learning as in the case of classroom based assessment, nor it certifies student qualification as done by examination in general. However, it provides information for the feedback to the overall system of education so that the quality of and equity in education system can be improved. As national assessment of student achievement, it assesses the education system and provides feedback; and the feedback ultimately links with student learning. Such an assessment generally answers the following questions (Greaney & Kellaghan, 2008):

- How well are students learning?
- Is there evidence of particular strengths and weaknesses in students' knowledge and skills?

- Do certain sub-groups of students perform poorly?
- What factors are associated with student achievement?
- Do the achievements of students change over time?

The answers to the above questions are based on the standard defined by the national curriculum, in which "national assessments provide rich information about learning outcomes according to nationally defined standards" (Sui-chu HO, 2013, p. 6). Therefore, national assessment is developed and carried out to assess the overall performance level against the curricular objectives or standards. Policy makers and practitioners may utilize the information about the overall achievement levels and contributing factors for learning – which becomes the feedback towards reform in policy and practices. Therefore, "well-designed national assessments can help countries make informed decisions about interventions to improve educational quality" (Murphy et al., 1996, p. 2).

Major focus of assessment is to make education systems accountable towards the leaning of students. However, test based accountability may not always contribute for student learning as "system-level accountability in isolation is generally unsuccessful" (Griffin, Care, Francis, & Claire, 2014, p. 73) and "assessment regimes in systems of high accountability do not necessarily improve the student learning when considered at an aggregate level" (ibid.). We should keep two points in mind while developing a system of national assessment: first, assessment should be linked with educational delivery; second, assessment should be used as a complementary activity of education delivery. While planning assessment, attention should be paid for balanced use of assessment system with other process of educational delivery so that there will be no over emphasis on testing and assessment and no negative effects on student learning.

Overview of Nepalese Practice on Students' Assessment

In Nepal, practices of public examination formally began alone with the establishment of SLC Board in 1934. National Education System Plan (NESP) (MOE, 1971) introduced classroom-based assessment, known as internal assessment. However, the assessment system did not continue after the implementation of revised curriculum of school level in the early 1980s as it dropped the system of internal assessment from school education. Government of Nepal reintroduced CAS (Continuous Assessment System) in primary grades of school education as a classroom based assessment system in the late 1980s. Since the middle of the 1990s, Nepal began large-scale national assessments of student achievement in some grades but not on regular basis. The government of Nepal, Ministry of Education/BPEP/DOE commissioned a number of assessments of student achievement from 1995 to 2011 in various grades. The student assessments conducted from 1995 to 2011 are the ones conducted by BPEP (1995, 1997), EDSC (1997), BPEP (1998), EDSC (1999), CERID (1999), EDSC (2001, 2003), CERSOD (2001), EDSC (2008),

Fulbright (2008), and EDSC (2011) (see: ERO, 2013; EDSC, 2011). As these assessments were commissioned by the MOE/BPEP/DOE and carried out mostly by external agencies, there was a lack of institutional setting and a regular system of national assessment. The tests were based on traditional testing theory with inadequate standardization of items and testing process in relatively small sample sizes. However, most of these assessments identified teacher, school and student related factors influencing student learning.

As envisioned in the School Sector Reform Plan (SSRP, 2009-2016) (MOE, 2009), Education Review Office (ERO) instituted in 2010. One of the major functions of ERO is to conduct large-scale system level assessment on a regular basis. Since its establishment, ERO has been conducting National Assessment of Student Achievement (NASA) for various grades (ERO, 2015a). It has completed two national assessments of student achievement for grade eight in 2011 and 2013 (see: ERO, 2013 and 2015) and one national assessment of grade five and three each in 2012 (see: ERO, 2015b). The second cycle of national assessment of grade three and five students is also in the process of accomplishment as the data analysis is underway. Each of the assessments conducted so far was based on large sample of schools and students representing various ecological zones and development regions of Nepal.

ERO has been following the uniform process and cycle for both rounds of National Assessment of Student Achievement of these grades. The general cycle used to carry out each of the NASAs is as follows (see, ERO, 2013, 2015): (1) assignment for NASA from the Ministry of Education in the form of the approval of budget and programme, (2) establishment of planning and execution team at the ERO – the NASA unit; (3) formation of subject committee and selection of item writers, (4) item writing, pre-testing and analysis of the results, (5) finalizing the test items based on the analysis of pre-test results, (6) sample selection, (7) test administration, (8) scoring, data entry, data verification and cleaning, (9) analysis of the data and equating the test scores using IRT modelling, (10) report preparation and dissemination of the results.

The above assessment procedure shows that a systematic and standard process has been followed for NASA. However, if we analyze each of the steps in the cycle, further improvement is felt necessary for developing an effective system of national assessment. For example, NASA unit at ERO has insufficient technical human resources as there is no any provision for the position of specific technical staffs, and the general staff will not retain long in a single office. Lack of adequate national capacity in item writing, item analysis, data analysis, particularly in IRT modelling is a challenge in developing effective system of national assessment in Nepal. As ERO has still been working on ad hoc basis, an institutional arrangement with proper legal mandate is yet to be established. The succeeding section discusses in detail about the national assessment system and the quality of assessment practices.

A Review of the Quality of NASA in Nepal

One way of understanding the quality of assessment system is the balanced practice of three types of assessment – classroom based assessment, examination, and large-scale assessment of student achievement (see, Clarke, 2012). Next, the overall quality of large-scale assessment of student achievement depends upon three quality drivers–enabling context, system alignment, and assessment quality (ibid.). Similarly, assessment quality can be assessed using five criteria for high quality assessment identified by Darling-Hammond et al. (2013). In this section, I will present a brief review of the quality of NASA conducted in Nepal based on above-mentioned three areas of quality concern.

Practice of three types of assessment

As mentioned in earlier section, the main objective of Classroom Based Assessment (CBA) is to improve learning or performance of students by improving teaching learning process. Whereas public examinations serve the purpose of selection and promotion of students for the next grade or level of education and certification of learning, which generally do not provide feedback to classroom teaching learning or the education system. Similarly, the main objective of periodic assessment of student achievement is to provide policy feedback to the education system and to generate evidence based information for monitoring the progress over the period. With regard to classroom-based assessment, there is a policy provision to implement Continuous Assessment System (CAS) in primary grades of Nepalese schools, which now has been extended to grade seven. However, in practice, some schools have been using CAS as a non-testing device while others have been using a combination of non-testing CAS, trimester and final examination, and some schools have only been using various types of tests for deciding students' grades and promotion. One of the weaknesses found in the practices of CAS in Nepalese schools is the detachment of assessment from classroom teaching learning; as a result, the use of assessment for formative purpose is minimal. In this regard, "assessment results have hardly been used in teaching-learning process; rather it has been a ritual of filling the forms in many schools" (Poudel, et al., 2015).

In Nepal, public examinations have been conducted at national level for grade 10, 11 and 12 – which mostly assess the lower order skills using the test items that are mostly not standardized. The recent practice of grade 8 district level examination is just a ritual in most of the cases. Regarding large-scale national assessment, some form of systematic and regular practice has begun since the establishment of ERO in 2010.

Three quality drivers and the NASA

Based on the framework proposed by Clarke (2012) about quality drivers, namely, enabling context, system alignment, and assessment quality, I briefly review the practice of national

assessment of student achievement in Nepal. We may assess the enabling context for NASA by examining: policies, leadership and public engagement, funding, institutional arrangement, and human resources. Similarly, system alignment can be assessed through learning/quality goals, curriculum, pre-service and in-service teacher training; and assessment quality includes quality design, instruments and process of administration and analysis, and effective use of results. I will discuss briefly the status and challenges for ensuring quality in the national assessment system in Nepal. Then I will review further the assessment quality of NASA conducted by ERO.

Existence of 'enabling context' for NASA is a pre-requisite for effective system of assessment. Although ERO has been established in 2010 and conducting NASA regularly, the institutional arrangement and policy provisions are ad hoc due to the lack of any legal provision. However, the Ministry of Education has formed a Steering Committee to facilitate the functions of ERO and to provide policy inputs. Similarly, to provide technical support to ERO, the Ministry of Education has formed a Technical Committee. As ERO has been established on temporary basis, there is a lack of functional autonomy including daily functional decisions, budgetary provision, procurement as well as office and personnel management. Even though the programme budgets are sanctioned and authorized by the secretary of education to the joint secretary of ERO, due to the lack of separate office management budget and account section at the ERO, unnecessary delay, burden and obstacles have been experienced even in the daily functions.

ERO has been working with the temporary arrangement of human resources, and some of them have some technical capacity required to conduct NASA and performance audit of schools and institutions. In the initial years of its establishment, the service of an international consultant was mobilised for student assessment. The provision of international consultant also contributed for capacity development of ERO technical staffs. Now, some of the trained technical staffs have already been transferred from ERO and there is no any provision of international consultants or long-term national consultants in ERO. In such a situation, on the one hand, ERO has been working with the limited human resources and at the same time, the decisions on staff transfer and redeployment have been made without adequately considering the nature of work to be performed and the capacity and interest of deputed staffs.

As discussed above, in Nepal the enabling context for large-scale assessment is not strong as required. Conducting NASA in poor enabling conditions is a challenging as well as risky move in many ways. For instance, there will be a lack of certainty in ensuring quality and sustainability of assessment; and there will be difficulties to use the assessment results properly in improving education system and practices.

National assessment system should be aligned with the other systems of education, including learning goals, standards, curriculum, as well as pre- and in-service teacher training opportunities (Fuhrman and Elmore, 1994; Smith and O'Day, 1991, quoted in Greaney & Kellaghan, 2008). Conducting assessment without aligning it with education system is an

ineffective way of engagement in assessment activities, as alignment is necessary not only for the use of assessment results but also for making assessment process smooth and ensuring support for and participation of relevant stakeholders. Aligning assessment activities with the rest of the education system also encourages using the information received through assessment in improving the quality of education by creating synergies between assessment and improvement in education system (Greaney & Kellaghan, 2008). While aligning assessment and other systems of education, considerations should be given to cover all learning domains including various levels of knowledge, skills and competencies as envisioned by the curriculum; to cover the whole population of students; and to make the results consistent and useful to stakeholders, learning goals and priorities (ibid.).

Since the first NASA, test items for the assessment have been based on the curricular objectives and competencies. To ensure test items and other tools are based on the curriculum, curricular objectives and competencies were analysed by a team of curriculum experts, teacher trainers and teachers. Various parts as well as domains of the curricular objectives and competencies were also tested using respective questions. Although participation of the faculties of universities responsible for teacher preparation, teacher training agencies and subject teachers were ensured in the initial drafting of the assessment questions as well as in pre-testing and revising the assessment tools, it was difficult to ensure one-to-one correspondence of assessment tools and process of participation. Accordingly, pre-service and in-service teacher training were not visible in the assessment process. Such alignments were not based on pre-determined assessment frameworks; as a result, there might have been some inconsistencies in the process.

The quality of assessment depends upon the design of assessment, process of conducting test and analysis of results as well as the use of assessment results. As part of the use of assessment results, the results were disseminated and discussed with a wider audience of stakeholders including the system of the Ministry of Education, various forms of media, regional and district levels. This year, it has been planned to disseminate school report card of NASA 2013 and the summary results and findings of all the four assessments to all the participating schools. During the dissemination at the Resource Centre level, each school will be facilitated to use the assessment results in improving teaching learning. At the same time, implementation of learning improvement plan developed based on assessment results will be piloted, so that Department of Education will mainstream the developed frameworks and process of learning improvement plan for schools. At central level, the Ministry of Education prepares and updates the action plan based on the NASA results and the implementation of the action plan is monitored during the regular monitoring from the Ministry of Education, and during the Joint Annual Review of SSRP organized between the government and development partners.

Quality of assessment refers to the quality in assessment design including instruments, process and analysis. It also refers to the effective use of assessment results. The design of national assessments was based on the sample of schools and students, which began with the

sampling methods and test design. The sampling design for each of the four NASAs conducted by the ERO was based on stratified random sampling with relatively large sample size. In order to represent three ecological zones and five development regions, stratified random sampling method was used. Due to the variations across ecological zones and regions and the requirement of disaggregated information such as gender, ethnicity, language, school types (public/private) and school location (rural/urban), relatively large sample size was taken. Quality of test items were ensured by constructing them based on the curriculum and by pilot testing and selecting the items with appropriate level of reliability. Similarly, item difficulty of each piloted item was calculated, and the items having very low and very high difficulty level were discarded.

All the processes under test administration and data generation were carried out by a selected consultancy company with a close monitoring of the ERO. The consulting company collected all the data (answer sheets and background questionnaires) with the support from respective District Education Offices and Schools. The consulting company marked students' answer sheets, tabulated the data (note that NASA 2011 and 2012 were done manually, but NASA 2013 and 2014 were done using OMR), and then provided the data to ERO in electronic form.

With the participation of external experts, data analysis was carried out by the ERO using One Parameter Logistic Model (OPLM) of Item Response Theory (IRT). IRT also used to calibrate items and to equate the scores of the various sets of items, international assessments like PIRLS (Progress in International Literacy Study) and TIMSS (Trends in International Mathematics and Science Study) and other assessments to make the results of these assessments comparable. During report writing and content editing, the inputs from external experts were included and the results were discussed to minimize the bias in the interpretation of results.

Five quality criteria of assessment and the NASA

In this sub-section, I discuss the assessment quality of NASA in detail following the five criteria for high quality assessment identified by Darling-Hammond et al. (2013). Darling-Hammond et al. (2013) identified five criteria for high quality assessment: (i) Assessment of high-order cognitive skills; (ii) high-fidelity assessment of critical abilities; (iii) standards that are internationally benchmarked; (iv) assessments that are internationally sensitive and educationally valuable; and (v) assessments that are valid, reliable and fair. When we compare NASA conducted so far, some similarities and some gaps can be observed. NASA used Bloom's taxonomy of cognitive domain while developing assessment tools and analysing the results from which it tried to maintain a balance among four levels of cognitive domain – namely knowledge, comprehension, application and higher ability. During the construction of items, although curricular objectives were analysed briefly, still there is a need for depth analysis of curricular competencies. At the same time, further discussion is needed on the appropriate categorization of the levels of learning among various domains. For example, the appropriateness of the use of Bloom's taxonomy (traditional or revised form) (see, Anderson, Krathwohl, et al., 2001), Solo taxonomy (see Biggs

& Collis, 1982), the four levels of depth of knowledge as proposed by Webb (2002) or any other classifications need to be discussed further before deciding the levels of learning. Assessments are based on paper-pencil tests in which some of the skills and competencies may not be possible to measure directly through it. They may need alternative tools, instead. Therefore, there is a need to identify and use varieties of assessment tools in order to assess the relevant skills and competencies. In order to make assessment results comparable, some of the items from PISA and TIMSS were taken and calibrated for grade eight assessments and some of the items from PIRLS and TIMSS were calibrated for grade three and five in respective subjects. However, without comparing contents and expected performance, maintaining international standards may not be possible using only some of the items. On the one hand, NASA test items were based on curriculum and therefore the assessments were instructionally valuable; on the other hand, the language used in the items and the examples presented in the test have not been scrutinized deeply; and therefore there might be some gaps. Items were based on the curriculum and pre-tested, from which biased and less reliable items were discarded to make the test more reliable, valid and fair.

From the above discussion based on the criteria of assessment as proposed by Darling-Hammond et al. (2013), assessments conducted by ERO have met some of the criteria; but some gaps were noticed. Moreover, in order to maintain the quality of NASA, there is a need for deeper analysis of curriculum, preparation of detailed framework for assessment, identification and use of alternative assessment tools and techniques rather than paper-pencil test, developing explicit criteria for constructing the tools and analysing the assessment results. The following preparation should be ensured before administering the national assessment in which these points need to be included in the assessment framework to be developed before administrating the test (Greaney, 1996, p 62).

- A detailed description of the constructs and the contents to be measured;
- An analysis of what each instrument or procedure purports to assess;
- A precise specification on how the instrument is to be administered, scored, and interpreted in the proposed national assessment;
- A set of supportive quantitative and qualitative evidence to justify the use of the particular instrument or procedure;
- An overall assessment of the validity of the use and, in particular, interpretation of the instrument or procedure.

Besides, the process of standardization of test items, including defining item parameters, ensuring appropriate difficulty level and discriminatory capacity of items should be described

during the planning phase of the assessment, and should be included in the developed assessment framework.

The above review of NASA policy and practice shows that the enabling context for NASA is still weak and therefore it is still in the 'emerging stage' considering the four-stage model of analysing assessment policies and practices in any education system. These four stages of development of assessment in general or NASA in particular are (see, Clarke, 2012): latent (almost absent or deviated from the right track); emerging (meeting minimum standard); established (acceptable minimum standard), and advanced (best practice).

When we see system alignment, it is still in the borderline between 'emerging' and 'established' stage. If we develop a framework for assessment, the alignment may arrive at the 'established' stage. Therefore, the quality criterion of the NASA suits the established stage. One of the main causes of overall stage of NASA remaining in the borderline between emerging and established stages could be the weak enabling context for NASA. However, for the high quality NASA, each of the three quality drivers has to be developed further to reach the 'established' level and moved ahead towards the 'advanced' stage.

The Way Forward for Improving the Quality of NASA

The above discussion indicates that there is a need for further strengthening the practice of national assessment of student achievement – a system level assessment – in Nepal. For strengthening NASA practice, we should move the overall national assessment system from 'emerging' stage to 'established' stage and continue the development towards the 'advanced' stage as suggested by Clarke (2012). The following are the major areas of intervention to improve the practice of NASA in Nepal:

- Developing and promulgating legislation for national assessment.
- Institutionalizing ERO (national authority for assessment) as a dedicated agency for national assessment with sufficient functional autonomy and human resources.
- Developing national and institutional capacity in test development, result analysis and report preparation, particularly strengthening ERO with technical capacity and technological infrastructures.
- Developing a pool of experts in assessment and psychometrics within the national system.
- Developing a clearly defined national assessment framework comparable to international standard.

- Defining criteria and standards for national assessment articulating the curricular competencies.

- Preparing national system to participate in international/regional level assessment.

- Establishing a mechanism for continued improvement in policy and practice towards enhancing the quality of and equity in education, based on the findings of national assessment of student achievement.

Some areas of suggested interventions should be initiated and coordinated by the Ministry of Education. Others should be worked out by ERO itself, and the rest should be planned and worked out together by the Ministry of Education and ERO. While planning and developing national assessment system, there is a need for considering quality of education for all to which national assessment could supplement for other system of educational process and help to increase the accountability of education system.

Besides, there is a need for improving the practices of classroom based assessments and public examination system, focusing respectively their roles as an assessment for learning and as an assessment of learning.

Conclusions

Assessment is a major tool to know "whether an education system is producing the desired outcomes for students, the economy, and society. Without effective assessment, it is impossible to know whether students are learning and whether reforms are working in the intended ways" (Clarke, 2012, p. 18). The information received about the achievement of students provides opportunity to improve student learning. Moreover, the feedback received from the assessment results encourages all relevant stakeholders to identify and reshape their roles for the enhancement of student learning. In this way, one of the major roles of assessment is to make education system accountable towards the leaning of students.

Three types of assessment are in practice in various education systems with three different focuses. Classroom Based Assessment (CBA) conducted by classroom teachers is an integral part of teaching learning, and it assesses each student. It facilitates student learning through regular feedback; and therefore it is formative in nature. The public examinations could be conducted at various levels including national, regional, district, resource centre and/or school levels with the purpose of assessing each student, mostly by external agency such as government or independent/ autonomous agency. Public examinations are generally detached from classroom, and they are summative in nature. On the other, the main objective of large-scale national assessment is to provide system level information to policy makers, program developers and implementers.

However, the ultimate aim of system improvement is to improve student learning. Systematic practice of national assessment in Nepal has been begun since the establishment of Education Review Office (ERO) in 2010, so far has accomplished two rounds of student assessment taking large samples of schools and students.

Carefully designed national assessment will support in improving system level accountability; and therefore it will contribute to improve the quality of and equity in education. Most importantly, national assessment should be designed in such a way that it should supplement other activities of education system. However, national assessment of student achievement does not provide every solution to make efficient education system and delivery process. Evidences suggest that unnecessary emphasis on assessment sometimes causes negative effects instead of improving student learning (see, Berliner et al., 2000; Volante and Ben Jaafar, 2008; Leighton, 2009, quoted in Griffin, Care, Francis & Claire, 2014). Therefore, the balanced practice of assessment and teaching learning could maximize learning if the assessment results are properly used for the improvement of classroom delivery process as well as overall education system. Moreover, to maximize the contribution for the improvement of quality of and equity in education by means of assessment system, the quality of NASA, school based assessment and public examination need to be improved simultaneously.

Analysis of enabling context, system alignment and quality of NASA in Nepal shows that the enabling context is still very weak as there is a lack of proper institutional arrangement. In the case of alignment with system, though the NASA tests were based on the curriculum prescribed by the government, assessments were carried out without properly defined assessment framework. As a result, there might be mismatch in alignment. To ensure quality in the assessment, tests were standardized in which items were pre-tested before finalizing, items having appropriate level of reliability were selected, and proper difficulty level of each item was assured. However, the depth analysis of curriculum to set explicit criteria for the standards is yet to be done; and proper institutional arrangement is also yet to be set with the necessary legal back up and enhanced institutional capacity.

References

Anderson, L.W. & Krathwohl, D. R., et al. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives.* New York: Longman.

Biggs, J. & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.

Clarke, M. (2012). *What matters most for student assessment systems? A framework paper.* Washington DC: The World Bank.

Darling-Hammond et al. (2013). *Criteria for high-quality assessment.* Stanford: Center for Opportunity Policy in Education, Stanford University.

EDSC. (2011). *A final report on national assessment of grade 10 students.* Sanothimi: Department of Education, GON.

ERO. (2013). *Report of national assessment of student achievement 2011, grade 8*. Sanothimi: Education Review Office.

ERO. (2014). *Results of national assessment of student achievement 2012, grades 3 and 5*. Sanothimi: Education Review Office.

ERO. (2015). *Report of national assessment of student achievement 2013, grade 8*. Sanothimi: Education Review Office.

ERO. (2015a). *Education Review Office: An introduction 2015*. Sanothimi: Education Review Office.

ERO. (2015b). *National assessment of student achievement: A glimpse of result.* Sanothimi: education Review Office (ERO).

Greaney, V. (1996). Stages in national assessment, in P. Murphy, V. Greaney, M. E. Lockheed & C. Rojas (eds). *National assessments: Testing the system, 51-66.* Washington D. C.: The World Bank.

Greaney, V. and Kellaghan, T. (eds.) (2008). *Assessing national achievement levels in education*. Washington DC: The World Bank.

Griffin, P., Care, E., Francis, M. & Claire, C.S. (2014). The role of assessment in improving learning in a context of high accountability; In Wyatt-Smith, Peta Colbert & Valentina Klenowski (eds.) *Designing Assessment for Quality Learning, 73-87.* New York: Spring.

IIEP. (2008). *National assessments of educational achievement*. Paris: The International Institute for Educational Planning (IIEP).

MOE. (1971). *National Education System Plan*. Kathmandu: Ministry of Education.

MOE. (2009). *The School Sector Reform Plan*. Kathmandu: Ministry of Education.

Murphy, P., Greaney, V., Marlaine, L., & Rojas, C. (1996). 'Introduction', in P. Murphy, V. Greaney, L. Marlaine & C. Rojas, (eds.) *National assessments: Testing the system, 1-8.* Washington DC: The World Bank.

Nwana, O. C. (1996). What are national assessments and why do them? In P. Murphy, V. Greaney, M. E. Lockheed & C. Rojas (eds). *National assessments: Testing the system, 21-32.* Washington D. C.: The World Bank.

Poudel, L. N. et al. (2015). *A background paper prepared for developing school sector development plan on student assessment and examinations (unpublished).* SSDP Thematic Working Group, MOE/ GON.

Sui-chu HO, E. (2013). *Student learning assessment (Asia-Pacific education system review series no. 5).* Bangkok: UNESCO.

Webb, N. L. (2002). *Depth-of-knowledge levels for four content areas.* Available at www. http://www.hed.state.nm.us/uploads/files.

Challenges in the National Level Criterion-Referenced Student Assessment

Jari Metsämuuronen, PhD,

Adjunct Professor and Senior Evaluation Specialist, Finnish Education Evaluation Centre

Abstract

This article discusses the challenges of developing a criterion-referenced student assessment system at national level. The basic concepts of norm-referenced and criterion-referenced assessment are introduced and some known challenges of criterion-referenced assessment are discussed. The main challenge in many practical situations is the lack of proper criteria as the basis for the criterion-referenced assessment. Different options, general and local criteria are discussed, and some solutions are suggested to face the challenges. An initial general framework for mathematics standard is introduced based on Common European Framework for Languages (CEFR).

Keywords: criterion-referenced assessment, norm-referenced assessment, performance standard, assessment standard, standardized test, general and local criteria

Introduction

It is evident that future personal development of individuals and the prospects of country largely depends on the knowledge and skills what students learn at school. Measuring such knowledge and skill is essential to tracking students' development and assessing the effectiveness of entire education programs and practices in order to rectify the course with additional intervention. Educational science has long been looking for devising an effective means that appropriately measures the students' learning. Though there are several means for measuring students' learning, standardized achievement tests either norm referenced or criterion referenced have widely been used. Since the late 20th century, there is a growing trend on the use of criterion-referenced test in order to assess students' mastery of knowledge and development of cognitive skills.

The aim of this article is to discuss challenges of developing and using criterion-referenced student assessment at national level. The article begins with the concepts of norm-referenced and criterion-referenced assessment and then discusses the challenges of criterion-referenced assessment. While presenting specific challenges it presents some practices on criterion as example as well as some dimensions of standards. It finally includes some concluding discussions. This article is developed mainly by reviewing relevant literatures as well as reflecting on the practices of criteria-based assessment.

Norm-Referenced vs. Criterion-Referenced Assessment

The assessment of learning outcomes of the students includes two main approaches: Norm-Referenced Assessment (NRA) and Criterion-Referenced Assessment (CRA). Essential in NRA is that the test items are carefully designed to maximize performance differences among test takers; the test, usually, is not meant to determine if students have achieved *specified* learning standards, or acquired *specific* skills (Norm-Referenced Test, 2015). NRA uses either the raw- or weighted total scores (or percentages of maximum scores) as the source of the decision. The common feature for NRA is that the final test score produces a *norm* with which the different groups (such as geographical areas or sexes) can be compared with each other. Hence, one may get to know that in a certain geographical area the results are *better* than in another area. However, in the norm-referenced testing one does not know how *good* the pupils in fact are, that is, what the real proficiency level is. For example, after analysing the Finnish results of PISA (Programme for International Student Assessment) where the Finnish students are at the very high level in the reading proficiency, and at the same time, the boys are much weaker in comparison to the girls. In this case, one may say, sarcastically though properly hitting the challenge in the norm-referenced testing, "*In Finland, there are the highest level weak students in the world*".

The classical and modern test theory is mainly involved with the Norm-referenced testing. The idea is to maximize the test reliability and validity, which involves with maximizing the reliability and validity of the test items. From the Test Theory viewpoint, within the NRA, the use of *raw scores* leads to the so-called (un-weighted) "alpha models" with alpha type of reliability estimates, that is, usually, to *Kuder-Richardson Formula 20* (KR20, Kuder & Richardson, 1937) or *Cronbach's alpha* (Gulliksen, 1950; Cronbach, 1951; Lord & Novick, 1968; see critics in Tarkkonen, 1987; Vehkalahti, 2000). These alpha reliabilities are, in practice, the most used way to estimate the reliability (Hogan, Benjamin, & Brezinski, 2000). The weighted scores lead either to the *factor scores* (Tarkkonen, 1987; Vehkalahti, 2000) or to the so-called modern test theory, that is, *item response theory* (IRT) modelling (i.e. Rasch,1960; Birnbaum, 1968; Lord & Novick, 1968; Mokken 1971; Stout, 2002).

The other approach in the student assessment is based on pre-set criteria, which is known as Criterion-Referenced assessment (CRA). Linn and Gronlund (2000, 42) define CRA as a type of assessment designed to provide a measure of performance that is interpretable in terms of a clearly defined and delimited domain of learning tasks. More precisely, criterion-referenced tests include items that are directly relevant to the learning outcomes to be measured, without regard to whether the items can be used to discriminate among students. No attempt is made to eliminate easy items or alter their difficulty. If the learning tasks are easy, then test items will be easy. The goal of the criterion-referenced test is to obtain a description of the specific knowledge and skills each student can demonstrate. (Linn & Gronlund, 2000, 43.) Hence, in CRA, the reliability issues are more or less secondary, but the validity issues are crucial; the test items need to be directly involved with the criterion.

A good example of functioning and active area of criterion-reference assessment is the language testing. There are several external criteria based on the known standards for language proficiency. Some of the well-known standards are Common European Framework for Reference of Language or CEFR, TOEFL, Cambridge Exam, and IELTS (see comparison of 57 different criteria at https://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages).

When the criterion or standard is known and in use, one of the many standard setting methods is used to determine the level of the applicant or test-taker. These methods for standard setting are *many* - Kaftandjieva calculated that there are more than 50 methods and many of those have several modifications (Kaftandjieva, 2004, 11). One of the recent ones is Metsämuuronen's Three-phased Theory-based and Test-centred method for the Wide range of proficiency levels (3TTW, Metsämuuronen, 2009; 2010; 2013) developed specifically for the use of national level assessment. From the Nepalese perspective, it may be specifically interesting because it has been used in Nepal in assessing the national proficiency level in Nepali (Acharya, Metsämuuronen, & Adhikari, 2013; Metsämuuronen, Acharya, & Aryal, 2013; ERO, 2014) and English (ERO, 2014). It has also been used in Finland in assessing the proficiency of Finnish for the Swedish speaking students (Toropainen, 2010), Finnish as the Second language (Kuukka & Metsämuuronen, 2016), Mathematics in Vocational Education (Metsämuuronen & Salonen, 2016), and Sustainable development in Vocational Education and training (Räkköläinen & Metsämuuronen, 2016).

Both approaches have their own place in assessment and evaluation. Norm-Referenced tests are relatively inexpensive to develop, simple to administer, and easy to score. When using wisely the multiple choice type of items, that is, the distractors are selected so that they indicate certain kinds of flaws in thinking, and Norm-Referenced tests can provide valuable information about student learning. Also, the quality of norm-referenced tests is usually high because they are developed by testing experts, piloted, and revised before they are used with students, and they are dependable and stable for what they are designed to measure. Norm-referenced tests can also help differentiate students and identify those who may have specific educational needs or deficits that require specialized assistance or learning environments. These tests are an objective evaluation method that can reduce bias or favouritism when making educational decisions. If there are limited places in a gifted and talented program, for example, one transparent way to make the decision is to give every student the same test and allow the highest-scoring students to gain entry. (Norm-Referenced Tests, 2015)

Criterion-Referenced tests are better suited to measuring *learning progress* than norm-referenced exams, and they may give educators information they can use to improve teaching and school performance. These tests may be fairer to students than norm-referenced tests because they do not compare the relative performance of students; they evaluate achievement against a common and consistently applied set of criteria. The tests apply the same learning standards to all students, which can hold underprivileged or disadvantaged students to the same high expectations as other students. (Criterion-Referenced tests, 2014)

Alone the fact that the criterion-referenced testing can be used in assessing the *progress* in learning outcomes in practical terms, while the norm-referenced testing cannot[1], gives an enormous advantage for CRA over NRA. Experiences in working with standards of language testing have made me; personally, favour CRA over NRA in the national level students' assessment in certain circumstances:

a) when the standards are available;

b) when the standards are expressed so precisely and exhaustively that the standards can be utilized in the item writing process;

c) when the standards show the practical differences between the levels; and

d) when the structure of the levels is nuanced so detailed that the students can be credibly be classified into the levels both in the beginning of the learning process as well as in the mastery level.

In many cases, however, the grade-wise aims in the national curricula are expressed so vaguely that the Criterion-Referenced testing is hardly possible in its strict form.[2] A vaguely wording in curriculum (type of "pupils are expected to write in a way which is *interesting*, conveying meaning *clearly* in the chosen form for an *intended* reader" [italics added by the writer]) leads to a situation where the criterion is interpreted differently by different scorers and across the different grades, depending on the programmes of study (Green 2002, p.7). It may also lead to a naïve criterion-referenced testing (type of "all curriculum-based assessment is criterion-based assessment"[3]) which can be criticized by using Angoff's (1974) words: "*one only has to scratch*

1 *Using linking items and IRT modelling within NRA, provides us the possibility to assess the change in learning outcomes. However, the output is still based on comparing the individuals into the norm. Hence, though we can say that the achievement level has raised or become lower, we still cannot say in the standard procedures, whether all the students have exceeded the level of "good" or whether all the students are below "fail".*

2 *Green (2002, p.7) uses term "true" criterion referencing and refers to Popham (1980). Green guesses that Popham (1980) would not accept criteria, which could allow a range of interpretations as is needed in loosely worded and categorized criterion.*

3 *From the Glossary of Education Reform (http://edglossary.org/) one easily finds the spirit of this kind of naïve criterion-based assessment: "Criterion-referenced tests created by individual teachers are also very common in American public schools. For example, a history teacher may devise a test to evaluate understanding and retention of a unit on World War II. The criteria in this case might include the causes and timeline of the war, the nations that were involved, the dates and circumstances of major battles, and the names and roles of certain leaders. The teacher may design a test to evaluate student understanding*

the surface of any criterion-referenced assessment system in order to find a norm-referenced set of assumptions lying underneath". If failing in the criterion-based assessment, "*we are in danger of implementing a system of tests that behave like thermometers, all pretending to measure on the Celsius scale, but which actually each have their own freezing point and each their own idea of what constitutes a nice summer's day."* (Pollitt, 1994, p. 69).

The basic thing in order to avoid the basic pitfalls of Criterion-Referenced assessment is to recognize its weak points. Some of these challenges are discussed in more details as follows:

Challenges in Criterion-Referenced Assessment

The succeeding section points out main challenges regarding CRA categorizing them general and specific.

General challenges in CRA

Both the NRA and CRA approaches include known challenges. Both can be used wrongly and both can lead to wrong conclusions when used improperly. The (unknown) writer in the Glossary of Education Reform puts is as follows:

"[The] perceived performance on a standardized test can potentially be manipulated, regardless of whether a test is norm-referenced or criterion-referenced…. For this reason, it is important to investigate the criteria used to determine "proficiency" on any given test—and particularly when a test is considered "high stakes," since there is greater motivation to manipulate perceived test performance when results are tied to sanctions, funding reductions, public embarrassment, or other negative consequences." (Norm-Referenced Test, 2015).

Some general, critical, notes on the criterion-referenced testing can be summarized as follows (Criterion-Referenced Tests, 2014):

of the criteria and determine a minimum passing score." (Criterion-Referenced Test, 2014) Though the history teacher in the citation may know something of the World War II, it remains more or less unknown what would be the true achievement level in history as a subject. Similarly challenging is Wiliam's (1993, p. 341) defence of the loose wording for criterion: "[N]o criterion, no matter how precisely phrased, admits of an unambiguous interpretation. … [W]e have to use norms, however implicitly, in determining the appropriate interpretations… [T]he criterion is interpreted with respect to the target population". Alone, the fact that in Language testing the international board for CEFR has been able to create standards independent on the target population makes Wiliam's argument seem an excuse for the low-levelled Criterion-Referenced testing.

- **The tests are only as accurate or fair as the learning standards upon which they are based**. If the standards are vaguely worded, or if they are either too difficult or too easy for the students being evaluated, the associated test results will reflect the flawed standards. Green (2002) notes that difficulties can arise when level descriptions do not give clear definitions of progress or do not relate to realistic progression. Cox (1995) pointed that, in England, the level descriptions in the national curriculum did not have the carefully defined progression that was necessary to allow reliable interpretations.

- **The process of determining proficiency levels and passing scores on criterion-referenced tests can be highly subjective or misleading**—and the potential consequences can be significant, particularly if the tests are used to make high-stakes decisions about students, teachers, and schools. Because reported "proficiency" rises and falls in direct relation to the standards or cut-off scores used to make a proficiency determination, it is possible to manipulate the perception and interpretation of test results by elevating or lowering either standards and passing scores. Even the reputations of national education systems can be negatively affected when a large percentage of students fail to achieve "proficiency" on international assessments. Even if not manipulating the interpretation, transforming the total score into the proficiency levels may lead to odd and implausible results as shown by Metsämuuronen (2013) and Metsämuuronen, Acharya, and Aryal (2013).

- **The subjective nature of proficiency levels allows the tests to be exploited for political purposes to make it appear that schools are either doing better or worse than they actually are**. For example, some states in the USA have been accused of lowering proficiency standards of standardized tests to increase the number of students achieving "proficiency," and thereby avoid the consequences—negative press, public criticism, large numbers of students being held back or denied diplomas (in states that base graduation eligibility on test scores)—that may result from large numbers of students failing to achieve expected or required proficiency levels.

Specific challenges in CRA – Creating the criterion from the scratch

The ultimate challenge for the Criterion-Referenced Assessment is the existence of the criterion. In languages, there are several options to use but what about Mathematics or History or Science? There are *no* internationally recognized criteria for any of the other school subjects than languages. When there are *no* criteria at all or the "criteria" are so vaguely worded in the curriculum that it is difficult (or impossible) to create the set of true criterion-referenced tests on the basis of those, what can one do? Where to start the process of setting the criteria? What to take into account? Some ideas are shared here; the Mathematics is taken here as an example of the standards.

The first challenge in creating or developing the criterion is to decide whether one is willing to select or create a **general standard** or a **local standard**. The second challenge is related to the **domains** of the new standards. The third challenge is related to the **hierarchical levels** in the standards. These are handled in what follows.

General Criteria

The first challenge in creating or developing the criterion is to decide whether one is willing to select or create a **general criterion** or a **local criterion**. An example of the *general criterion* is the aforementioned CEFR classification. The original CEFR classification includes six stages of language proficiency based on action-driven linguistic theories. In the CEFR classification, the original set of proficiency levels are fixed to six levels: Breakthrough or beginner (A1), Way stage or elementary (A2), Threshold or intermediate (B1), Vantage or upper intermediate (B2), Effective Operational Proficiency or advanced (C1), and Mastery or proficiency (C2). Naturally, the contents differ in different areas of language (reading, writing, listening, and speaking). In Finland, it was noticed that six basic levels were not a fruitful basis for student assessment in schools. Hence, before the Finnish National Board of Education (FNBE) started to use the CEFR levels in the student assessment of languages in the core curriculum of the year 2004 (FNBE, 2004), the national experts of CEFR levels divided the classification into more precise levels which are now used in teaching and student assessment in Finnish schools (Table 1). The levels higher than C1.1 are not defined in the Finnish system because it is not expected for anyone to reach fluency in the foreign language acquisition within the school years. The summarized descriptions of the contents for the topics of reading and writing are seen on Table 2.

Table 1. CEFR levels used in the Finnish core curriculum (FNBE 2004, 2015)

CEFR level	Short Description
A1.1	First stage of elementary proficiency
A1.2	Developing elementary proficiency
A1.3	Functional elementary proficiency
A2.1	First stage of basic proficiency
A2.2	Developing basic proficiency
B1.1	Functional basic proficiency
B1.2	Fluent basic proficiency
B2.1	First stage of independent proficiency
B2.2	Functional independent proficiency
C1.1	First stage of fluent proficiency

From the mathematics viewpoint, the general criteria are not necessarily that transparent as in languages. On the other hand, one can think that mathematics learning shares somewhat the same logic as learning languages – all in all, mathematics *is* an ultimate language with its own syntax and logic. Also, as in language learning, the new material is more or less cumulative also in

mathematics. Thus, let us see what would be a parallel systemic for Mathematics as is the FNBE extension of CEFR.[4]

a) Dimensions in general standards

The first thing is to decide what would be the dimensions of the standards. There are several possibilities to select. The *content area-wise division* would lead to a direction of "Algebra", "Arithmetic", "Geometry", "Percentages", "Sets", "Statistics", and so on. This division is supported by the fact that there are content-wise specialties to learn – problems in one content area do not necessarily correlate with those in other content area (Räsänen, 2015). A *competence-wise division* may divide mathematics learning into 6–8 dimensions of competencies (Hannula, 2015): in Europe, some popular classifications are Niss-Jensen-Højgaard model (Niss & Jensen, 2002; Niss & Højgaard, 2011) or rather its further adaption by Lithner and colleagues (2010) and a further reduced model by Säfström (2013). The original Niss-Jensen-Højgaard model comprises eight competencies (Mathematical thinking competency, Problem tackling competency, Modelling competency, Reasoning competency, Representing competency, Symbol and formalism competency, Communicating competency, and Aids and tools competency). Lithner and colleagues (2010) reduced the competencies into six (Problem solving competency, Reasoning competency, Applying procedures competency, Representation competency, Connection competency and Communication competency).[5] Säfström (2013) reduced further the competencies into five (Representation competence, Competence in procedures, connection competence, Reasoning Competence, and Communication Competencies). The potential challenge in the competence-based classifications above is that they are not developed for criterion-referenced testing but rather for educational purposes. Though the categorization seems relevant, the original classification and its use seems not taking into account differences in achievement levels.[6] A third direction would

4 *This exercise, though not necessarily being very serious, is boosted by discussions with several Finnish experts of learning. A set of inspiring email discussions of the matter was participated by Professor Markku Hannula (Pedagogics of Mathematics in Helsinki University), Doctor Laura Tuohilampi (Pedagogics of Mathematics in Helsinki University), Senior Researcher Pekka Räsänen (Neuropsychology from the Niilo Mäki institute for learning disabilities in Jyväskylä University), Research Specialist Markus Mattsson (Cognitive psychology in Helsinki University), and Professor Jari Lavonen (Head of Teachers education faculty in Helsinki University, specialized in History, Social Studies and Science). All shared somewhat the same opinion that creating a universal criterion for mathematics (or any other school subject) would be an interesting though demanding task.*

5 *Lithner and colleagues use here the word 'ability' instead of 'competence'. I have used here the 'competence' for consistency reasons.*

6 *Originally, in both the models of Niss-Jensen-Højgaard and Lithner et al. use three-levelled grading in the dimensions: Interpret, Do and use, and Judge. Now, think about a first grade student with the problem of 1 + 2 = ?. When (s)he solves the problem (that is, "interpreted" the task properly, "did and*

be to utilize the *hierarchical structure of proficiency in mathematics*. This direction is elaborated in what follows.

b) Hierarchical structure in the general standards

A tentative classification – let us call the systemic as General Framework in Reference of Mathematics (GFRM) – based on the universal hierarchical structure in proficiency in mathematics may divide achievement into three dimensions: 1) proficiency in concepts, 2) proficiency in operations, and 3) proficiency in mathematical abstractions and thinking (Table 2).[7] Two first dimensions are to do with mechanical calculation and the third one is to do with changing the problems in a mathematical form. The last one may borrow the areas of competency from Niss-Jensen-Højgaard model, for example.

Table 2. Short descriptions of the GFRM levels

GFRM level	Short Description
A1.1	First stage of elementary proficiency
A1.2	Developing elementary proficiency
A1.3	Functional elementary proficiency
A2.1	Developing of basic proficiency
A2.2	Functional basic proficiency
B1.1	First stage of advanced proficiency
B1.2	Developing advanced proficiency
B2.1	First stage of Functional advanced proficiency
B2.2	Functional advanced proficiency
C1	Basic Professional level
C2	Advanced Professional level

used" proper mathematical tools, and "judges" and evaluates the validity of a solution), the beginner mathematician seem to get the highest level grading in all six or eight dimensions even though (s)he, apparently, is quite far from the real mastery in mathematics.

7 *One may see here parallelism with the competence-based systemic of Niss-Jensen-Højgaard (Niss & Højgaard, 2011; Niss & Jensen, 2002) and Lithner et al. (2010) where their original levels of competence can be rephrased as the levels of Recognizing, Doing and Using, and Evaluating and Judging (Hannula, 2015).*

The rationale of the two first dimensions is somewhat obvious. In order to master even the most simplest and mechanical mathematical task, certain concepts are needed: the concepts of numbers and their representations (like 'five' = 5 = ✋ = *****) as well as their consecutive nature. For the geometry, certain basic shapes such as triangle, square, and circle should be recognized and remembered. Some of the relevant concepts are collected in table 3 and initially divided into levels parallel to CEFR levels. The concepts are not independent of the operations. Primarily the operations can be taken as the engine of the two: one can ask what are the concepts needed in order to master the mathematical operations. On the other hand, the operations are more or less hierarchically organized in the normal educational process: for example, in order to be able to manage the powers, the procedures of multiplication is wise to learn first. In addition, it is wise to start learning mathematics with concrete things such as summing and subtracting the natural numbers.

Though the basis of GFRM comes from the CEFR levels and the logic seem to follow the basic logic of CEFR classification, one may note that the names of the levels are mainly different from CEFR – only the names of the elementary basics (A1.1 to A1.3) are the same. In GFRM, the idea is, contrary to CEFR systemic, that the A level is more or less the basic level with the relevance in the everyday life. The B level is advance level with less relevance with the everyday life but with high relevance with further studies in several professional areas like statistics, engineering, or economics. The C level is left for the professional level mathematics needed either in the practical fields (like for Statisticians, Advanced Researchers, Economist, or Engineers) or in the theoretically oriented fields (like for the professors or researchers of pure mathematics, physics, astronomy, or chemistry). It is good to remember that Mathematics as a scientific domain is so wide that no one masters all the areas any more. In a congress of mathematics, the researches in the neighbouring room do not necessarily understand anything of the other room's contents. It is thus too much to assume that any professional at the level C would remember all the things are the level B by heart. However, after participated the courses, it is possible to keep the minimum level on each proficiency area: to remember the basic concepts and to remember which operation should be selected to a problem. Most of the facts and nuances can easily be recall from the textbooks when knowing the connection of the concepts and the operations.

c) Initial suggestion as a general framework of standards in reference of mathematics

The table 3 is an attempt to show what kind of set of criteria a general standard could be. More precise descriptions for different content areas should be added in order to use it fully in all grades of the compulsory education. The proficiency-based criteria in table 3 is not very practical when it comes to assess the professional mathematicians or university level mathematics students' proficiency levels. However, it may fulfil quite reasonable the needs of compulsory education up to +12 grade. This kind of standard could be used easily as a basis for the item writing; the items in different levels can be identified. One may note that, in each level, the items can be easy, medium or demanding on the topic. The hierarchical nature of the standards makes it happen that even the easiest item of level B1.1 cannot be mastered by the students at the very elementary

level. This leads to a possibility that more or less standard (norm-referenced type of) test can be administered within the level, the score and cut-offs can be used in standard setting process. The advantage of the classification is that the proficiency level can be defined by knowing which kind of items the test-taker manages.

Preparing these kinds of standards is a huge task and it requires involving experienced mathematics teachers and curriculum developers, trade unions, psychologists and politicians to participate the process. Without a large consensus, it is difficult to convince the audience of the rationale behind the levels. The challenge in creating the general standard is that if and when there are no explicit and strict criteria for the different school subjects or they are worded vaguely in the curricula (as, for example, in Mathematics), it easily takes lots of time to convince all the relevant players in the field of the standards. In Finland, developing criteria just for *one* level "good", took several years and lots of discussions between the different stakeholders. The European Standard for Criterion for Languages (CEFR) took 10 years to build up.

Table 3. An initiative sketch for proficiency-based criterions for mathematics

Proficiency level	Condensed description of the proficiency	1) Proficiency in Concepts	2) Proficiency in Operations	3) Proficiency in abstractions and mathematical thinking
A1.1 First stage of elementary proficiency	• Is familiar with the numbers, but the use in mathematical operations is very limited. • Recognizes the basic two-dimensional shapes (circle, square, and triangle) and their three-dimensional counterparts (ball, box, and pyramid) and can tie their name with pictures. • can express some limited mathematical expressions, such as order of the numbers	• knows the numbers 1 – 10 • understands the consecutive nature of numbers • recognizes the basic shapes (circle, square, triangle, ball, box, and pyramid) • do not necessarily understand the concept of zero	• can form the consecutive order of the numbers • can categorize the shapes in to groups (do not mess with different sizes, colors, and positions) and can tie their name with pictures. • cannot use even the plus or minus operations	• have the basic understanding of adding, taking away, dividing something equally, and multiplication by using adding as a rationale • have the basic understanding of unseen numbers (for example, what number is missing in the consecutive order) • cannot formulate a mathematical abstraction and operands in a formal way • cannot judge which of two pictures/computations corresponds with the oral mathematical task. • cannot recognize whether the logical inference is correct or not (type of: "I have a number which is bigger than 2. is it smaller than 5?") • Cannot connect the numbers with geometry (triangle – three angles; square – four angles)

Proficiency level	Condensed description of the proficiency	1) Proficiency in Concepts	2) Proficiency in Operations	3) Proficiency in abstractions and mathematical thinking
A1.2 Developing elementary proficiency	• can use natural numbers with fluency at range 1 - 100 • Can operate with basic operations of summing and subtracting • Understands the concept of zero	• knows the numbers 1 – 100 • can use the knowledge of consecutive nature in numbers • recognizes the basic shapes and can connect them in basic mathematical operations • have the basic understanding of the concept of zero • Do not have the concept of Rationale number	• can form simple series • can use the basic shapes in simple calculations in geometry • can use plus and minus operations in range 1 – 100 • cannot use multiplication and division in a formulated way	• can judge which of the possible strategies in summing up and subtraction is most effective (like 6 + 7 can be solved as 6 + 4 + 3 or 6 + 6 + 1 or 5 + 5 + 1 + 2) • Can do basic logical inference of hidden numbers (type of X < 5 and X > 3, X = 4) • can formulate a mathematical abstraction and operands from an oral task • can judge which of two pictures/ computations corresponds with the oral task. • can recognize whether the logical inference is correct or not (type of: “I have a number which is bigger than 2. is it smaller than 5?”) • Can connect the numbers with geometry (triangle – three angles; square – four angles)

Proficiency level	Condensed description of the proficiency	1) Proficiency in Concepts	2) Proficiency in Operations	3) Proficiency in abstractions and mathematical thinking
A1.3. Functional elementary proficiency	• can use natural numbers with fluency at range - ∞ - +∞ • Can operate with basic operations of plus, minus, multiplication, and division with Natural numbers • Understands the concept of Rationale number	• Understand the negative numbers • understands the numbers with fluency at range - ∞ - +∞ • Knows elementary specific terms of geometry, such as rectangle, catheti, hypotenuse… • have the basic understanding of the concept of Rationale number • Have a basic understanding of the concept of brackets but cannot necessarily use it without flaws	• can use basic operations plus, minus, multiplication, and division • Can understand and operate with series with natural numbers • can do the basic calculations of rectangular and circle • understands the meaning of the order in the calculations but may make mistakes in the actual calculations	
A2.1 Developing basic proficiency	• understands and applies percentages • Understands rations and probability • masters the order in calculations	• concept of remainder in division • concepts of ratios, proportions, and percentages • Have a basic understanding of the concept of brackets and can use it without flaws	• division and reminder with rational numbers • calculation of percentages • no mistakes in order of calculations • Cannot operate with Powers and Roots	• Logical inference on the basis of percentages
A2.2 Functional basic proficiency	• Masters the basics of Powers and its connect to multiplications • Masters simple roots up to 3 and its connection to practical situations.	• concepts of Powers and Roots	• Powers (on the basis of multiplication) • Roots up to 3 (practical application with the box)	

Proficiency level	Condensed description of the proficiency	1) Proficiency in Concepts	2) Proficiency in Operations	3) Proficiency in abstractions and mathematical thinking
B1.1 First stage of advanced proficiency	• masters the Numbers and sets of numbers Polynomial functions, and advanced Geometry	• understands the concept of sets of numbers • understands the polynomial functions • understands the concepts of sine and cosine	• Numbers and sets of numbers (A1) • Polynomial functions (A2) • Advanced Geometry (A3)	
B1.2 Developing advanced proficiency	• masters the Vectors, Analytical geometry, and Derivatives	• understands the concept of the vector • understands the concept of derivatives	• Vectors (A4) • Analytical geometry (A5) • The Derivatives (A6)	
B2.1 First stage of Functional advanced proficiency	• masters Trigonometric functions and number sequences, Radical and logarithmic functions, and Integral calculus	• understands the concept of the trigonometric functions • understands the concepts of logarithm function • understands the concept of the integral function	• Trigonometric functions and number sequences (A7) • Radical and logarithmic functions (A8) • Integral calculus (A9)	
B2.2 Functional advanced proficiency	• masters the Advanced Probability and Statistics, Number theory and logic, Advanced Algorithms in mathematics, and Advanced differential and integral calculus	• understands the concept of combinatorics • understands the concepts of the open statement and quantifier • understands the concepts of the absolute error and the relative error • understands the idea of differentiability of functions	• Advanced Probability and Statistics (A10) • Number theory and logic (A11) • Advanced Algorithms in mathematics (A12) • Advanced differential and integral calculus (A13)	
C1-2 Professional level – not relevant within the Basic education	• Specified professional level in different domains in mathematics	• Concepts of plains • Concepts of vectors and matrices • …	• Projections in plains • matrix and vector algebra • …	

Local Standards

Another option is to create a local criterion for the assessment. Then the criterion can be very free – the main thing is that it satisfies the local needs. Ultimately, an individual teacher can prepare his or her own standards based on the core curriculum, the District can create its own standards, the Development region can create its own criteria, or the National agencies, like ERO can create their own way of assessing the student performance.

The local standard can be formulated grade-wise or, as in Finland, so-called joint-point-wise (in mathematics for the end of grade 2, grade 6, and grade 9). The latter takes less time but the criteria cannot be used for each grade. The local standards can be simple, such as giving the passing-failing signal, or they can be more or less nuanced. A simple division is to divide the proficiency into "satisfactory", "good", and "excellent". The latter one can easily be broken to 6- or 9-level classification. The first challenge is, what would be the composition of the panel of experts, which would decide what kind of performance is "good" and which is just "satisfactory". The second challenge is, how is this panel would judge which of the students would be ranked as "failed"? that is, how to determine the cut-off scores for the level?

In Finland, in assessing the mathematics proficiency in vocational education and training, the curriculum explicates which kind of achievement is needed in order to show "satisfactory", "good", and "excellent" mathematics proficiency. Experienced mathematics teachers classified the items into "baskets" on the basis of their subjective experience of their students and teaching: all items were analysed on the basis of what kind of ability level would be needed in order to solve the task. At the end, all the items were classified into the baskets of "satisfactory", "good", and "excellent". The final standard setting was done by using Metsämuuronen's (2013) 3TTW method (Metsämuuronen & Salonen, 2016).

Green (2002) suggests somewhat the same kind of process. Green notes that difficulties can arise when level descriptions do not give clear definitions of progress or do not relate to realistic progression. Hence, she proposes a crude steps for creating the criteria:

a) As an initial guide, draft criteria could be selected in consultation with teachers.

b) The teachers could suggest how many of their children could be expected to achieve each level.

c) Those criteria could be evaluated using a pre-test, possibly using data from an initial pilot year.

d) The children's actual performance (e.g. answers to questions as well as mark distributions) would be used to improve the draft criteria.

e) The initial normative approach could then give way to criterion referencing once the scale has been determined.

Green reminds us that if descriptions are written without reference to empirical data on pupil performance, there is a danger that unrealistic standards will be set.

Concluding Discussion

Though the Criterion-referenced student assessment is preferred over the Norm-Referenced assessment because the progress can be monitored, there is a real danger in turning to naïve or quasi criterion-referenced testing when there is not developed a shared understanding of holistic descriptions of performance which may lead to the problems of 'unreliable' human judgements. Very often, the curriculum is vaguely worded when it comes to the criterion or the curriculum lack totally the criterion references for critical levels, such as "failing", "sufficient", "good", or "excellent" proficiency. Even if the curriculum includes such goal references and criterion, reminds Greene (2002), those involved in assessment, test development, teaching and curriculum development need to understand levels of performance and the nature of progression in the curriculum and to develop an understanding of standards of performance within a community of practice. Such a body of knowledge would help to increase the credibility of valid, reliable assessment of what students know, understand and can do in the context of transparency, clarity and shared understanding.

Experienced actors in the field of criterion-referenced testing (e.g. Green 2002, p.10; Wiliam, 1996) remind us that resources and time are needed to develop a shared understanding of progression in the curriculum and to enable reliable assessment to become part of an integrated educational process; it is difficult to see possibility of establishing comparability of standards other than through the professional judgement of a community of experts.

The initiative attempt to create a basis for hierarchically structured proficiency standard for mathematics should be revisited by the experts in mathematics teaching. As it is now – more or less a rough sketch, I hope, it can courage the reader to bring some ideas how to develop it further.

References

Acharya, D. R.; Metsämuuronen, J. & Adhikari, H. (2013). Nepali Achievement in NASA 2011. In J Metsämuuronen & BR Kafle (eds). *Where Are We Now? Student achievement in Mathematics, Nepali and Social Studies in 2011*. Kathmandu: Ministry of Education.

Angoff, W. H. (1974). Criterion referencing, norm referencing and the SAT, *College Board Review*, 92, 3-5.

Birnbaum, A. (1968). Estimation of ability. In F. M. Lord & M. R. Novick, *Statistical Theories of Mental Test Scores.* Reading, Mass.: Addison–Wesley Publishing Company.

Cox, B. (1995). *The Battle for the English Curriculum* London: Hodder & Stoughton.

Criterion-Referenced Test. (2014). In S. Abbott (Ed.), *The glossary of education reform*. Retrieved from http://edglossary.org/criterion-referenced-test/

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. http://dx.doi.org/10.1007/BF02310555.

ERO. (2014). *Student achievement in Mathematics, Nepali and English at grade 3 and 5 in 2012*. Sanothimi: ERO, Manuscript.

FNBE. (2004). *National Core Curriculum*. Yliopistopaino: Helsinki.

FNBE. (2015). *National Core Curriculum*. Yliopistopaino: Helsinki.

Green, S. (2002). *Criterion Referenced Assessment as a Guide to Learning – The Importance of Progression and Reliability.* A paper presented at the Association for the Study of Evaluation in Education in Southern Africa International Conference Johannesburg, 10 – 12 July 2002.

Gulliksen, H. (1987/1950). *Theory of Mental Tests.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Hannula, M. (2015). Email discussion at 2015 November 6th.

Hogan, T. P., Benjamin, A. & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*(4), 523–531. http://dx.doi.org/10.1177/00131640021970691.

Kaftandjieva, F. (2004). Standard Setting. Reference Suplement, Section B in S. Takala (Ed.) (2009). *Relating Language Examinations to the CommonEuropean Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual.* Language Policy Division, Strasbourg.

Kuder, G. F. & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2, 51–160.

Kuukka, K. & Metsämuuronen, J. (2016, forthcoming). *Student achievement in Finnish as Second Language.* FINEEC.

Linn, R. L. & Gronlund, N. E. (2000). *Measurement and assessment in teaching (8th ed.).* Upper Saddle River, NJ: Prentice Hall.

Lithner, J., Bergqvis, E., Bergqvist T.; Boesen, J., Palm, T. & Palmberg, B. (2010). Mathematical competencies: A research framework. In C Bergsten, E Jablonka & T Wedege (Eds.), *Mathematics and mathematics education: Cultural and social dimensions*. Proceedings of MADIF 7, the Seventh Mathematics Education Research Seminar, Stockholm, January 26-27 (pp. 157–167). Linköping, Sweden: Svensk förening för matematikdidaktisk forskning, SMDF.

Lord, F. M. & Novick, M. R. (1968). . Reading, Mass.: Addison – Wesley Publishing Company.

Metsämuuronen, J. & Salonen, R. V. (2016, forthcoming*). Student achievement in Mathematics in Vocational Education.* FINEEC.

Metsämuuronen, J. (2009). *Ratkaisuprosentin muuntaminen taitotasoksi – A-Finska-oppiaineen aineiston erityiskysymyksiä [Changing the percentage of correct answers into proficiency level – Specific questions of A-Finnish assessment]. Unpublished memorandum* 28.12.2009 [In Swedish].

Metsämuuronen, J. (2010). Omvandlingen av poängtal till färdighetsnivåer inom det receptive delområdet av provet i A-lärokursen och i den modersmålsinriktade lärokursen. [*Changing the percentage of correct answers into proficiency level in Finnish language for Swedish speakers A-course and native speakers' course test].* Appendix VII in O. Toropainen (2010). *Utvärdering av läroämnet finska i den grundläggande utbildningen. Inlärningsresultaten i finska enligt A-lärokursen och den modersmålsinriktade lärokursen i årskurs 9 våren 2009*. Uppföljningsrapporter 2010:1. Utbildningsstyrelsen. Helsinki. Retrieved from http://www.oph.fi/publikationer/2010/utvardering_av_laroamnet_finska_i_den_grundlaggande_utbildningen [Swedish], 164–168.

Metsämuuronen, J. (2013). A New Method to Setting Standard for the Wide Range of Language Proficiency Levels. *International Education Research*, 1(1), http://www.todayscience.org/IER/v1/Jari%20Metsamuuronen.pdf.

Metsämuuronen, J; Acharya, S.P. & Aryal, H. (2013). Setting Standard in Nepali Language – CEFR levels. In J Metsämuuronen & BR Kafle (eds). *Where Are We Now? Student achievement in Mathematics, Nepali and Social Studies in 2011*. Kathmandu: Ministry of Education, 393–422.

Mokken, R. J. (1971). *Theory and Procedure of Scale Analysis*. New-York: De Guyter.

Niss M. & Højgaard. T. (Eds.) (2011). *Competencies and Mathematical Learning. Ideas and Inspiration for the development of mathematics teaching and learning in Denmark.* English Edition, October 2011. Roskilde University, Department of Science, Systems and Models, IMFUFA. nr. 485 – 2011. Available at http://rudar.ruc.dk//bitstream/1800/7375/1/IMFUFA_485.pdf.

Niss, M. & Jensen TH (Eds.) (2002). *Kompetencer og matematiklæring: Ideer og inspiration til udvikling af matematikundervisning i danmark.* København: Undervisningsministeriets forlag.

Norm-Referenced Test. (2015). In S. Abbott (Ed*.), The glossary of education reform*. Retrieved from http://edglossary.org/norm-referenced-test/

Pollitt, A. (1994). 'Measuring and evaluating reliability in national curriculum assessments' in Hutchinson, D. and Schagen, I. eds, (1994) *How reliable is national curriculum assessment?* London: NFER.

Popham, W. J. (1980). Domain specification strategies. In Berk, R.A. ed, *Criterion-referenced measurement: the state of the art*. Baltimore: John Hopkins University Press, 15-31.

Räkköläinen, M. & Metsämuuronen, J. (2016, forthcoming). Student achievement in Sustainable Development in Vocational Education. FINEEC.

Räsänen, P. (2015). Email discussion at 2015 November 9th.

Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research.

Ridgway, J. (1988). *Assessing Mathematical Attainment* London: NFER-Nelson.

Säftström. A. I. (2013). *Exercising Mathematical Competence. Practising Representation Theory and Representing Mathematical Practice.* Thesis for the Degree of Doctor of Philosophy. Göteborgs Universitet. Available at http://hdl.handle.net/2077/32484.

Stout, W. (2002). Psychometrics: From Practice to Theory and back. 15 Years of Nonparametric Multidimensional IRT, DIF/Test Equity, and Skills Diagnostic Assessment. *Psychometrika* 67(4), 485–518. http://dx.doi.org/10.1007/BF02295128.

Tarkkonen, L. (1987). *On Reliability of Composite Scales. An Essay on the measurement and the properties of the coefficients of reliability-an unified approach*. Tilastotieteellisiätutkimuksia 7. Helsinki: Finnish Statistical Society.

Toropainen, O. (2010). Utvärdering av läroämnet finska i den grundläggande utbildningen. Inlärningsresultaten i finska enligt A-lärokursen och den modersmålsinriktade lärokursen i årskurs 9 våren 2009. Uppföljningsrapporter 2010:1. Utbildningsstyrelsen. Helsinki. Retrieved from http://www.oph.fi/publikationer/2010/utvardering_av_laroamnet_finska_i_den_grundlaggande_utbildningen [In Swedish]

Vehkalahti, K. (2000). *Reliability of Measurement Scales. Statistical Research Reports 17*. Finnish Statistical Society. Retrieved from http://ethesis.helsinki.fi/julkaisut/val/tilas/vk/vehkalahti/

Wiliam, D. (1993). Validity, dependability and reliability in National Curriculum assessment, *The Curriculum Journal*, 4.3, 335-350.

Raising Students' Learning Achievement: Concerns and Context of Nepal

Gopal Prasad Bhattarai, Under Secretary, Education Review Office

Abstract

Understanding the significant role of formal schooling in fostering knowledge and skills among youngsters to achieve individual, social, economic development and generating employability, countries around the world have expanded schooling as an effective developmental strategy that succeeded in achieving nearly universal enrolment and attainment of schooling. The implicit assumption behind the effort was that universal attainment of schooling would yield expected level of knowledge and skills among the youths, which have impacts on individual's success in labour market outcomes, economic growth of nation, sustainable development in society and the reduction in social inequality. However, even after achieving universal attainment of basic schooling did not result in the expected level of learning, recent educational agenda include a stronger focus on raising the level of learning achievement at basic level to all as the main concern. Conceptualizing the learning achievements, this paper first discusses the main concerns behind improving learning achievement. Then, in the succeeding part, it tries to present a glimpse of learning achievement level in the context of Nepal drawing upon the evidences from large-scale assessments carried out so far.

Keywords: cognitive skills, educational attainment, learning achievement, literacy skills, schooling

Introduction

Since the early times, significance of formal schooling has long been well established for its role in equipping youngsters with knowledge and skills essential for individual development and economic productivity. Its importance lies first in developing specialized knowledge in core subjects such as language, mathematics, science, history, geography and the like that helps to broadening the understanding and forming concepts on the lived physical, socio-cultural and economic world. With these knowledge and understanding on core subjects, individuals learn to use their minds well so that they are prepared to handle practical problems of daily life, be a responsible citizen, and to grab a productive employment in the economy (Patrick, 1991). Along with the subject knowledge and understanding, schooling also fosters among youngsters social skills and values that help them succeed in the adult life. Most importantly, its significance lies particularly in developing basic cognitive skills in the forms of literacy and numeracy skills--the reading writing ability--which in turn form the strong, foundation for the development of more complex skills such as problem solving, reasoning, information processing, advanced thinking, communication skills, and ability to innovate or adopt new technology associated with production functions. The skills developed through schooling also determine the young people's prospects of entering job market, enhance their productivity and enable them to adopt better to rapidly changing economy (King & Palmer, 2006 cited in Rolleston & James, 2014).

Understanding the importance of these knowledge and skills on which individual success in the job market outcomes, economic growth of a nation and the effort for promoting equality in society largely depend; nation around the world made significant commitments for ensuring universal schooling as a central part of development strategy. Since achieving universal primary education has been on the global agenda as the Universal Declaration of Human Rights 1948, the Convention on the Rights of the Child 1989 and other international conventions recognize children's right to free and compulsory education. Therefore, over the past 20 years, it has developed into an international priority. In 2000, the United Nations adopted the Millennium Declaration and laid out a road map for achieving the Millennium Development Goals (MDGs) with a series of development targets for countries around the world. These goals, among others, include achieving "universal primary education" to "ensure that all boys and girls complete a full course of primary schooling"—a target that is often measured through primary school enrolment, completion, and the literacy rate among 15 to 24 year-olds. With these imperatives, a dramatic growth in enrolment along with higher level of school attainment has been phenomenal across the world. The motif behind this was also that achieving universal school attainment would lead students in learning the desired knowledge and skills that in turn generates employability of youth, raise national productivity and reduce inequality.

Despite the success in school attainment, the experience of many developing countries is that children do not master basic literacy and numeracy even after four or five years of schooling. As an instance to it, UNESCO (2012) showed that as many as 130 million children including those who spent at least four years in school were found unable to read, write and count, failing to meet the minimum learning standard. Given the context, it becomes evident that only the access to schooling and attending it more years, though necessary, is not sufficient condition for ensuring learning the desired knowledge and skills. For this to happen there has to be a significant gain in knowledge and skills in the forms of literacy numeracy, scientific knowledge, thinking and reasoning ability that generates gains in terms of employment and productivity (Hanushek & Woessmann, 2008).

Being cognizant of the importance of knowledge and skills learned in school for social, economic and personal development; what and how much students learn are policy concerns for various reasons, ranging from ensuring human rights to reducing inequality and improving individual life outcomes, competitiveness, economic growth, and development outcomes. Research evidences suggest that returns to learning may be even higher in developing countries than in developed countries. Therefore, understanding what and how students are learning is also important concern for all parents, teachers and general public to know how well the school education systems equip youths with knowledge and skills they need to better their lives, to play a role in building more peaceful and equitable societies (Matsuura, 2004) and to be able to face future challenges. Besides, measurement of these knowledge and skills is also essential to tracking development and assessing the effectiveness of educational policies and practices (Boreman, Hews, Overman & Brown, 2003; Hanushek & Rivkin, 2010).With this concern, measuring and monitoring students learning achievement by means of large scale assessment developed world wide during the last decade of 20th century, which is important not only for

determining existing level of learning and finding the gaps. It also plays crucial role in providing feedback for improving quality of learning.

Understanding the importance of learning for individual development, nation's economic growth and equitable social development, stronger focus is on raising learning achievement for the reason that success of achieving universal schooling lies when it raises the learning standard at least at minimum level. This paper discusses the main concerns for raising the learning achievement at least at minimum level describing how it contributes to other goals such as individual, social, economic development and employability. Before doing so, it also tries to conceptualize what learning achievement is that is relevant for the theme under reference. Succeeding part presents a glimpse of learning achievement level of Nepali students drawing up the evidences from large-scale assessment carried out so far.

Understanding Learning Achievement

A universal definition of the term may not be available because of diverse context and expectation of society as well as demand of job market and national expectation of particular country. However, taking the purpose of schooling into account it can be understood as the success in learning the subject matter taught, acquisition of knowledge and cognitive skills as expected by curriculum or the mastery over the given content knowledge within the curricula. Given the purpose of schooling, it can also be defined in terms of knowledge, cognitive skills, and abilities that students have attained as a result of their involvement in a particular set of educational experiences (Yuba Community College District Academic Senate, 2005). In the similar vein, the World Bank (2011, cited in Naomee & Tithi, 2013) conceives it as the particular knowledge, cognitive skill or behaviour that a student is expected to exhibit after a period of study, which reflect a nation's concern with the level of knowledge acquisition among its student population. Measuring learning achievements provides information on what particular knowledge (cognitive), skill or behaviour (affective) students have gained after the instruction is completed.

Similarly, cognitive skills, the main component of learning achievement constitute the individual capacities to logically combine, analyse and apply informational symbols (Gintis, 1971) which are nurtured throughout the learning process channelled in schooling and influenced by teaching and curriculum. In specific terms, they are reading, writing, mathematics skills, problem solving and communication skills along with vocabulary and background knowledge.

Based on the understanding it can be concluded that learning achievement is the growth and proficiency in knowledge and understanding in subject matter on the one, and levels of cognitive skills in reading, writing, mathematics and science developed through schooling on the other (Student Learning, Student Achievement Task Force, nd.), usually indicated by test score or numerical value. In essence, it is increased academic performance in the form of cognitive skills - what students know and can do in social, economic and political life. Having conceived learning achievement from this perspective, the succeeding section deals with the main concerns for raising learning achievement.

Main Concerns for Raising Learning Achievement

Although schooling has multi-dimensional objectives including transferring social values, developing life skills, loyalty to the nation, its success is judged mainly by their ability to impart basic literacy, numeracy and analytical skills to students (Riddel & Green, 2009). Since, these skills are rewarded in job market for the reason that either they are ready-to-use skills for productivity or indicate potential productivity or trainability on which individuals are selected (Arrow, 1973; Spence, 1973; Thurow,1975). Additionally, these skills also lay strong foundation for individuals to learn throughout their life helping people adapt changes to meet future challenges. Moreover, certain prestige is also associated with these skills so that one can enjoy a dignified life in the society. To enable individuals deserve these ends, students need to learn in school.

Acquiring learning at least at minimum standard is central concern also for reasons that range from ensuring human rights to improving individual life outcomes; raising competitiveness, economic growth, and development outcomes; and reducing inequality. Succeeding paragraphs deals each of the concerns separately.

Concern for job market outcomes

As education acquired in school has been one of the main determinants of job opportunity for a person, ensuring learning to the desired level is the prime concern to prepare individuals for the job market in order to improve lifetime earnings. By means of schooling, individuals acquire a wide array of reading, writing, analytic, reasoning, communication and so on skills to a different degree, which are always rewarded by the job market for two reasons. First, these skills are ready to use in the job or are easily conversable to work specific skills (Becker, 1962). In this sense, these are considered a stock of productive skills and hence enable individuals to employment opportunity for earning. Second, these skills are positional good that indirectly signals productivity of the person as they make individuals more easily trainable at the work place, thereby reducing training cost of the employers (Thurow, 1975).

While these skills are the basic requirement in getting employment, higher level of achievement in these skills is also associated with higher level of earning. As shown by UNESCO (2010), one extra year of schooling increases an individuals' earning by up to 10 percent. Similarly, in the context of OECD countries, a 1 percent rise in literacy scores relative to the international average is seen associated with an eventual 2.5 relative rise in labour productivity and 1 percent rise in GDP per head (Gigure, 2006). Given the benefits of higher level of skills in school graduates, higher level of schooling is urged for the reason that individuals with higher level of schooling have better market opportunities than the people with lower levels of schooling.

While higher level of schooling is strong predictor of higher cognitive ability, mere attaining higher level of schooling has not ensured the learning achievement to the desired extent due to the difference in learning context in school, which has caused wider differences in skills

distribution ultimately leading towards unequal distribution of income. Given instance is evident that it is the quality not the quantity of schooling that explains variation in labour market outcomes between individuals despite having the same level of schooling. Recent studies show that cognitive skills explain substantial part of variation in income variations across individuals. Particularly in the context of developing countries, such wider variations in income level is more pronounced because of the unequal distribution of learning achievement. In Pakistan, for instance, Behrman, Ross & Sabot (2008) estimated that 1 standard deviation increase in cognitive achievement is associated with a 25 percent increase in individual earning. Similarly, the studies conducted in the United States also show a 12 percent increase in earnings for every one standard deviation increase in math test scores (Mulligan 1999; Murnane et al, 2000; Lazear 2003). This impact is thought to increase with work experience; that is, educational attainment may help workers to hire, but it is the recognition of their skill-related performance that may cause their earnings to rise once they are on the job (Altonji & Pierret, 2001).

This being the time of technology led knowledge base economy where every sphere of individual's life is surrounded by flow of written information; mastery over the literacy and numeracy to understand written information and to keep track of the numbers involved in particular tasks have been the necessary qualification for even the simple jobs of production and distribution (Dreze & Sen, 2013). At the most fundamental level, literacy and numeracy constitute a foundation for developing higher order cognitive skills such as analytic reasoning skills and are essential for gaining access to an understanding the specific domains of knowledge. Besides, these skills are also relevant across the range of life contexts, from education through home to work and social life and interaction with public authorities. The capacity to manage and solve problems in technology-rich surroundings – that is, to access, evaluate, analyse and communicate information through the use of digital device and application– is becoming necessity as information communication technology permeates every sphere of social, economic and cultural life including work place. In this context, the individuals who are proficient in the advanced literacy and numeracy skills are likely to succeed to grab the opportunities in labour market for earning and those with lower skills are seen at considerable risk of losing out the opportunity that is competitive (OECD, 2013).The results discussed so far show a strong ground for raising learning achievement in order to promote income equality.

Concerns for achieving economic growth

Apart from the high returns to schooling in labour market outcomes for individuals, abundance of literatures have established a positive relationship between educational attainment level or the quantity of education and growth rates – a relation that is widely accepted in development circles. The relation between education and economic growth can imply even greater gains for society as a whole. Examining the relation between education and economic growth, significant body of works have also established that a higher rates of schooling generates more skilled and productive work force, who in turn increase economy's output of goods and services (Barro & Lee, 2001). So, in a country with an abundance of educated work force, economic growth remains high.

Educated work force with cognitive skills at least at basic level acquires knowledge of adopting new technology, has the capacity to absorb the processes and production function there by promotes growth. At the same time, these knowledge and skills also facilitate the diffusion and transmission of knowledge needed to understand and process new information and to implement new technologies devised by others, again promoting growth. Because of the increased capability from schooling to effectively use production technology, it has been estimated in some countries that each additional year of schooling attained by the labour force has raised average annual GDP growth by 0.37 percent points (UNESCO, 2010).

As the level of cognitive skills of a nation's students has a large effect on its subsequent economic growth rate, increasing the average number of years of schooling attained by the labour force boosts the economy only when increased levels of school attainment boost such cognitive skills. While basic level cognitive skills in school graduates increases the productivity of existing labour force because of the technology assimilation capability, higher level of cognitive ability in them leads to rapid economic growth by increasing the innovative capacity of the economy as it leads to the creation of new technology, products and processes. The studies show that the gains in cognitive skills are thought to occur through the accumulation of benefits to individuals, the increase in rates of invention and innovation, and the introduction of new technologies and improved production methods (Hanushek & Woessmann, 2008)

It is not clear, however, whether years of education lead to economic growth or economically healthy countries tend to prioritize education. Pritchett's (2001) research on the relation between educational attainment and economic growth suggests that the quality of education—not just the quantity—may play a key role. His findings—which suggest that mere schooling without acquisition of cognitive skills does not contribute to increased economic growth, because schooling generates higher wages while not generating higher productivity or skills—are seen as a mandate to improve the learning achievements. Inferring the context of Latin American countries, Hansushek & Woessmann (2012) noted that the countries are lagging behind East Asia, Middle East and North American (MENA) countries in economic growth because of lower cognitive abilities in students of the region despite of its relatively equal or higher-level school attainment.

Indeed, new research on the relation between higher learning achievement and growth suggests that years of schooling may be a less important contributing factor to economic growth than the quality of education, as represented by scores on international assessments (Lee & Lee 1995; Hanushek & Kimko 2000; Barro 2001). Using cross-country data from 1960 to 1990, Hanushek & Kimko (2000) examine what they call "the quality of the labour force," as measured by math and science scores. They find that a one standard deviation difference in test scores is associated with a 1 percent difference in annual growth rates of per capita GDP. As this added growth compounds, it can lead to large increases in national income (Hanushek, 2004).

Using data on 15 countries in the Organization for Economic Cooperation and

Development (OECD) from the International Adult Literacy Survey (IALS), Coulombe & Tremblay (2006) confirm these findings. They conclude that the higher learning achievement in schools, as expressed by student test scores, is more important for overall economic growth than years of schooling and that returns for improving literacy skills are higher for women than for men. They argue that improving the overall literacy skills of society has a greater effect on growth than does concentrating on developing highly educated elite.

While looking at the macro level, students' learning achievements, especially in language, mathematics and science, have been found to have a significant effect on economic growth. For instance, Hanushek & Woessmann (2008) estimated that a rise of 1 percent standard deviation in student test scores on international assessment of literacy and mathematics is associated with 2 percent increase in annual growth of per capita GDP. More recently, the OECD study noted that raise in Program for International Student Assessment (PISA) test scores may have very large impacts on the future wellbeing of countries by dramatically improving national labour force skills. It is estimated that bringing all OECD countries up to the average performance of Finland, the top performer on PISA, would boost aggregate OECD GDP by US$ 260 trillion– six times the current GDP of OECD countries (OECD, 2010). These all studies, thus, emphasize that the quality of learning outcomes– gain in cognitive ability not the quantity or length of schooling – makes the difference (OECD, 2010).

Concerns for human right

All the Declarations, Covenants, Conventions from the Universal Declaration of Human Rights 1948, International Covenant on Economic, Social and Cultural Rights 1966 and to UN Convention on the Rights of the Child 1989, which are also binding international laws, recognize education as a fundamental human right. Reasserted in the Jomtien and Dakar Declarations, this right has since been incorporated into most national constitutions.

These Declarations, Covenants and Conventions urge to ensure the right to free and compulsory education for all for the “the development of the child’s personality, talents and mental and physical abilities to their fullest potential” (Article 29). Similarly, the United Nations considers education a prerequisite for exercising other civil, political, economic, and social rights, viewing it as “the primary vehicle by which economically and socially marginalized adults and children can lift themselves out of poverty and obtain the means to participate fully in their communities” (UNESCO, 1999).To achieve these end results, the EFA and MDG goals set the target for all countries to ensure all children complete primary school.

The central concern behind these urges for ensuring universal completion of primary schooling is learning the desirable knowledge and skills needed to live a decent life. So, universal completion of primary school has always been only a means to the actual goal of universal education. Every youth should make the transition to adulthood equipped with the minimal set of competencies—including both cognitive and non-cognitive skills—needed to function adequately

in the economic, social, and political spheres of a modern society. The recent *World Development Report 2007: Development and the Next Generation* builds on this notion of childhood and youth as a time to prepare for transitions for which schooling has to improve the skills of young people for work and life—making education opportunities more relevant to the needs of young people as future workers, parents and citizens (World Bank, 2006). To enable children attain basic level of skills and knowledge required to live productively in the life, the minimum threshold for universal primary cycle to complete, usually of 5 or 6 years has also been set. The underlying rationale for setting the 5-6 years of schooling goals has always been broad learning goals that includes a set of minimally adequate knowledge, skills, attitudes, values, behaviours to which could broadly be called the "competencies", to be acquired through schooling. The duration and curriculum of primary or basic schooling were then set so that completion of the cycle with at least some mastery of the curriculum implied acquisition of the universally necessary competencies.

To put it another way, universal completion of primary education – the output goal is a means to the end results – the outcome goal of universal competencies. The implicit assumption in the EFA and MDG is that attaining full course of primary schooling would lead the average and even low performing students reach the threshold of basic skills. There is little basis for working towards universal primary schooling if students emerge from the schooling cycle without an adequate education. So, ensuring a child's right to education implies guaranteeing all students an equal opportunity to learn going beyond simply providing access to schools.

Concerns for social development

Another concern for raising learning achievement is accruing wider social benefits as educational attainment and learning are seen tied to a number of development outcomes beyond individual incomes. Education has been shown to affect numerous social development outcomes--reduction in maternal and infant mortality, lower fertility, prevention of unwanted pregnancy, delayed marriage, increased civil participation in community development, lower crimes in society and reduction in violent and risky behaviors (UNESCO, 2014; Bledsoe et al., 1999). These social returns to education thus exceed the private returns.

Additionally, researchers have also established the relation between a variety of health and well-being outcomes as well as educational attainment and learning outcomes. Higher reading and math scores are associated with lower fertility rates in Ghana (Oliver, 1999) and South Africa (Thomas, 1999). As with individual economic returns, on which test scores show an even stronger impact than mere educational attainment, cognitive skills have stronger effects on the number of children per household than do mere years of schooling. In Africa, education has also been associated with lower prevalence of HIV and improved reproductive health, among both men and women (UNESCO, 2004).

A mother's education also has a strong impact on her child's health. The link between a mother's years of schooling and her children's health is well established empirically (Behrman,

1996; Strauss & Thomas, 1998; Cutler & Lleras-Muney, 2006). Although it is not clear which aspects of education account for this relation, studies from developing countries have associated it with health knowledge and math scores (Glewwe, 2002).

Apart from these wider social benefits, learning achievement at least at basic level at par with the grade level to a higher level is prerequisite to attain higher level of education. Particularly, improved achievement at early grades results in higher-level participation in further grades that ultimately leads towards abundance stock of human capital in the society. As is evident in the developing countries that higher rate of drop out and repetition at early grades due to the low learning achievement has resulted lower participation in upper grades which in turn has led to the short falls of educated human resources. There is a demonstrable link between early performance in school and subsequent rates of high school graduation, college attendance and completion, and ultimately earnings. While this does not mean that individual students who perform poorly early on cannot improve their performance and subsequent outcomes, the pattern of success leading to success is strong.

The discussion so far shows that the schooling practices that enhance student achievement in the form of cognitive skills have a larger effect in long run. So getting children into school and letting them thorough it is not an end in itself but only a means of delivering these skills. So, success or failures in achieving education for all lies not just on children attain more years of schooling but on what and how well they learn in school.

Given that the knowledge and skills learnt at school, have a larger impact on labour market, economic and social developmental outcomes; parents, students, teachers, governments and the public – all stakeholders – need to know how well their education systems prepare students for real-life situations and for meeting the future challenges. Measuring such knowledge and skills is also essential to tracking students' development and assessing the effectiveness of educational policies and practices. In response of this need, many countries have embarked primarily on standardized achievement test of large-scale type for assessing student achievement in terms of knowledge and skills that students have acquired as a result of their exposure to schooling. Such national assessments, on the one hand, attempts to describe national levels of learning achievement as a whole at a particular grade level or age, especially in key subject areas, and on the other, compares the achievement levels of key subgroups (such as boys and girls, ethnic groups, urban and rural students, students in different geographical areas, and public and private school students). It also provides evidence that enables the system to monitor the standards of student achievement whether they are rising or falling over time (Greaney & Kellaghan, 2008). Additionally, it provides information on the strengths and gaps in the system based on which measures for further reform can be identified.

Nepal also has been adopting standardized achievement test on large scale for assessing the student achievement at particular grade level since 2011 in order to know whether the students have acquired desired levels of knowledge and skills in reading literacy, mathematical

literacy and scientific literacy that are needed to succeed in the 21st century society and economy. The focus of these achievement tests have been the students of grade 3, 5 and 8. The published result of three achievement tests, two times for grade 8 and one for grade 3 and 5 so far, have depicted existing learning level of these students. Drawing upon the results of National Assessment of Student Achievement (NASA) 2011 and 2013 for grade 8 and 2012 for grade 3 and 5 (ERO, 2013; 2015), succeeding paragraphs present the learning achievement of Nepalese students.

Learning Achievement Level of Nepalese Students

Higher learning achievements of students reflects not only the efficiency of schools and quality of teaching and learning but also is one of the important indicators of a wide range of skills, abilities and knowledge that workforce of a nation hold to contribute for national development. Since, the availability of human knowledge and skills is critical in determining a country's rate of economic development and its competitiveness in an international market. Though high or low is a relative concept, regarding the achievement of students, the mostly used indicator of student achievement is the test score obtained out of 100 by students in a standardized test. It means that the more the student obtains out of 100, the higher is the achievement level. The results in the achievement tests are reported mainly as percentage of maximum marks where 100 percentages represent all task solved and the 0 percentage represents none.

While analysing the achievement level of Nepalese students on Nepali language, Mathematics and Science based on the achievement tests conducted so far; poor performance, wider inequality among the students, unbalanced learning across the curricula contents and low ability in higher order skills, among others, can be observed. Using the test scores of those assessments, each of them is described hereafter.

Low level of learning

While the national average score, depending upon the subjects, ranges from 35 to 49 for grade 8, 53 to 60 for grade 5 and 60 to 63 for grade 3; there were also as many as 4.35-6.33 percent schools performing below 30 percent whereas only 13-19 percent schools obtained above 80 percent. Similarly, 0.2 to 4 percent students studying in primary grades were not able to solve single test item. Regarding the grade 8, the average achievement of students in the lowest performing schools was below 15% while the students from the highest performing schools gained, on average, over 90 percent. The average score in a school is below 15 percent means that in those schools many students fall below that level. The national average in each subject for both the grades shows that students are comparatively poorer in Mathematics and English than in Nepali.

Table 1. Learning achievement by gender

Gender	Grade 3 (2012)		Grade 5 (2012)			Grade 8 (2011)			Grade 8 (2013)		
	Nepali	Math	Nepali	Math	English	Nepali	Math	Social Study	Nepali	Math	Science
Girls	64	60	61	53	53	49	41	49	48	33	39
Boys	63	60	59	54	55	48	45	50	48	38	43
National	**63**	**60**	**60**	**53**	**54**	**49**	**43**	**49**	**48**	**35**	**41**

Source: ERO, 2015a

Above data is also evident that in comparison to lower grades, the achievement at the upper grade in all the subjects is falling down with remarkable differences. On this ground, one can infer that mere higher level of schooling is not the indicative of higher level achievement. This level of achievements of 8th graders which is below than 50 percent for all the subjects in both years, is low which means that our system is not efficient enough to deliver the expected learning outcomes. Moreover, it is also noticed that the delivery system is getting deteriorated gradually than it is in the previous years as the achievement is going further down instead of improving. Most frustrating fact is that about 15 percent of the sample population of students is found to have reached the grade 8 without basic proficiency even in the basic literacy of minimum level.

While comparing the achievement level of Nepalese students with that of international students of the same level, Nepali students lag far behind with their international colleagues in terms of their proficiency in each subject. The table below the average performance level of Nepali students in comparison to international mean (0).

Table 2. Average performance level of Nepali students

Subjects	Grade 8 (2011)	Grade 8 (2013)	Grade 5	Grade 3
Nepali	-0.87	NA	-1.58	-1.78
Math	0.25	-0.5	0.08	-0.88
Science	NA	-0.8	NA	NA
English	NA	NA	-1.48	NA

Source: ERO. 2013, 2015a, 2015b, 2015c

Compared the data set of Mathematics with the international standard of it against Trend in International Mathematics and Science Studies (TIMSS), Nepali students are, on an average, one year behind the international average where the 5th graders are somewhat at the level of grade 4 students and 3rd graders at the level of grade 2. Even for Nepali and English, the average reading proficiency of grade 5 students is much lower than the average international standards of grade 4 through Progress in International Reading Literacy Study (PIRLS). In Nepali and English, the 5th graders are far below even the grade 4 at the level of grade 3 in comparisons

to international averages. Similarly, the estimated level of 3rd graders is round -1.78 standards units lower than the international mean which means they are at the level just above grade 1 students of the international level. The same kind of results are found also in grade 8 in NASA 2011 where Nepali students are found -0.25 below in Mathematics and -0.87 below in Nepali than the international average performance. This level of performance is estimated to be the level of grade 7 in international standard. Most distressing is the fact that Nepali students are falling further down in mathematics from around 0.25 point above average performance in 2011 down to the level -0.5 in 2013. Given context implies that Nepali students are likely to fall behind their international colleagues to compete in the global labour market.

An unbalanced learning across the curricula contents and cognitive domains

Apart from the low level of learning, wider variances in learning achievement across the curricula contents within the subjects in each and across the cognitive domains are also evident from these assessment results. Against the expectation of curricula, the results are evident that certain contents of the curricula are learnt less effectively than others. For instances, in Mathematics, the achievement level in Algebra and Numeracy is remarkably lower than Arithmetic and Geometry. More or less similar is the situation in grade 8 too where students 'performance is lower in those areas in comparison to others. In Nepali and English, the Reading and Writing skills are poorer in comparison to the achievement in Vocabulary. Such unbalanced learning across the contents is also visible in other subjects. For instances, the Biology in Science and Politics in Social Studies are seen less learnt area than other areas. The content wise learning achievement level as presented in the table shows that unequal level of leaning is persistent for all the grades and subjects.

Table 3. Learning achievement in various content areas by subjects

Subject	Grade	Year	Reading	Writing	Grammar	Vocabulary		
Nepali	8	2011	56	48	51	43		
	8	2013	46	46	49	55		
	5	2012	56	58	64	70		
	3	2012	64	54	65	60		
English	5	2012	50	49	57	58		
Science	8	2013	Biology	Chemistry	Physics	Astrology & Geology		
			34	43	43	43		
Social Studies	8	2011	Geography	Civics	Economics	History	Politics	
			57	51	58	50	46	
Math	8	2011	Algebra	Geometry	Arithmetic	Numeracy	Sets	Stat
			48	37	49		38	48
		2013	27	34	37		37	38
	5	2012	49	57	54	44		
	3	2012	40	60	61	54		

Source: ERO, 2015a

As confirmed by the results, reading proficiency of Nepali students is poor, which has not improved over the years. The low level of reading proficiency among the students has kept them in a weak state of comprehending implied meaning, solving complex problem, abstracting of deeper ideas, producing open-ended text not only in Nepali subjects but also in other subjects too, leaving them less able to perform the tasks demanding higher cognitive ability. Their performance is found comparatively poorer in the ability to solve problems, to analyse, deduce logic, generalize, justify an argument or viewpoint, and in the ability to transfer learning from one context to another. Circumscribed with such unbalance learning on some domains, the entire system is shifted towards a low performing making it less effective to yield better results.

A remarkably high number of students in primary grades were able to solve only 35-40 percent practical problems of Nepali and Mathematics. Of the total, 4-20 percent of the students in Mathematics, 3-18 percent in Nepali, and 18 percent in English could not solve any of the problems requiring higher ability. In Mathematics, students are able to do basic calculations, but are weak in reasoning, problem solving, proving theory or formula, and in constructing shapes and figures. In many cases, the students did not even attempt to complete the open-ended questions of higher cognitive level. Also they are much weaker in producing fluent texts or essays, and in preparing synthesis and abstracts from a text. In case of applying the gained knowledge in new situation, they are also found very poor. From the table below, one notices a more discouraging fact that the higher the grade, the lower is the ability except in Nepali.

Table 4. Performance level by cognitive domains

Cognitive Domains	Grade 3		Grade 5			Grade 8					
	2012					2011			2013		
	Nepali	Math	Nepali	Math	English	Nepali	Math	Social Study	Nepali	Math	Science
Knowledge	72	59	71	65	64	74	68	52	57	46	49
Comprehension	65	64	63	59	50	45	55	66	56	42	49
Application	56	53	61	52	49	46	38	40	45	35	32
Higher Ability	37		47	40	35	42	21	34	48	25	23

Source: ERO, 2015a

Although cognitive skills are seldom taught explicitly in schools, various researches indicate that schooling through teaching knowledge and skills in language, mathematics and science need to promote cognitive ability. As cognitive skills and academic performance complement each other, the success of schooling is also measured in terms of cognitive skills. In this regard, our system is somehow weak in achieving fundamental goal of education--to equip students to think critically, solve complex problems and succeed in the society and economy of 21st century (Fin et al, 2014).

Inequality in learning level between regions, locations and types of school

While looking at the results from equality perspectives, the data sets of all grades confirm that there is a wider inequality between ecological belts, development regions and districts for students' opportunity to reach an adequate level of learning in all subjects. The table below shows that the students from the Mountain and Kathmandu valley are far ahead of all and the students from the Tarai rank the lowest in most of the subjects. In terms of Development region, the Western region outperforms the rest where Eastern region has achieved the lowest with a significant difference. Similarly, a rural and urban variation in achievement level is also noticeable. This level of inequality in achievement level indicates not only the disparity in learning outcomes among students but also an enduring social inequality.

Table 5. Learning achievement by regions and locations

Regions/ Locations	Grade 3		Grade 5			Grade 8					
	2012		2012			2011			2013		
	Nepali	Math	Nepali	Math	English	Nepali	Math	Social Study	Nepali	Math	Science
Mountain	61	60	61	56	51	49	41	48	47	21	32
Hill	60	55	57	49	49	50	38	47	39	29	38
Valley	80	77	78	68	79	64	55	59	66	53	56
Tarai	57	55	53	49	46	42	44	48	37	27	32
Eastern	54	52	52	45	39	43	33	46	36	25	32
Central	59	56	54	51	49	49	50	48	47	30	34
Western	67	64	63	57	59	52	45	49	47	32	41
Mid-Western	54	46	52	45	44	43	39	48	41	22	34
Far-Western	59	58	58	51	49	42	37	48	38	29	32
Urban	75	71	74	65	73	55	48	52	57	52	51
Rural	59	56	56	50	48	46	41	48	45	28	37

Source: ERO, 2015a

While looking at the district wise results, similar level of differences is seen between the districts to give equal opportunities in reaching the preset educational goals. The average achievement in the Kathmandu Valley is very high in comparison with the lowest performing districts. In primary grades, differences in the mean scores between the lowest and highest scoring districts vary from 33 percent points (in Math's grade 5 and Nepali grade 3) to 43 percent points (in Nepali grade 5) and ultimately to 51 percent points in English. In English, the difference is found to be connected with the proportion of institutional schools in the district as the medium of language for instruction in most of the private schools is English which has helped enhance the achievement. In other subjects, as well the results are seen to be the higher in the private schools. The results of some districts showed higher achievement without including any institutional

schools in the sample. The more crucial reality is that the average achievement is very low in the lowest performing districts, which concentrate in comparison to some other parts of the country. It is not known how much lower achievement is there in some other districts, which have not been selected as the sample.

The variations in learning achievements of 8th graders are also great among the districts. In the lowest performing 7 districts the average achievement of math is below 25 ranging from below 12 to 24, followed by other 13 districts having the highest achievement 33 and the lowest 28. In another extreme the highest performing districts from Valley range from 51 to 59. More or less the same is the situation for science too. In Nepali the difference is wider ranging from the lowest 29 to highest 68. From the equality viewpoint, this is not a positive sign both for the system and the society at large in achieving social justice and equity.

Another level of inequality in achievement is seen between the types of schools-community and institutional schools. As shown in the table below the difference is the wider ranging from 20 to 36 percent points.

Table 6. Learning achievement by school type

Type of School	Grade 3		Grade 5			Grade 8					
	2012		2012			2011			2013		
	Nepali	Math	Nepali	Math	English	Nepali	Math	Social Study	Nepali	Math	Science
Community	57	54	54	49	66	46	39	46	42	26	34
Institutional	80	75	78	67	83	62	63	63	65	57	57

Source: ERO, 2015a

This level of difference is obvious because of intake of high achieving students, higher socio economic status and rigorous teaching learning process in the institutional school. In the long run, this will have a remarkable effect to the community schools and inequality in the society. Especially, when the highest achieving students are picked away from the community schools, the average achievement level is automatically reduced and the remaining students do not find competitive environment to learn. This will ultimately cause little prospects for the students within the community schools because of harsh competition for the study of upper level. When only the less-motivated students from the family with low-socioeconomic status remain in the community schools, the burden will come to the teachers to motivate in learning and to upgrade the low-achievers to higher ones. Ultimately, the trend will cause the inequality between the citizens with high and low socio-economic status which will, in effect, lead to conflicts between the different groups in society.

Given situation of low achievement, unequal level of learning in some contents and domains, and inequality in learning opportunity between regions, districts and locations call for some reform measures immediately that have long-term impacts. These reform measures would include establishing well-defined goals, setting and assessing performance standards, promoting reading ability and improving assessment system.

Conclusion

This paper discussed the main concerns for raising students' learning achievements in the form of improved cognitive skills to the desired level. Though more years of schooling is believed to lead to long term improvements in cognitive ability, the mere schooling is not enough to ensure the expected skills to develop. It is only a means to develop them in systematic and organized way through teaching the prescribed subjects. For this to happen, schooling needs to lead effective learning in students. So the target set under both the EFA movement and MDG to achieve universal schooling can be achieved only when the students acquire the desired level of knowledge and skills in the form of reading, mathematical and scientific literacy, and technological know-how. They are the basic requirements to live a life in modern society and economy. This basic level of knowledge and skills are fundamental not only for generating individual employability, rising income and sustaining economic productivity but also are critically significant for the development of the child's personality, talents and mental and physical abilities to their fullest potential. In Dreze & Sen's (2013) terms, they are essential for individuals to have freedoms to understand the world, to lead an informed life, to communicate with others, and to be generally in touch with what is going on, ultimately to be free from any kind of incarceration.

Besides ensuring basic level of knowledge and skills for the masses, the students need raising higher-level achievement to produce the pool of human resources capable of technological diffusion and innovation to further accelerating the growth and development of country. If the country has more workers who have at least basic skills, the economic growth sustains as they can make the use of the technology developed elsewhere. The economic growth further spurs when there are pool of highly skilled human resources because of their capability to adapt new technology in new situation. They can diffuse, innovate technology and further generate intellectual prosperity by developing basic skills to adopt technology. It is the precondition to sustain economy and developing higher skills are essential for technological advancement and innovation to spur the growth and development. Furthermore, as they both complement and reinforce each other, achieving basic literacy for all may well be a precondition for identifying those who can gain higher skill or dignified status. In other words, tournaments among a large pool of students with basic skills may be an efficient way to obtain a large share of high-performers.

While looking at learning achievement level of Nepalese students on these grounds, the schooling system is seen less adequate to develop the skills even at the basic level in majority of students, as low achievement is apparent. As shown, Nepalese students lag behind the international standard; it is also likely to fall behind in taking competitive advantage in the global market.

Given the large disparities in learning outcomes between location and types of school one attends, region and area one resides; the economic, social, and political inequalities is also likely to be perpetuating for some years to come. If higher learning achievement do indeed raise personal income, spur overall economic growth, and raise social indicators, a strong argument can be made that provides equal level of learning opportunity to all can contribute to reduce social and economic inequalities in the country.

References

Altonji, J. G. & C. R. Pierret. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics, 116 (1), 313–50.*

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics, 2(3), 193-216.*

Barro, R. J. (2001). Human capital and growth. *American Economic Review Papers and Proceedings 91 (2), 12–17*.

Barro, R.J. & Lee, J. W. (2001). International data on educational attainment: Updates and implications. *Oxford Economic papers, 3, 541-563.*

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *The Journal of Political Economy, 70, 9-49.*

Behrman, J. (1996). The impact of health and nutrition on education. *World Bank Research Observer, 11 (1), 23–37.*

Behrman, J. R., Ross, D. & Sabot, R. (2008). Improving the quality versus increasing the quality of schooling: Evidence for rural Pakistan. *Journal of Development Economics, 85, 94-104.*

Bledsoe, C. H., Johnson-Kuhn, J. A. & Hagga, J. G. (1999). Introduction. In: C.H. Bledsoe, J. B. Casterline, J. A. Johnson-Kuhn & J.G. Haaga (Eds), *Critical Perspectives on Schooling and Fertility in Developing World, 1-22.* Washington DC: National Academy Press.

Borman, G. D., Hewes, G. M., Overman, L. T. & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73(2), 125-230.*

Coulombe, S. & Tremblay, J. F. (2006). Literacy and growth. *Topics in Macroeconomics 6 (2).* Berkeley Electronic Press. ttp://www.bepress.com/bejm/topics/vol6/iss2/art4/

Cutler, D. & Lleras-Muney, A. (2006). Education and health: Evaluating theories and evidence. *NBER Working Paper 12352.* Cambridge, MA: National Bureau of Economic Research.

Dreze, J. & Sen, A. (2013). *An Uncertain Glory: India and Its Contradictions.* New Delhi: Allen Lane an Imprint of Penguin Books.

Education Review Office. (2013). *Where Are We Now? Student Achievement in Mathematics, Nepali and Social Studies in 2011*. Kathmandu: Author

-------- (2015a). *National Assessment of Student Achievement: A Glimpse of Results*. Sanothimi: Author.

--------(2015b). *Report on National Assessment of Student Achievement, 2013 (Grade 8: Mathematics, Nepali and Science).* Sanothimi: Author

--------(2015c). *National Assessment of Student Achievement (NASA) 2012 (Grade 3 and 5).* Sanothimi: Author.

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., Sheridan, M. A., Gabrieli, C. F. O & Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science, 25(3), 736-744*.

Giguere, S. (2006). An introduction to skills upgrading: Why a shift in policy is needed. In: OECD, *Skills Upgrading: New Policy Perspectives, 23-32*. Paris: OECD Publishing.

Gintis, H. (1971). Education, technology and the characteristics of worker productivity. *The American Economic Review, 61(2), 266-279.*

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature 40 (2), 436–82.*

Greaney, V. & Kellaghan, T. (2008). *National Assessment of Educational Assessment: Assessing National Achievement Levels in Education, Vol 1.* Washington DC: World Bank.

Hanushek, E. A. (2004). Some simple analytics of school quality. *NBER Working Paper 10229.* Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A. & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100(2), 267-271.*

Hanushek, E. A. & D. D. Kimko. (2000). Schooling, labour-force quality, and the growth of nations. *American Economic Review*, *90 (5)*, 1184–1208.

Hanushek, E.A. & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46(3), 607-668.*

--------(2012). Schooling, educational achievement and Latin American growth puzzle. *Journal of Development Economics,99, 497-512.*

Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review 10 (3), 179–214.*

Lee, D. W. & T. H. Lee. (1995). Human capital and economic growth: Tests based on the international evaluation of educational achievement. *Economics Letters, 47 (2), 219–25.*

Matsuura, K. (2004). Forword. In: *UNESCO, EFA Global Monitoring Report 2005. Education for All: The Quality Imperative*.Paris: UNESCO.

Mulligan, C. B. (1999). Galton versus the human capital approach to inheritance.*Journal of Political Economy 107 (6)*, S184–S224.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y. & Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management 19 (4), 547–68*.

Naomee, I. & Tithi, U. M. (2013). Reflection of Bloom's taxonomy in the learning outcomes of secondary social science curriculum of Bangladesh. *International Journal of Science and Research, 2(2), 550-559.*

OECD. (2010). *Education at a Glance 2010: OECD Indicators*. Paris: Author.

-------- (2013). *OECD Skills Outlook 2013: First Result from the Survey of Adult Skills*. Paris: OECD Publishing House. http://dx.doiorg./10.1787/9789264204256-en

Oliver, R. (1999). Fertility and women's schooling in Ghana. In: P. Glewwe (Ed), *The Economics of School Quality Investments in Developing Countries, 327–44*. New York: St. Martin's.

Patrick, John J. (1991). Student achievement in core subjects of school curriculum. *ERIC Digest. ED 332930*. Retrieved from: www.ericdigests.org/pre9219/core.htm*on* May 15, 2015.

Pritchett, L. (2001). Where has all the education gone*? World Bank Economic Review, 15 (3), 367–91.*

Riddel, W. C. & Green, D. A. (2009). Understanding educational impacts: The role of cognitive skills. Retrieved from: www.researchgate.net/…/255646496.

Rolleston, C. & James, Z. (2014). Schooling and cognitive outcomes from childhood to youth: A longitudinal analysis. In: M.Bourdillon & J. Boyden (Eds), *Growing up in Poverty: Findings from Young Lives, 117-138*. Hamisphire UK/New York: Palgrave Macmillan.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics* 87, 355–374.

Strauss, J. & Thomas, D. (1998). Health, nutrition, and economic development. *Journal of Economic Literature 36 (2), 766–817.*

Student Learning, Student Achievement Task Force. (nd).Student learning, student achievement: How do teachers measure up? Arlington VA: National Board for Professional Teaching Standard.

Thomas, D. (1999). Fertility, education and resources in South Africa. In: C. H. Bledsoe, J. B. Casterline, J. A. Johnson-Kuhn & J. G. Haaga (Eds), *Critical Perspectives on Schooling and Fertility in Developing World*, *138-180*. Washington, DC: National Academy Press.

Thurow, L. (1975). *Generating Inequality*. New York: Basic Books.

UNESCO. (1999). Right to education: scope and implementation. http://portal.unesco.org/education/en/file_download.php/c144c1a8d6a75ae8dc55ac385f58102erighteduc.pdf

-------- (2004*). EFA Global Monitoring Report 2005. Education for All: The Quality Imperative*. Paris: Author.

-------- (2010). *Education Counts Towards the Millennium Development Goals*. Paris: Author

……(2010). *EFA Global Monitoring Report 2010. Education for All. Reaching the Marginalized*. Paris: UNESCO/Oxford University Press.

-------- (2012). *EFA Global Monitoring Report 2012. Education for All. Youths and Skills: Putting Education to Work*. Paris: Author.

-------- (2014). *EFA Global Monitoring Report. Education for All. Teaching and Learning: Achieving Quality for All*. Paris: Author.

World Bank. (2006). *World Development Report 2007: Development and the Next Generation*. Washington, DC: Author.

Yuba Community College District Academic Senate. (2005). Student learning outcomes. Retrieved from: http/www.imt.liu/se/edu/Bologna/LO/slo.pdf.on 2015/10/5.

Comparing the Achievement of Mathematics for Grade 3 and 5*

Education Review Office

Abstract

Drawing entirely upon the data base of National Assessment of Student Achievement (NASA) 2012 for 3rd and 5th graders in Mathematics, this paper compares students' learning achievement considering various diversity and equality factors such as ecological belts, development regions, districts, types and location of schools, caste/ethnicity and language of students. Gender perspective crosscuts across the analysis. While doing so, an attempt has been made to analyse the results comparatively on the basis of latent theta (θ) produced by IRT modelling and presents the comparable scenario of students' proficiency in Mathematics for both grades. Not only the 5th graders have comparatively higher level knowledge, comprehending capabilities and skills of solving higher level tasks than the 3rd graders, their achievement is also higher while assessing their respective curricular competencies. The first part of the paper compresses basic results, the second part describes achievement in terms of diversity factors followed by summary and conclusion.

***Keywords**:* assessment, cognitive skills, latent ability, learning achievement, Mathematics.

Introduction

Achievement in Mathematics of 3rd and 5th graders are assessed systematically and frequently in Nepal since the mid 1990s. It was started during the implementation of the Basic and Primary Education Project (BPEP), though the results of the previous National Assessments (see BPEP, 1995;1997; 1998; CERID, 1998; EDSC, 1997; 1999; 2001;2003;2008; CERSOD, 2001; Fulbright, 2008) are not fully comparable in absolute sense with each other because of the missing linking procedure between the tests. In NASA 2012, a linking procedure is created between grade 3 and 5 tests by using the principles of Item Response Theory (IRT) modelling. The latent ability (theta, θ) was used in the comparison with equating three versions of test items for both grades, calibrating all of them with each other with Trends in International Mathematics and Science Study (TIMSS) scale in which TIMSS average was set as theta θ = 0, and the latent theta was used to make a comparison between grade 3 and 5 results.

** This paper is the part of the report "National Assessment of Student Achievement (NASA) 2012 (Grade 3 and 5)" prepared by Shyam Prasad Acharya, Section Officer, ERO and Jari Metsämuuronen (PhD), Senior Evaluation Specialist, Finnish Education Evaluation Centre, which is reorganised in this form by Gopal Prasad Bhattarai Under Secretary, ERO.*

Although a separate analysis of grade 3 and 5 Mathematics has already been published, it has not compared the percentage of the raw scores (See ERO, 2015). In this paper, the results are compared on the basis of latent ability (theta) presenting the scenario of students' proficiency in Mathematics. Obviously, the 5th graders have comparatively higher level in knowledge, comprehending capabilities and skills of attempting higher level tasks than the 3rd graders. In this regards, this comparative analysis gives a synopsis of the existing situation of Nepalese education system as to what extent 5th graders are better than 3rd graders. The comparison is made here mainly from the perspective of the quality and equality – gender, caste-ethnicity, language, various geographical regions, and school types and locations. Therefore, the entire variables which were included in the main report of NASA 2012 will not be compared here.

This article first tries to compare basic results of Mathematics by analysing the distribution of student population in terms of achievement score, basic results by contents, types of items, gender, types of schools and cognitive domains. Then it moves to compare the region wise (Ecological and Development regions), district wise and school location wise variations in achievement vis-a-vis the differences between case/ethnicity and language of students.

Methodology

This article has been developed from the secondary data of NASA 2012 generated from the test conducted in 28 districts of Nepal representing all Ecological zones, Development regions including the Valley as a separate stratum. For the methodological questions, the reader is asked to go through the main report of NASA 2012 (ERO, 2015), particularly the chapter 2 for the technical matters of sampling, test instruments and analytical tools which are not handled here.

The test for the assessment was administered in 1690 schools (843 schools for grade 3 and 557 schools for grade 5) among 19252 students in grade 3 and 13714 students in grade 5 using random sampling method representing each stratum. The test was administered in three versions of standardized test items using linking items in each set. Using IRT modelling, items were calibrated and the versions of test items were made comparable by equating the scores. Use of three versions of test items facilitate for wider coverage of curricular contents as well as for ensuring high reliability in test administration. Standardized items were developed covering the learning objectives of curriculum of respective grades which were pilot tested in order to test their difficulty levels and were finalized and selected having appropriate difficulty level. Teachers, subject specialists and professionals participated in the process of test construction and item selection. Validity of the test items were ensured by analysing and covering of curricular contents. Some linking items from an international assessment TIMSS were calibrated to compare the results with the international assessment. Data were tabulated using Optical Mark Reading (OMR) sheet and analysed using One Parameter Logistic Model (OPLM). The overall assessment results of grade 3 and 5 Mathematics as well as some

disaggregated results were compared. Basic comparison is done by calculating latent ability of students (Theta, θ) and also calculated effect size[1].

Comparison of Basic Results

Altogether 32,966 students (51 percent boys and 49 percent girls) participated in the Mathematics test (13,714 from grade 5 and 19,252 from grade 3) from 28 districts. The latent ability of both grades was calibrated based on grade 4 TIMSS scale (θ = 0). As expected, in the whole population, the grade 5 students are better (average θ = 0.159) than grade 3 (average θ = – 0.484) and the difference is notable i.e., 0.642 standard units which is presented in figure 1 below.

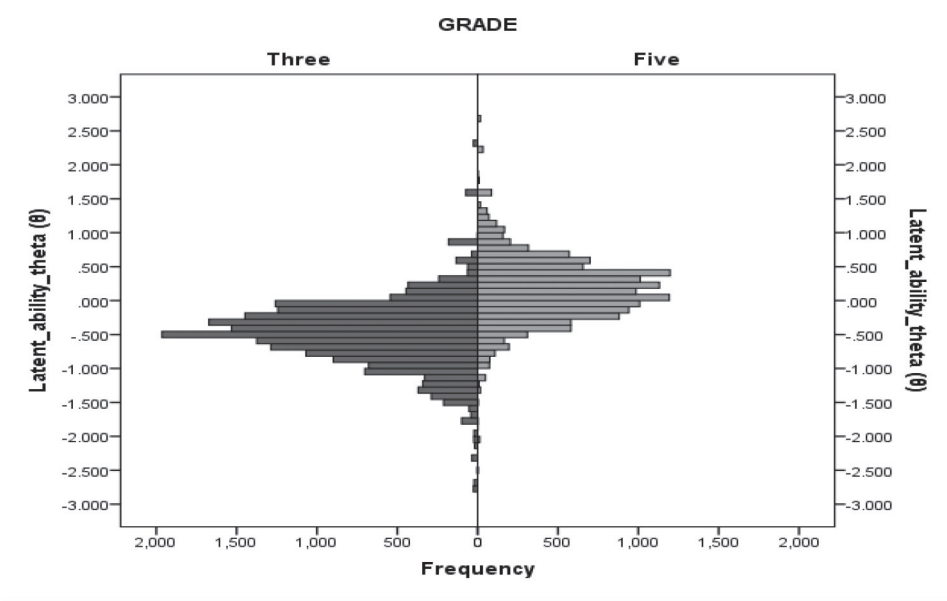

Figure 1. Distribution of the students' achievements by grade

Figure 1 compares the latent ability (Theta value=θ) for 3rd and 5th graders and shows that the students at grade 3 and 5 are distributed normally based on latent ability, but the grade 5 students are concentrated more on higher value of Theta than the students of grade 3. For the comparison of the latent ability, theta, corresponding to each grade, was grouped into ten groups based on percentile shown in table 1.

1 *Cohen's d and f (Cohen, 1988)*

Table 1. Comparison of the percentiles of latent ability in Mathematics

Percentiles	Grade 3	Grade 5
	Average (θ)	Average (θ)
10	-1.086	-0.402
20	-0.856	-0.213
30	-0.688	-0.074
40	-0.564	0.033
50	-0.461	0.157
60	-0.36	0.26
70	-0.254	0.382
80	-0.126	0.519
90	0.049	0.707

The table indicates that few (10%) of the students from grade 3 and most of the students from grade 5 (70%) show better ability ($\theta > 0$) compared to the international TIMSS average of grade 4 ($\theta = 0$) value. However, a large number of students from grade 5 (30%) are still below TIMSS grade 4 ability. These figures also indicate that a big number of students are below the average in Mathematics proficiency to their international colleagues. This situation calls for an extra efforts in pedagogical processes to raise the learning ability of large number of students especially who are too below the average.

Gender-wise Comparison of Results

Gender-wise comparison is also made under various headings below. However, the total comparison is made here as it reflects the whole scenario of population. Gender-wise achievement scores are presented in the table 2.

Table 2. Gender-wise comparison of achievement

Gender	Grade 3					Grade 5			
	N	Mean (θ)	SD	CV		N	Mean (θ)	SD	CV
Boys	8,670	-0.49	0.49	100.0		6,570	0.18	0.47	38.3
Girls	8,304	-0.48	0.50	96.0		6,302	0.15	0.49	31.6
Total	**16,974**	**-0.48**	**0.50**	**96.0**		**12,872**	**0.17**	**0.48**	**35.4**

SD = Standard Deviation, CV = Coefficient of Variance

Overall, the gap between girls' and boys' achievement is not significant in 3rd graders ($p > 0.05$) where as it is significant ($p < 0.001$) in 5th graders. However, the gap is very low ($\eta^2 = 0.001$) and the gender explains only 0.1% of the ability. It is seen that boys are slightly better in

Mathematics skills only in grade 5. Datasets from NASA 2011 grade 8 also showed that difference in latent ability of girls and boys was significant ($p < 0.001$, $\eta^2 = 0.009$, $f = 0.10$) and, hence, gender explains 0.9% of the student variation (ERO, 2013).[2] This indicates real but small gap in Mathematics skill at the higher grades, the ability gap between the gender is seen to have increased slightly in upper grades.

Above dataset of grade 3 shows that the differences between boys and girls in Mathematics proficiency are not significant, which is a good sign from the equality point of view. The difference in grade 5 is significant but is very small. The tendency shows that the boys are slightly out-performing the girls in Mathematics skills. This gap in latent ability is wide in grade 5 in comparison to grade 3.

School Types and Achievement

Generally, students' performance in institutional schools is better than the students in community schools though there are also high performer students in community school as well.

Table 3. School type and Achievement in Mathematics

School Type	Grade 3				Grade 5			
	N	Mean (θ)	SD	CV[1]	N	Mean (θ)	SD	CV
Community	14,476	-0.58	0.49	118.4	10,657	0.08	0.47	17.0
Institutional	4,775	-0.20	0.43	46.5	3,057	0.44	0.44	100.0
Total	19,251	-0.48	0.51	94.1	13,714	0.16	0.48	33.3

1) *Absolute value*

On the basis of table 3 it is evident that, on average in both grades, the students from the institutional schools perform higher than the students from the community schools. Again, they acquire better mathematical skill when they are in the upper grade. But the gap is narrower in grade 3 (0.36) which is slightly wider in grade 5 ($\theta = 0.38$).[3] It means that the students in institutional school are ahead from the lower grade and they are acquiring more learning skills in grade 5 than the students in community school. The differences for both grades are statistically significant ($p < 0.001$) and the effect size is remarkably high in both grades (Cohen's $d = 0.79$ in grade 3 and $d = 0.77$ in grade 5).

2 Reanalysis of the NASA 2011 dataset shows that the latent ability for boys was $\theta = 0.37$ and for girls $\theta = 0.17$, that is, the difference is 0.20 standard units indicating that the difference does not change much during whole school time.

3 In grade 8, in NASA 2011 dataset, the difference is 1.13 standard units, that is, there seems to be a tendency for the gap to grow within the years.

Table 4. School type and gender-wise Achievement

School Type	Gender	Grade 3					Grade 5				
		N	Mean θ	SD	Difference B-G[1]	Cohen's *d*	N	Mean θ	SD	Difference B-G[1]	Cohen's *d*
Community	Boys	6,514	-0.57	0.49	0.01	0.01	4,935	0.10	0.46	0.03*	0.07
	Girls	6,384	-0.57	0.48			5,016	0.07	0.47		
Institutional	Boys	2,156	-0.23	0.41	0.03*	0.07	1,635	0.43	0.42	0.03*	0.07
	Girls	1,920	-0.20	0.45			1,286	0.46	0.45		

*1) G-girls, B-boys, and * significant*

From the gender perspective (table 4), the difference in ability is significant and slightly higher in the institutional school than the community schools (0.01 and 0.03 in grade 3). The difference is equal in grade 5 in both types of schools (0.03). Though there are differences, the effect size indicates the very small differences in both grades (see Cohen's *d* from table 4).

As a whole, institutional school students outperform the community school and the gap of learning ability widens more in the upper grade. The gap in latent ability between boys and girls is equal in 5th graders in both types of schools. However, in both grades, the discrepancies between girls and boys are pronounced more in the institutional schools than community schools. From the equality view point , it is a positive aspect in the community schools that, although the ability level is not as high as compared to the institutional schools, the gap between boys and girls is narrower.

Various Content Areas and Achievement

The latent ability of students based on the content areas is analysed here. The dataset shows that students' latent ability in mathematical skills is below the average (i.e., 0) in all the four content areas (table 5).

Table 5. Various content areas and achievement

Content areas	Grade 3 (N = 19,252)			Grade 5 (N = 13,717)		
	Mean(θ)	SD	CV	Mean (θ)	SD	CV
Algebra	-0.45	1.57	28.7	0.05	0.53	9.4
Arithmetic	-0.62	1.5	41.3	0.16	0.48	33.3
Geometry	-0.75	1.45	51.7	0.16	0.58	27.6
Numeracy	-0.61	1.42	43.0	0.16	0.52	30.8
Total	**-0.61**	**1.48**	**41.2**	**0.13**	**0.53**	**24.5**

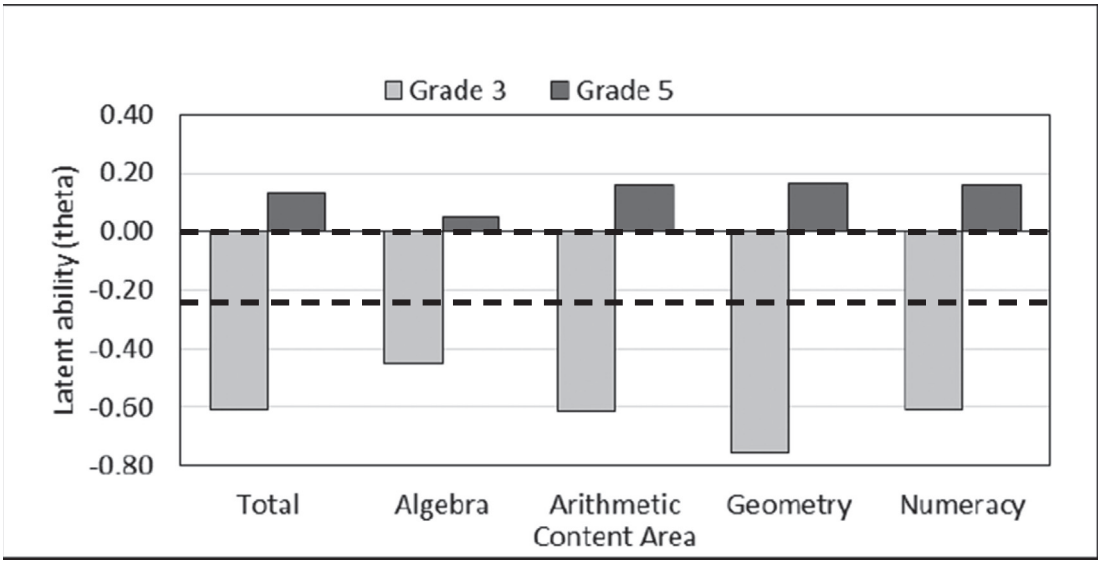

Figure 2. Content wise comparison of achievement for grade 3 and 5

Table 5 and figure 2 illustrate that students have definitely gaining additional learning abilities in grade 5 than in grade 3. For grade 3 the Algebra and Numeracy are seen to be easier areas compare to other. Students were not able to attempt difficult items from international TIMSS linking items of Geometry and hence the low ability level in it, however score was high because of most of the easy items came from Algebra and Numeracy. Also, students of grade 3 are remarkably weak in higher skill items in Geometry. Like in national level, in all the content areas boys' mathematical skill is better than girls in grade 5. However, there is no significant difference between boys and girls in any content areas. In both objective and subjective items, in both grades girls and boys do not differ much though boys outperform girls in both type of items; though effect size is small (Cohen's d <0.10). In knowledge level items, girls slightly outperform boys in grade 3 (Cohen's d = 0.04). Though the difference is statistically significant in grade 5, the effect size is smaller in all the content areas and in all hierarchical level items. From the average reference points of grade 3 (- 0.61) and grade 5 (0.13) theta score, the grade 4 ability level can be estimated to be in the middle of grades 3 and 5. This is shown in figure 2 by the dotted line.

Though the achievement level in different content areas varies depending on grades, in grade 3, the students are seen to be poorer in mathematical skills in comparison to the international grade 4 students and the grade 5 students are also not much better. There is no remarkable difference between boys and girls in grade 3 whereas boys are slightly outperforming girls in all the content areas in grade 5. Students in grade 3 are weaker in application type items in Geometry than the students in grade 5. Boys are better in application and higher ability skills than girls. However,

girls slightly outperform boys in Algebra and Numeracy in grade 3, as well as in knowledge and comprehension type of items. Additional efforts are needed to put to improve the learning ability to reach at the TIMSS average.

Ecological Zone and Variation in Achievement

As in the main report, the paper also compares the results for the four strata – Mountain, Hill, Tarai and the Kathmandu Valley considering the ecological feature of Nepal. The variation in achievement by the Ecological zones is presented in figure 3.

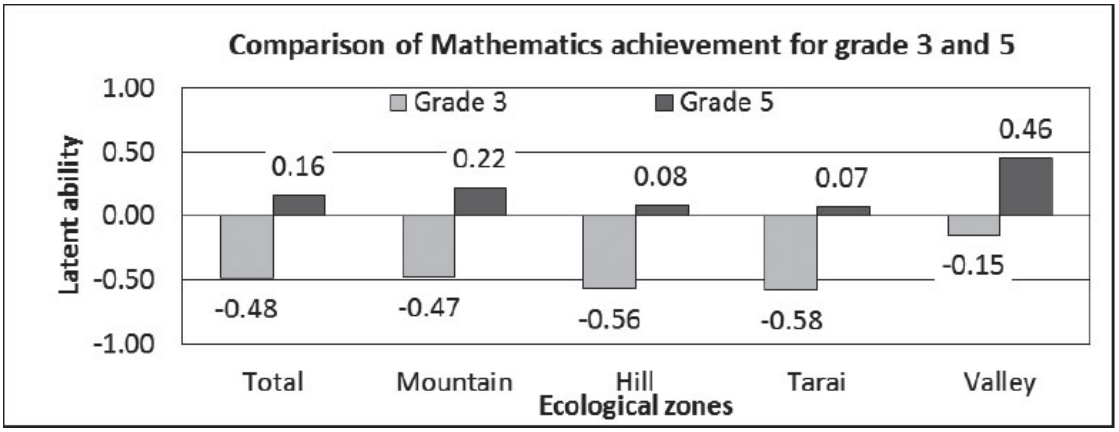

Figure 3. Ecological-Zone wise comparison of achievement for grade 3 and 5

Figure 3 shows that students from the Valley outperform students from other Ecological zones. The low Coefficient of Variation in the Valley also exhibits the same. Obviously, the students in grade 5 perform better than grade 3 in all the Ecological zones. However, the difference between the grades in Mountain zone (0.69 standard units) is seen higher than that of students in other zones (0.61 – 0.65); the difference is the smallest in the Valley. This means that, in the Valley, either the grade 3 students are comparatively somewhat better than in the other zones, and the gain is the same in all zones, or the grade 5 students do not gain as much as in the other zones. The General Linear Modelling (GLM) shows that the achievement in the zones differs significantly ($p < 0.001$). Ecological zones explain 8% of the variance in the dataset; the explaining power is somehow higher in grade 5 ($\eta^2 = 0.08$) than grade 3 ($\eta^2 = 0.09$). The effect sizes are medium ($f = 0.32$ in grade 3 and $f = 0.31$ in grade 5). When the Valley is excluded, the effect sizes are small ($f = 0.06$ in grade 3 and $f = 0.10$ in grade 5). The comparison of Ecological belt and gender wise achievement is shown in table 6.

Table 6. Gender differences in achievement in various Ecological zones

Eco zone		Grade 3				Grade 5			
		N	Mean	SD	CV[1]	N	Mean	SD	CV
Mountain	Male	1064	-0.45	0.47	95.7	673	0.25	0.50	50.0
	Female	847	-0.46	0.49	93.9	669	0.19	0.47	40.4
Hill	Male	3864	-0.56	0.49	114.3	2919	0.09	0.44	20.4
	Female	3691	-0.54	0.48	112.5	2846	0.07	0.46	15.2
Tarai	Male	2404	-0.55	0.47	117.0	1809	0.13	0.46	28.3
	Female	2346	-0.60	0.47	127.7	1621	0.03	0.47	6.4
Valley	Male	1338	-0.19	0.41	46.3	1169	0.45	0.43	104.6
	Female	1420	-0.13	0.45	28.9	1166	0.48	0.45	106.7

1) Absolute value

From the gender point of view, in most of the cases, boys are slightly ahead of girls. However, the girls in Mountain, Tarai and Valley outperform boys in grade 3. In both grades, there is significant difference between Ecological zones ($p < 0.001$) and effect size is moderate or high (Cohen's $f = 0.31$ in grade 3 and $f = 0.32$ in grade 5).

Regional Variation in Achievement

While analysing the Development region-wise achievement, the Valley is taken as a separate stratum. Hence, there are 6 strata. The achievements for each Development region are given in table 7.

Table 7. Achievement in various Development regions

Dev Region	Grade 3				Grade 5				Difference
	N	Mean	SD	CV	N	Mean	SD	CV	
Eastern	2,533	-0.61	0.5	122.0	2,037	-0.01	0.50	2.0	0.60
Central	4,127	-0.54	0.47	114.9	2,900	0.12	0.45	26.7	0.66
Western	2,784	-0.41	0.41	100.0	2,056	0.23	0.46	50.0	0.64
Mid-Western	2,158	-0.72	0.46	156.5	1,499	0.02	0.37	5.4	0.74
Far-Western	2,614	-0.50	0.52	96.1	2,045	0.12	0.46	26.1	0.62
Valley	2,758	-0.16	0.43	37.2	2,335	0.46	0.44	104.5	0.62
Total	**16,974**	**-0.48**	**- 0.50**	**96.00**	**12,872**	**0.17**	**0.48**	**35.42**	**0.65**

The students from the Eastern and Mid-Western region in grade 3 and 5 are seen to be lagging behind the other regions. Excluding the Valley, students in the Western region has the highest ability level in both grades. The latent ability gap between grade 3 and 5 is wide in Mid-Western region.

The main effect of GLM indicates that the Development region explains 13% of the variance in grade 3 and 10% in grade 5 (η^2 equals 0.13 and 0.10 in grade 3 and 5 respectively). The differences are quite high in both grades but it is more apparent in grade 5 ($f = 0.37$) than in grade 3 ($f = 0.34$).

The above data shows that there is big difference in ability of the students in Mathematics between grades 3 and 5 among various Development regions. Students in Mid-western and Eastern regions are showing low ability whereas students from Central and Western regions are performing better compared to other regions, and the ability of students in Mathematics in both grades are high in the Valley.

Districts Variations in Achievement

Some variation is seen in the in achievement level among and between Ecological zones and Development regions and it is also more apparent between and among districts as well. Though the inputs put from the government side are similar in all districts, the output varies from each other. However, the variation between grade 3 and 5 is not found in the same pattern (see figure 4).

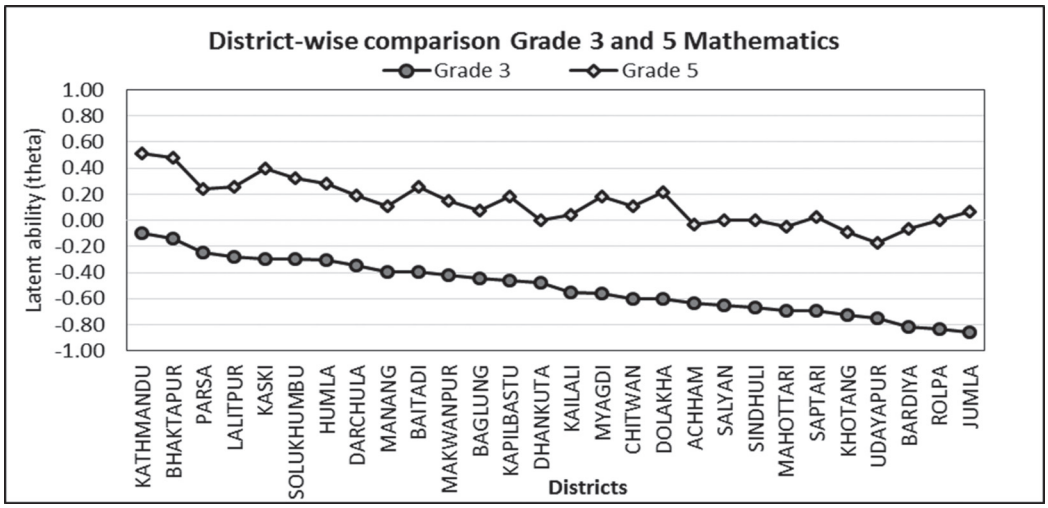

Figure 4. District variation in students' achievement for grade 3 and 5

The difference in achievement of grade 3 and 5 among district is not the same as it varies from narrow to wide. The differences are wider in Jumla, Rolpa, Bardiya, Dolakha, Saptari, Myagdi, Baitadi, Kaski and Chitwan compared to other districts. Table 8 shows the district-wise comparison of latent ability of grades 3 and 5 students and also effect size of difference in latent ability for both boys and girls by using independent sample t-test.

Table 8. District variation in achievement

Districts[1]	Grade 3				Grade 5				Difference[3]
	N	Mean	SD	Cohen's d[2]	N	Mean	SD	Cohen's d[2]	
Jumla	184	-0,86	0.46	-0.52	157	0.07	0.55	0.01	0.93
Rolpa	697	-0,83	0.42	-0.03	425	0.00	0.29	0.14	0.83
Dolakha	708	-0,60	0.46	0.00	448	0.22	0.45	0.22	0.82
Bardiya	563	-0,82	0.45	-0.08	375	-0.06	0.34	0.20	0.76
Myagdi	376	-0,56	0.36	-0.16	271	0.19	0.29	0.07	0.75
Saptari	832	-0,70	0.47	-0.09	453	0.03	0.43	0.10	0.73
Kaski	1,001	-0,30	0.37	0.14	722	0.4	0.5	-0.07	0.70
Chitwan	831	-0,60	0.44	-0.16	590	0.10	0.41	0.16	0.70
Sindhuli	824	-0,67	0.46	0.06	591	0.00	0.41	-0.05	0.67
Baitadi	742	-0,40	0.52	0.11	597	0.26	0.52	-0.01	0.66
Kapilbastu	854	-0,46	0.41	0.00	549	0.19	0.5	0.18	0.65
Salyan	649	-0,65	0.44	-0.01	503	0.00	0.31	0.04	0.65
Mahottari	660	-0,70	0.45	-0.11	449	-0.05	0.64	0.23	0.65
Khotang	581	-0,72	0.53	0.05	508	-0.09	0.46	0.03	0.63
Bhaktapur	557	-0,14	0.47	0.05	374	0.48	0.47	0.03	0.62
Solukhumbu	397	-0,30	0.36	0.18	288	0.32	0.61	0.10	0.62
Kathmandu	2,042	-0,10	0.45	0.21	1,551	0.51	0.45	-0.11	0.61
Achham	703	-0,63	0.53	-0.29	512	-0.03	0.44	0.16	0.60
Kaikali	1,193	-0,55	0.52	-0.11	739	0.04	0.42	0.20	0.59
Humla	198	-0,30	0.35	-0.31	127	0.28	0.45	0.41	0.58
Udayapur	693	-0,75	0.5	0.28	523	-0.17	0.47	0.01	0.58
Makwanpur	963	-0,42	0.48	0.08	635	0.15	0.42	0.09	0.57
Lalitpur	733	-0,28	0.35	0.14	502	0.26	0.38	0.08	0.54
Darchula	495	-0,35	0.49	0.09	411	0.19	0.44	0.08	0.54
Baklung	732	-0,45	0.44	0.06	574	0.08	0.38	0.15	0.53
Manang	19	-0,40	0.34	0.31	18	0.11	0.27	0.02	0.51
Parsa	617	-0,24	0.4	-0.10	448	0.24	0.41	0.24	0.48
Dhankuta	407	-0,48	0.51	-0.11	374	0.00	0.43	-0.01	0.48

1)Districts are ordered on the basis of the latent ability of grade 3 2) For the difference between boys and girls. Since boys are marked 1 and girls 2, effect size (Cohen's d) for the difference in the particular district is marked negative sign when girls outperform boys and vice versa. 3) Difference between grade 3 and 5 in standard units.

The difference in achievement due to the district is statistically significant ($p < 0.001$) in both grades. The variation explained in achievement due to the district is $\eta^2 = 0.188$ in grade 3 and $\eta^2 = 0.162$ in grade 5, which means that the district explains around 19% in grade 3 and 16% in grade 5 of the variation in the data. Effect size ($f = 0.48$ in grade 3 and $f = 0.44$ in grade 5) indicates remarkably high difference between the lowest and highest performing districts. In grade 5, Achham, Mahottari, Khotang, Udayapur and Bardiya districts are performing lower than international TIMSS grade 4 level. Whereas Jumla, Rolpa and Dolakha are the districts where the grade 5 students are notably higher level than the grade 3 students.

Though, in general, boys perform better than girls in Mathematics at national level, in some districts boys lag behind the girls. In grade 5, girls outperform boys in Kathmandu, Kaski, Sindhuli, Dhankuta and Baitadi. Similarly, in grade 3, girls outperform boys in 13 districts, which are indicated with negative Cohen's *d* value. Out of these, Jumla, Humla, Achham and Chitwan are the districts with moderately high difference (Cohen's $d = -0.52$ indicating remarkable difference to moderate difference $d = 0.16$) between boys and girls. Manang is the district where boys outperform girls with moderately high difference in grade 3. In grade 5, the boys in Humla outperform girls remarkably high (Cohen's $d = 0.41$) and moderately high (Cohen's $d = 0.31$ in grade 3) in Manang.

In some districts the gap between learning abilities at lower and upper grade is prominent (like Solukhumbu and Bhaktapur which are also high performers) whereas in some districts, the difference is moderate. Dataset shows that out of 28 districts, girls outperform boys in 13 districts in grade 3 and in five districts in grade 5. Girls perform better than boys in the Valley and in some other urbanized districts like Kaski and Chitwan in grade 3. In grade 5, also, the result is quite close. However, boys outperform girls in rural districts like Humla and Manang. Urbanization is seen to be favouring girls over boys in the development of Mathematics ability.

School Location and Student Achievement

Previous assessment studies had always shown the urban and rural disparity. The NASA 2012 results do not deviate from this phenomenon. The achievements of the students in rural and urban schools are presented in table 9.

Table 9. Achievement score in rural and urban areas

Location	Grade 3				Grade 5				Difference
	N	Mean	SD	CV	N	Mean	SD	CV	
Rural	14,553	-0.55	0.5	-110.0	9,940	0.09	0.46	19.6	0.64
Urban	4,074	-0.27	0.47	-57.4	2,638	0.40	0.49	81.6	0.67
Total	**18,843**	**-0.48**	**0.51**	**-94.1**	**12,796**	**0.16**	**0.48**	**33.3**	**0.64**

The difference between grade 3 and 5 student's achievement levels is more prominent in the urban area than in rural. Again, the students' performance gap between urban and rural is found more wider in grade 5 than in grade 3. The difference is statistically significant at $p < 0.001$. The effect size is medium (Cohen's $d = -0.56$ in grade 3 and $d = -0.67$ in grade 5). Generally, institutional schools are located in the urban location, which is inflating the achievement score of the urban areas. Nearly 90 percent community schools in this assessment of grade 3 and grade 5 are located in rural areas. However, among the community schools 75 percent with grade 3 and 78 percent with grade 5 are located in rural areas. While excluding the Valley from the analysis, still the difference is significant ($p < 0.001$) and effect size is medium (Cohen's $d = -0.39$ and $d = -46$ respectively in grade 3 and 5).

The medium difference between rural and urban schools is probably due to the fact that the institutional schools are more often located in the urban areas. In any case, the difference in achievement only due to the location of school is not good sign from the equality view point of equality particularly for the rural schools.

Language at Home and Student Achievement

In the test, 41 percent in grade 3 and 36 percent students in grade 5 have the mother tongue different from Nepali. The mother tongue reflects variance in many ways in learning achievements, especially in lower grades. From the equality perspective, students' mother tongue need not be an obstacle in learning. The table 10 shows to what extent the influence of language has effects in Mathematics achievement.

Table 10. Latent ability based on mother tongue

Language	Grade 3				Grade 5				Difference
	N	Mean	SD	CV	N	Mean	SD	CV	
Nepali	10,191	-0.45	0.48	93.7	8,179	0.19	0.47	40.4	0.64
Non- Nepali	5,383	- 0.49	0.48	102.1	4,996	0.10	0.50	20.0	0.59
Total	**15,574**	**-0.46**	**0.48**	**95.8**	**13,175**	**0.16**	**0.48**	**33.3**	**0.62**

Student with Nepali as the mother tongue have shown higher latent ability than the non-Nepali speakers. Although the difference is significant, effect size is very small (Cohen's $d = 0.09$) in grade 3 and in grade 5 ($d = 0.20$), favouring more to those having Nepali mother tongue in Mathematics ability. The difference between the grades is seen to be higher (0.64 standard units) with the students having Nepali mother tongue than the others (0.59). This means that the students having Nepali as mother tongue gain more in Mathematics in both the grades. The language effect was found to be higher in both grades when divided into two groups 'Nepali' and 'Non-Nepali' speakers. Nepali speakers were found high in latent ability in Mathematics, and difference is wider in upper grade.

Ethnic/Caste and Variation in Students' Achievement

Ethnicity/caste is another important affecting factor in Mathematics achievement after home language. Historically, the Brahmans and Chhetris had higher level of educational attainment and other groups like Dalits and certain groups of Janajatis had their low participation in education. However, since the last decade the educational participation from those minority groups has dramatically increased, but still their learning quality need to be ensured to create the equitable knowledge society in Nepal. The results concerning the caste/ethnicity and achievement are presented in table 11.

Table 11. Achievement variation among various ethnic groups

Ethnic/Caste group	Grade 3			Grade 5			Difference
	N	Mean	SD	N	Mean	SD	
Brahman	1,822	-0.32	0.44	1,568	0.32	0.50	0.64
Chhetri	3,155	-0.40	0.48	3,107	0.20	0.42	0.60
Dalit	2,114	-0.58	0.44	1,538	0.11	0.46	0.69
Janjati	5,491	-0.45	0.48	3,667	0.13	0.46	0.58
Madhesi	1,570	-0.51	0.49	802	0.11	0.52	0.62
Others	1,010	-0.60	0.52	2,188	0.13	0.53	0.73
Total	**16,818**	**-0.50**	**0.50**	**13,175**	**0.16**	**0.48**	**0.66**

The latent ability of all the caste/ethnic groups differs from grade 3 to 5 from 0.58 in Janjati to 0.64 in Brahman, except the 'Others' group. Brahmans are ahead in grade 3 and Madhesi in grade 5. Madhesi, Dalit and Others groups are below the national average in both grades. The overall difference between the groups is statistically significant ($p < 0.001$) and the effect size is medium ($f = 0.31$) in grade 3 and ($f = 0.17$) in grade 5. The division of students according to their ethnic/caste background explains 9% and 3% of the student variation in grade 3 ($\eta^2 = 0.09$) and 5 ($\eta^2 = 0.027$) respectively.

Brahmans are ahead among the other group whereas Madhesi and Dalit students are poorer in Mathematics ability. The effect size of caste/ethnicity of high performer and low performer is medium in both grades. However, the variation explains less in grade 5 (3%) than 3 (9%). Gender difference is not seen in high achiever Brahman and low achiever Dalit. Except in Madhesi group, in all ethnic/caste group girls' Mathematics ability is better than boys.

Summary

Having described the achievement status, the main results for the both grade can be summarized in the following points:

- The differences in proficiencies between boys and girls are not significant in grade 3. From the equality point of view this is a positive sign. Though the difference is significant in grade 5, it is very small. It is the tendency that the boys are slightly outperforming the girls. This gap in latent ability is widening in upper grades.

- The students in institutional schools outperform the students in community schools and the gap of learning ability expands more in the upper grade. The gap in the latent ability between boys and girls is seen equal in grade 5 in both types of schools. However, the differences between girls and boys are more pronounced in institutional schools in grade 3 than in the community schools. From the equality view point, it is a positive aspect in the community schools that, although the ability level is not as high as compared to the institutional schools, the gap between boys and girls is small.

- In both grades, the achievement level across the content areas varies depending on grades. Students' mathematical skills are lower than the level–international grade 4 students. There is no remarkable difference between boys and girls in grade 3 whereas boys are slightly outperforming girls in all the content areas in grade 5 so the gap is wider in grade 5. Boys are also better than girls in the tasks requiring application and higher ability. However, girls slightly outperform boys in Algebra and Numeracy in grade 3, as well as in tasks requiring recalling and comprehenson. Lots of efforts need to put to improve the achievement level to reach the average international standard.

- There are only small differences among the Ecological zones when the Valley is excluded in both grades where Mountain outperforms other three Ecological zones. The Valley students outperform the others in all the areas. After Valley, Mountain zone is the second highest among the Ecological zones. The differences extends moderately within the years. Less or no gender gap is noted among zones.

- With regards to the Development regions, there are wide differences in ability of the students. The students from Mid-Western and Eastern regions are at relatively low ability level where as the students from the Central and Western regions perform better. Again the Valley is at the top of all. In the Valley, which is incomparable with other zones, students outperform the students in all other regions. Eastern region is recorded as the lowest. Less or no gender gap is noted among regions.

- In some districts the gap in learning abilities between lower and upper grade is prominent whereas in some districts, the difference is medium. Dataset of grade 3 shows that out

of 28 districts, girls outperform boys in 13 districts in grade 3 and 5 districts in grade 5. Girls perform better than boys in the Valley and in some other urbanized districts like Kaski and Chitwan in grade 3. In grade 5 too, the result is quite close. However, boys also outperform in rural Districts like Humla and Manang. Urbanization is seen to be favoring girls over boys in Mathematics ability.

- The medium difference between rural and urban schools is seen due to fact that the institutional schools are located more often in the urban areas. In any case, the difference is not good from the view point of the equality for location.

- The language effect was found to be higher in both grades when divided into two groups 'Nepali' and 'Non-Nepali' speakers. Nepali speakers were found higher in achievement and difference is wider in upper grade. It indicates that geographical factor explains more than language factor in students ability though there are many other factors associated with.

- Brahmans and Chhetris are ahead of all the group. The ethnicity-wise effect size of high performer and low performer is medium in both grades. However, the variation explains less in grade 5 (3%) than in grade 3 (9%). Gender difference is not seen in high achiever Brahman and low achiever Dalit.

Conclusion

Having analysed the above description on comparative results, some conclusions can be drawn regarding the Mathematics ability for third and fifth graders. While looking at the results from gender perspective, one notices no disparity in Mathematics ability between girls and boys at lower grade as both have developed same level of proficiency and in lower order skills which is a positive indication toward equality. However, there is still more to do on the part of girls to achieve equality when knowing that boys outperform girls in grade 5 and in higher order ability implying that gender disparity widens more in upper grades. Though the difference is not so wide, results still favours slightly to boys .

While looking at region or district wise results, a remarkable variance is observed between Ecological and Development regions showing an inequality to reach at the expected level of ability. Despite the equitable intervention across the different part of the country, Eastern region and Tarai still record the lowest ability of all lagging far behind the Mountain and the Valley. Similarly, the differences in achievement from lowest to highest level between the districts also show an inequality in learning opportunity for children across the country which require further attention to achieve the equality goal. Besides the geography, location of school the rural–urban differences in achievement is noticeable favouring more to urban students. Though urban favours more to girls than boys, urban schools are seen to be offering better learning opportunities than the rural ones. This level of rural urban differences might have been created due to the increased

number of private schools in the urban location that always have higher achievement than the community schools.

Inequality also pertains in certain caste/ethnic and language groups though it is not clear whether it is due to their ethnic or linguistic background. Looking at the dataset, still Dalits and some ethnic groups including Madhesi groups lag behind Brahman and Chhetri, perpetuating another kind of inequality in society though it narrows down in upper grade. Similarly, students with Nepali mother tongue are seen higher in latent ability than other non-Nepali speakers indicating an influencing factor for the achievement.

As achievement in different content areas and cognitive domains varies from lower to high level, inequality also persists in providing similar level of learning opportunity across all curricular contents where children are found weaker in application and higher ability skills. Besides such unbalanced level of learning, Nepali students are also seen poorer in mathematical skills in comparison to the international standards. Given situation requires additional efforts to put to improve the learning ability of Nepali students.

This comparison of grade 3 and 5 shows that the learning achievement in grade 3 are lower than that of grade 5 in Mathematics. The situation urges that more concentration on improving the pedagogical process is needed in lower grades than the upper ones, particularly for grades 1 to 3 in order to improve students' learning achievement.

References

BPEP. (1995). *Basic and Primary Education Project Achievements Study*. Kathmandu: Author.

----- (1997). *The Effect of New Curriculum on the Achievement of Grade 4 Student.* Kathmandu: Author.

----- (1998). *The Effect of New Curriculum on the Achievement of Grade 5 Student*. Kathmandu: Author.

CERID. (1998). *Performance Levels of Grade V Students.* Kathmandu: Author.

CERSOD. (2001). *District Level Assessment 2001.* Sanothimi: Department of Education (DOE).

Cohen, J. (1988). *Statistical Power for the Behavioural Science (2nd Edition).* Hillsdale NJ; Lawrence Erlbaum.

EDSC. (1997). *National Assessment of Grade 3 Students.* Kathmandu: BPEP.

----- (1999). *National Assessment of Grade 5 Students.* Kathmandu: BPEP.

----- (2001). *National Assessment of Grade 3 Students.* Sanothimi: DOE.

----- (2003). *National Assessment of Grade 5 Students*. Sanothimi: DOE.

----- (2008). *National Assessment of Grade 8 Students.* Sanothimi: DOE.

Education Review Office (ERO). (2013). W*here Are We Now? Results of Student Achievement in Mathematics. Nepali and Social Studies in the Year 2011*. Kathmandu: ERO.

-----(2015). *National Assessment of Student Achievement(NASA) 2012 (Grade 3 and 5).* Kathmandu: ERO.

Fulbright Consultant (FBC) & Cooperative Hands in Restoration (CHIRAG). (2008). *A Final Report on National Assessment of Grade-V Students*. Sanothimi: DOE.

OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing.

Comparing the Achievement of Nepali for Grade 3 and 5*

Education Review Office

Abstract

Drawing entirely upon the database of National Assessment of Student Achievement (NASA) 2012 for Grade 3 and 5, this paper further analyses comparatively the achievement in Nepali considering various diversity and equality factors such as ecological belts, development region, districts, types and location of schools, caste/ethnicity and language of students. Gender perspective crosscuts throughout the analysis. While doing so, an attempt is also made to analyse the results comparatively on the basis of latent ability (θ) produced by Item Response Theory (IRT) modelling and presents the comparable scenario of students' competency in Nepali for both grades. Nepali language proficiency of Nepalese students has also been described comparing with the Progress in International Reading Literacy Study (PIRLS) standard by using the Common European Framework of Reference for Language (CEFR) levels. Obviously, the 5th grader have comparatively higher level knowledge, comprehending capabilities and skills of solving higher level tasks than the 3rd graders in the subjects in terms of latent ability . The first part of the paper compares basic results, the second part delves into describing Nepali language proficiency in comparison with international standards. At the end, it presents summary and conclusion.

***Keywords**:* assessment, cognitive skills, latent ability, learning achievement, language skills

Introduction

Learning achievement in Nepali subject for grade 3 and 5 are assessed systematically and frequently in Nepal since the mid 1990s along with the beginning of the Basic and Primary Education Project (BPEP) though the results of the previous national assessments (see BPEP, 1995; 1997;1998; CERID, 1998; EDSC, 1997; 1999; 2001;2003; 2008; CERSOD, 2001; Fulbright, 2008) are not fully comparable in the absolute sense (as, for example, percentages of correct answers are not mentioned) with each other because of the missing linking procedure between the tests. In NASA 2012 assessment, a linking procedure is created between grade 3 and 5 tests by using the principles of Item Response Theory (IRT) modelling. The latent theta (θ) is used in the comparison of results in three ways: first, in equating the three versions of tests of both grades; second, it is also used in calibrating all the items of both grades with PIRLS (Program in International Reading Literacy Study) scales (see ERO, 2015, Chapter 2 for detail) in which the international average is set as theta $\theta = 0$; third, the latent theta is used to make a comparison between grade 3 and 5 results.

---

* *This article is the part of the report "National Assessment of Student Achievement (NASA) 2012 (Grade 3 and 5)" prepared by Sunita Shakya, Section Officer, ERO, which is further edited and re organized in this form by Gopal Prasad Bhattarai, Under Secretary, ERO.*

Although there is separate analyses of grade 3 and 5 Nepali (Chapter 4), it does not compare the percentage of the raw scores (See ERO, 2015). Here in this paper, the results are compared on the basis of latent ability(theta) presenting the scenario of students' proficiency in Nepali. In this regards, this comparative analysis gives a synopsis of the existing reality of the Nepalese education system as to what extent 5th graders are better than 3rd graders. Further, by using the Common European Framework of Reference for Language (CEFR) standards, it gives an additional scenario of the Nepali language proficiency of the lower and upper primary levels. The comparison is made here mainly from the perspective of quality and equality – gender, caste/ethnicity, language, various geographical regions, and school types and locations.

This article first tries to compare basic results in Nepali subject analyzing the distribution of student population in terms of achievement score, basic results by subjects, contents, types of items, gender, location and types of schools and cognitive domains. Then it proceeds ahead to compare across ecological belts and development region, variations based on district, school location in achievement vis-à-vis the comparison between case/ethnicity and language of students followed by the description of Nepali language proficiency of Nepalese students in relation to the CEFR standard.

Methodology

This article entirely draws the secondary data of NASA 2012 generated from the test conducted in 28 districts of Nepal covering all Ecological zones and Development regions including the Kathmandu Valley as a separate stratum. For the detail methodological questions, the reader is asked to go through the main report of NASA 2012 (ERO, 2015), particularly the chapter 2 – the technical matters of sampling, test instruments and analytical tools.

The test for the assessment was administered in 1690 schools (847 schools for grade 3 and 570 schools for grade 5) with 19501 students in grade 3 and 13971 students in grade 5, using random sampling method representing each stratum. For the test, three versions of standardized test items were used with some linking items in each set. Using IRT modelling, items were calibrated and the versions were made comparable by equating the scores. Use of three versions of test items facilitate for wider coverage of curricular contents as well as for ensuring high reliability in test administration. Items were standardized by developing items based on curriculum of respective grades, their difficulty level was pretested through a piloting and only the items having appropriate difficulty level were selected with their further analysis. Teachers, subject specialist and professional participated in the process of test construction and item selection. Validity of the test items were ensured by analysing and covering of curricular contents. Some linking items from international assessment, PIRLS, were calibrated to compare the results with the international standards and to describe the proficiency based on CEFR level. Data were tabulated using Optical Mark Reader (OMR) sheet and analysed using One Parameter Logistic Model (OPLM). The overall assessment results in Nepali for grade 3 and 5 were compared considering some variables. Basic comparison is done by calculating latent ability of students (Theta, θ) and also by calculating effect size.

Comparing Basic Results

Altogether 33,372 students (51 percent girls and 49 percent boys) participated in the Nepali test (19,501, grade 3 and 13,971, grade 5) from 28 districts. As expected the grade 5 students achieve higher (θ = –1.40) than grade 3 students (θ = –1.57). The difference can be seen from figure 1.

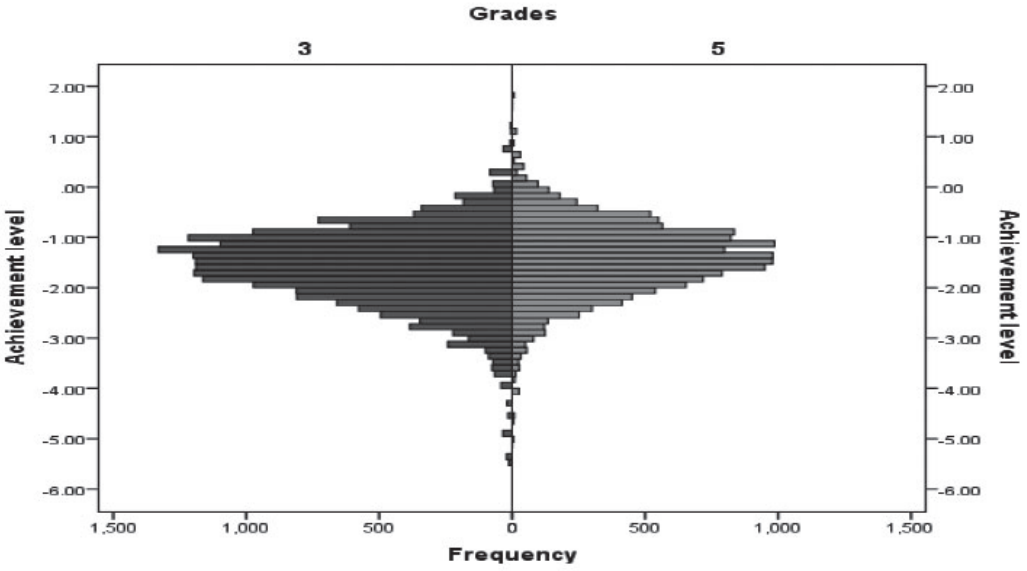

Figure 1. Distribution of the students' achievements by grade

Comparing the results of grade 3 and 5, the figure 1 shows that the students at grade 3 (left side of figure 1) are widely distributed with longer lower tail whereas grade 5 students (right side of figure 1) tend to be more concentrated within normal distribution curve. In both grades, the average students are located in between the θ value of –2 to –1 (around 52 to 53 percent of students in grades 3 and 5). However, when looking more closely, there are more students with higher score in grade 5 than in grade 3, i.e., above θ = –1; 29 percent in grade 5 and 21 percent in grade 3. And it is just reversed below the theta value –2 (around 27 percent in grade 3 and 19 percent in grade 5). Likewise, the number of students above the international average (θ = 0) is 2 percent in grade 5 and less than 1 percent in grade 3. Note that the majority of students are below the international average of grade 4 students in Nepali language proficiency. This situation urges for extra efforts to put in pedagogical processes to raise the learning ability of large number of students.

Gender and Achievement

Gender-wise comparison is made under various sub-headings it follows. However, the main comparison is made here as it reflects the whole population scenario(table 1).

Table 1. Gender-wise comparison of achievement

Gender	Grade 3				Grade 5			
	N	Mean(θ)	D	V[1]	N	Mean(θ)	D	V[1]
Boys	8,586	-1.59	.76	7.8	6,449	-1.43	.72	0.4
Gils	8,861	-1.54	.76	9.6	6,977	-1.37	.73	3.4
Total	**17,447**	**-1.57**	**0.76**	**48.8**	**13,546**	**-1.40**	**.73**	**1.9**

1) Absolute value

Overall, the gap between girls' and boys' achievement is small. The difference is more or less same in both grades favouring slightly (0.05– 0.06 standard units) the girls over the boys. It is seen that girls are better in Nepali language skills from lower grades and they are able to maintain their status in upper level as well. The same trend – the girls perform slightly over the boys – was seen also in grade 8 dataset (ERO, 2013, p. 141, Table 3.2.20).[1]

The differences between boys and girls in Nepali proficiency are very moderate, especially in grade 3 which is a positive sign from the equality point of view. The tendency also hints that the girls are slightly out-performing the boys in Nepali language skills. It is noteworthy to understand that girl's language skill is also better across the OECD countries (Moss, Francis & Skelton, 2009, p. 3).

School Types and Achievement

Generally, students' performance in institutional schools is better than in the community school though there are also high performing students in community schools.

1 *NASA 2011 dataset shows that the latent ability for boys was θ = -0.92 and for girls θ = -0.84, that is, the difference is 0.08 indicating that the difference does not change much during whole school time.*

Table 2. School type and achievement

School Type	Grade 3				Grade 5			
	N	Mean(θ)	D	V	N	Mean(θ)	D	V
Community	14,712	-1.78	.74	1.6	10,842	-1.58	.67	2.7
Institutional	4,789	-1.02	.58	6.5	3,129	-0.80	.58	2.0
Total	**19,501**	**-1.59**	**.78**	**8.7**	**13,971**	**-1.40**	**.73**	**1.9**

On the basis of table 2, it is evident that, on average in both grades, the students from the institutional schools perform higher than the students from the community schools. Again, they acquire better language skill when they are in the upper grade. But the gap is narrower in grade 3 (0.76) and it is slightly wider in grade 5 (0.78).[2] It means that the institutional school students are ahead in the lower grade and they are acquiring more learning skills in grade 5 than students of community schools. The differences in both grades are statistically significant ($p = < 0.001$) and the effect size is somewhat higher in grade 5 (Cohen's $d = 1.15$) than in grade 3 ($d = 1.08$).

Table 3. School type and gender wise achievement

School Type	Gender	Grade 3					Grade 5				
		N	Mean θ	SD	Difference G-B[1]	Cohen's *d*	N	Mean θ	SD	Difference G-B[1]	Cohen's *d*
Community	Boys	6,340	-1.78	0.73	0.05	-0.07	4,884	-1.60	0.67	0.06	-0.10
	Girls	6,651	-1.74	0.72			5,524	-1.55	0.67		
Institutional	Boys	2,246	-1.05	0.58	0.10	-0.20	1,565	-0.87	0.58	0.16	-0.32
	Girls	2,210	-0.95	0.57			1,453	-0.71	0.56		

1) G=girls and B=boys

From the gender perspective (table 3), the difference in achievement between girls and boys is greater in institutional schools (0.1 in grade 3 and 0.16 in grade 5) than in the community schools (0.05 and 0.06 in grade 3 and 5) in both grades and, again, the gap is slightly higher in grade 5. Though there are differences, the effect size indicates the small difference in both grades, i.e., Cohen's $d = -0.20$ at highest in grade 3. However, in grade 5 the gap in institutional schools looks narrower to medium size ($d = -0.32$).

The students from institutional school outperform students from the community school as a whole and the gap of learning ability widens in the upper grade. However, in both grades, the difference between girls and boys are pronounced more in the institutional schools than community schools. From the equality viewpoint, it is a positive aspect in the community schools that, although the ability level is not as high as in the institutional schools, the gap between boys and girls is narrow.

2. *In grade 8, in NASA 2011 dataset, the difference is 0.80, that is, there seems to be a tendency for the gap to grow moderately within the years.*

Content Areas and Achievement

The latent ability of students based on the content areas is analysed here. The dataset shows that students' latent ability in language skills is below the average (i.e., 0) in all the four content areas (table 4).

Table 4. Different content areas and Achievement

Content areas	Grade 3 (N=19,501)			Grade 5 (N=13,971)		
	Mean	**SD**	**CV**	**Mean**	**SD**	**CV**
Reading	-1.61	0.87	185.1	-1.41	0.84	167.9
Writing	-1.66	0.85	195.3	-1.45	0.82	176.8
Grammar	-1.67	0.79	211.4	-1.38	0.81	170.4
Vocabulary	-1.67	0.83	201.2	-1.37	0.85	161.2
Total	**-1.59**	**0.78**	**203.8**	**-1.40**	**0.73**	**191.8**

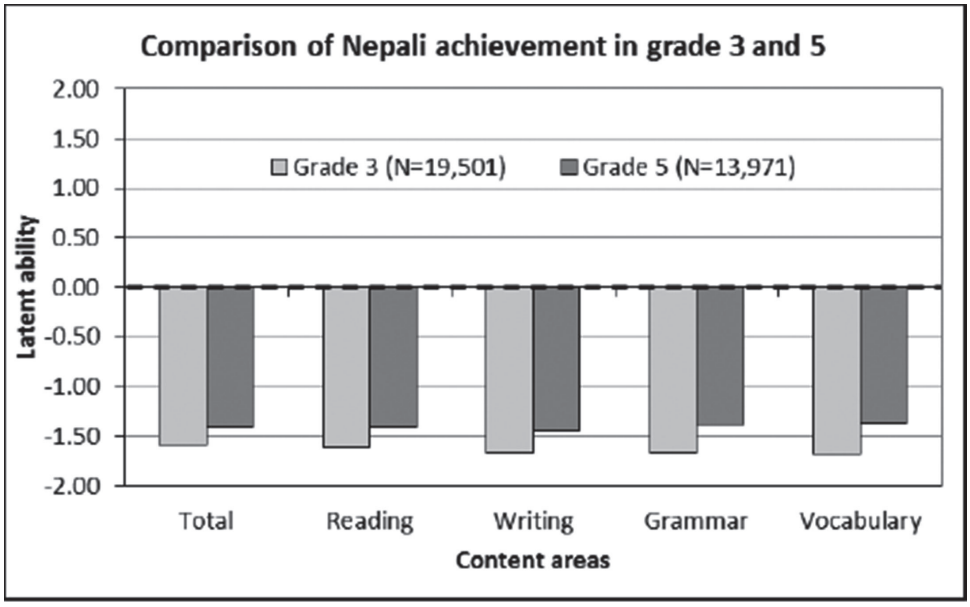

Figure 2. Content wise comparison of achievement for grade 3 and 5

Table 4 and figure 2 illustrate that students have definitely gained additional learning abilities in grade 5 than grade 3. Contrary to grade 3, grade 5 students are found to be performing better in those content areas. Nevertheless, writing tends to be difficult for both graders, obviously

mainly in grade 3. Like at the national level, in all the content areas girls' language skill is better than boys in both grades in community as well as institutional schools. Though the difference is statistically significant, the effect size is smaller in all the content areas (Cohen's $d < 0.2$). However, the gap in higher ability is seen to be wider in grade 5 (girls $\theta = -1.41$ and boys $\theta = -1.53$) than in grade 3 (girls $\theta = -1.61$ and boys $\theta = -1.65$). This means that the girls are comparatively better in language skills. From the average reference points of grade 3 (–1.59) and 5 (–1.4) in theta score, the grade 4 ability level can be estimated to be in the middle of grades 3 and 5. This is shown in figure 2 by the dotted line.

In both grades, there is not much differences in various content areas. In comparison to the international standard of grade 4 students, the proficiency level is much lower. Girls perform better than boys in all the content areas in both grades, however, the gap is widened in grade 5 calling for additional of efforts to put to improve the achievement level to meet the level of international average.

Results Based on Ecological Zone

The result is disaggregated into three strata – the Mountain, Hill, Tarai and the Valley considering their different ecological features. The variation in the Ecological zones is condensed in figure 3.

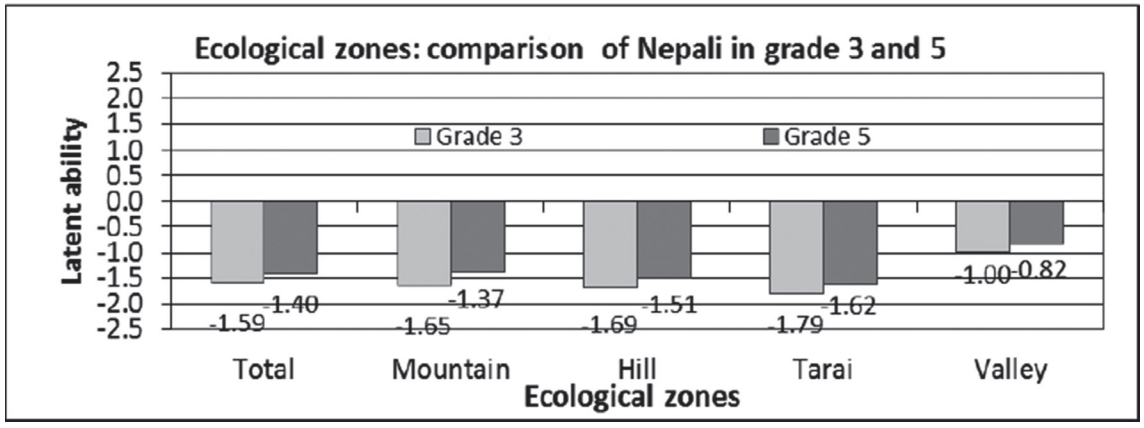

Figure 3. Ecological zone wise comparison of achievement in grade 3 and 5

Figure 3 shows that students from the Valley outperform students from other Ecological zones. Obviously, students in grade 5 perform better than grade 3 in all the Ecological zones. However, the difference in Mountain zone between grade 3 and 5 (0.28) is seen higher than in other zones. The lowest difference between two grades is noticed in Hill region (0.17) followed by Tarai region (0.18). The GLM indicates that the achievement in the zones differs significantly ($p < 0.001$). Ecological zones explain 15% of the variance in the dataset and the explaining

power is somewhat higher in grade 5 ($\eta^2 = 0.149$) than grade 3 ($\eta^2 = 0.127$). Likewise, the effect size (Cohen's *f)* is also higher in the grade 5 ($f = 0.42$) than grade 3 ($f = 0.38$). The effect sizes are smaller in both grades if the Valley is taken out of the analysis (grade 3, $f = 0.07$ and grade 5, $f = 0.11$ respectively). The comparison of Ecological zone and gender wise achievement is shown in table 5.

Table 5. Ecological zone and gender wise variation in Achievement

Ecological zone		Grade 3				Grade 5			
		N	Mean	SD	CV	N	Mean	SD	CV[1]
Mountain	Male	736	-1.7	0.7	40.7	564	-1.3	0.6	47.1
	Female	702	-1.6	0.7	42.7	715	-1.3	0.7	50.9
Hill	Male	4,075	-1.7	0.7	44.2	2,934	-1.6	0.7	42.4
	Female	4,200	-1.6	0.7	45.4	3,107	-1.5	0.6	44.1
Tarai	Male	2,325	-1.7	0.7	42.0	1,838	-1.6	0.7	46.2
	Female	2,382	-1.7	0.7	40.1	1,881	-1.6	0.7	44.7
Valley	Male	1,450	-1.0	0.6	59.8	1,113	-0.9	0.6	69.6
	Female	1,577	-0.9	0.6	64.7	1,274	-0.8	0.6	79.7

In both grades, girls are either slightly ahead than boys or they are equally competitive with boys. In Tarai in either grades and in Mountain in grade 5, the gender gap is not notable. From the performance point of view, Mountain region is better after Valley. The gender difference in Ecological zones is statistically significant at $p < 0.001$. The high effect size of the Ecological zone is notable in both grades (Cohen's $f = 0.36$ for boys and $f = 0.42$ for girls in grade 3; $f = 0.39$ for boys and $f = 0.45$ for girls in grade 5), expanding the differences moderately within the years. The variation between the Ecological zones in language proficiency is remarkable. The Valley students outperform the others in all the areas. After Valley, Mountain zone is the second highest achieving among the Ecological zones. The differences is seen to be expanding moderately over the years.

Results Based on Developmental Region

Results have been analysed considering the 6 strata: the five development regions and the Kathmandu Valley, which, though, administratively lies within the Central development region. Hence, there are altogether 6 strata analysed in this section. The achievements in the Development regions are given in table 6.

The students from the Eastern and Mid-Western regions are seen to be lagging behind other development regions. Excluding Valley, students in Western region have the highest ability level in both grades. The students in grade 3 are at much higher level compared with the other regions. The difference between grade 3 and 5 in Western region (0.14) is less than Valley (0.18). The latent ability gap between grade 3 and 5 is widened more in Eastern, Mid-Western and Far-Western development regions.

Table 6. Achievement in the development regions

Dev Region	Grade 3				Grade 5				Difference
	N	Mean	SD	CV	N	Mean	SD	CV	
Eastern	3,043	-1.89	0.77	245,5	2,207	-1.66	0.79	210,1	0.23
Central	4,568	-1.72	0.74	232,4	3,106	-1.57	0.67	234,3	0.15
Western	3,023	-1.47	0.68	216,2	2,185	-1.33	0.64	207,8	0.14
Mid-Western	2,328	-1.84	0.71	259,2	1,702	-1.63	0.67	243,3	0.21
Far-Western	3,164	-1.70	0.75	226,7	2,274	-1.48	0.63	234,9	0.23
Valley	3,375	-1.00	0.63	158,7	2,497	-0.82	0.61	134,4	0.18
Total	**19,501**	**-1.59**	**0.78**	**203,8**	**13,971**	**-1.40**	**0.73**	**191,8**	**0.19**

The main effect of GLM indicates that the Development region explains 15% of the variance in grade 3 and 16% in grade 5 (η^2 equals 0.15 and 0.16 in grade 3 and 5 respectively). The effect size is quite high in both grades but it is more apparent in grade 5 (f = 0.44) than in grade 3 (f = 0.42). As in the Ecological zones, in the Development regions, excluding Valley, the effect sizes are much smaller (f = 0.19 and f = 0.17 in grade 3 and 5 respectively). The comparison of the Development region and gender wise achievement is shown in table 7.

Table 7. Gender and achievement in the development regions

Dev. Region		Grade 3				Grade 5			
		N	Mean	SD	CV	N	Mean	SD	CV
Eastern	Boys	1,339	-1.9	0.8	237.5	1,028	-1.7	0.8	212.5
	Girls	1,423	-1.8	0.7	257.1	1,113	-1.6	0.8	200.0
Central	Boys	1,826	-1.7	0.7	242.9	1,418	-1.6	0.7	228.6
	Girls	2,018	-1.7	0.7	242.9	1,519	-1.5	0.6	250.0
Western	Boys	1,431	-1.5	0.7	214.3	1,023	-1.4	0.6	233.3
	Girls	1,436	-1.4	0.7	200.0	1,100	-1.3	0.7	185.7
Mid-Western	Boys	1,109	-1.8	0.7	257.1	819	-1.6	0.7	228.6
	Girls	1,130	-1.8	0.7	257.1	847	-1.6	0.7	228.6
Far-Western	Boys	1,431	-1.7	0.8	212.5	1,048	-1.4	0.6	233.3
	Girls	1,277	-1.7	0.7	242.9	1,124	-1.5	0.6	250.0
Valley	Boys	1,450	-1.0	0.6	166.7	1,113	-0.9	0.6	150.0
	Girls	1,577	-0.9	0.6	150.0	1,274	-0.8	0.6	133.3
Cohen's *d*	Boys	0.40				0.42			
	Girls	0.47				0.46			

Except for the Far-Western development region of grade 5, in all the Development regions girls perform either somewhat better or they have the same ability level as that of the boys in Nepali language. Excluding the Valley, in both grades, the students from the Western region are seen to be performing highest which is also true for both boys and girls from gender perspective.

GLM shows that the difference of girl is more or less similar in grade 3 and 5 (η^2 = 0.179 in grade 3 and η^2 = 0.173 in grade 5) but, in the case of boys, the difference is wider in grade 5 than in grade 3 (η^2 = 0.139 in grade 3 boys and η^2= 0.15 in grade 5 boys). Cohen's *d* also shows a higher effect size of girls than boys in both grades (table 7).

Ecological zone, Development region, school location and district variations have played a crucial role in Nepalese education The Cohen's *d* shows high effect sizes, that is, the differences are remarkable. In the Valley, which is incomparable with other zones, students outperform the students in all other Ecological zones followed by Mountain zone, the second highest. In Development regions, also, the Valley students outperform the other regions and Eastern region is recorded as the lowest. Less or no gender gap is noticed among zones and regions.

Results Based on Districts

The variation in achievement level as found in various Ecological zones and Development region is also seen more distinctly at district level too. Though the inputs from the government side are the same in all districts, the output varies from each other. The variation between grade 3 and 5 is seen to be in the same pattern (figure 4).

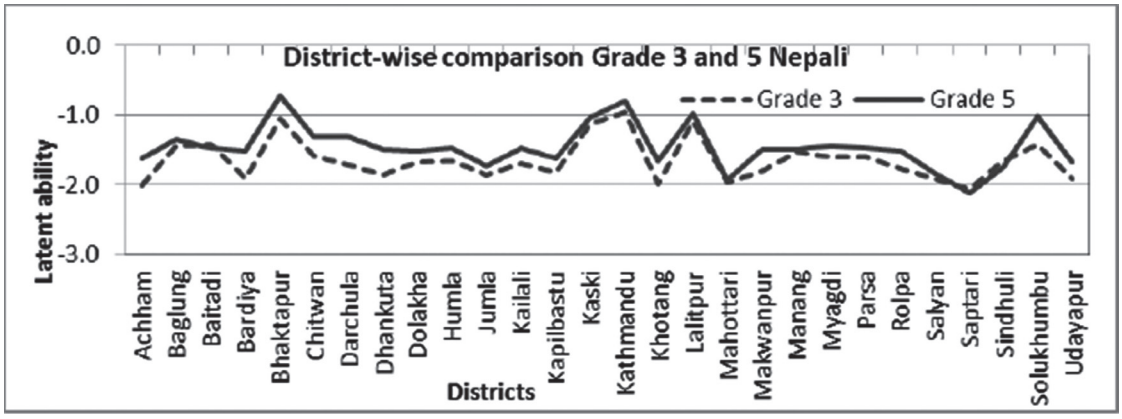

Figure 4. District-wise students' achievement in grade 3 and 5

In some district, the difference between the students from grade 3 and 5 is much smaller than in other districts. The less difference (less than – 0.1) between latent abilities of grade 3 and 5 is found mainly in Mahottari, Baitadi, Manang, Saptari, Salyan, and Sindhuli districts. It means that the achievement level of lower and upper primary grades do not differ much in those districts. However, in some districts (Darchula, Bardiya, Solukhumbu, Achham, Dhankuta, Bhaktapur, Khotang and Makwanpur) students' achievement level in grade 5 has improved much compared to the grade 3. While looking at the difference in the highest performing districts

(i.e., Kathmandu, Bhaktapur, Lalitpur, Kaski, and Solukhumbu), the difference in performance is pronounced more in Solukhumbu and Bhaktapur district than other three top performing districts. Even in the highest performing districts like Kathmandu, Lalitpur and Kaski, the variance in the ability levels in 3rd and 5th grade are quite small.

Table 8. Achievement in the sampled districts

District[1]	Grade 3				Grade 5				Difference[3]
	N	Mean	SD	Cohen's f^2	N	Mean	SD	Cohen's f^2	
Darchula	535	-1.72	0.76	0.10	372	-1.31	0.55	0.19	0.41
Solukhumbu	375	-1.42	0.61	0.08	294	-1.02	0.66	0.11	0.40
Bardiya	548	-1.91	0.80	0.00	361	-1.50	0.80	0.05	0.40
Achham	725	-2.00	0.75	0.03	503	-1.61	0.73	0.07	0.39
Dhankuta	521	-1.86	0.78	0.09	387	-1.50	0.66	0.14	0.36
Bhaktapur	544	-1.05	0.49	0.13	408	-0.73	0.51	0.04	0.32
Khotang	659	-1.98	0.81	0.00	504	-1.66	0.74	0.26	0.31
Makawanpur	979	-1.80	0.78	0.00	603	-1.50	0.63	0.08	0.30
Chitwan	842	-1.59	0.68	0.00	606	-1.31	0.59	0.04	0.28
Rolpa	699	-1.77	0.68	0.07	501	-1.50	0.49	0.13	0.27
Kailali	1,120	-1.70	0.68	0.00	847	-1.47	0.64	0.12	0.23
Udayapur	768	-1.91	0.70	0.00	495	-1.68	0.70	0.07	0.23
Kapilbastu	817	-1.82	0.70	0.03	599	-1.62	0.72	0.05	0.20
Humla	214	-1.66	0.66	0.00	156	-1.48	0.59	0.04	0.18
Dolakha	587	-1.68	0.67	0.07	417	-1.51	0.66	0.34	0.18
Myagdi	426	-1.60	0.70	0.00	302	-1.43	0.53	0.18	0.17
Kathmandu	2,110	-0.96	0.66	0.08	1,558	-0.79	0.61	0.10	0.16
Parsa	638	-1.61	0.57	0.18	453	-1.48	0.55	0.04	0.13
Lalitpur	721	-1.09	0.60	0.12	531	-0.97	0.63	0.08	0.12
Jumla	190	-1.86	0.50	0.19	159	-1.74	0.73	0.19	0.12
Kaski	974	-1.13	0.55	0.08	711	-1.03	0.53	0.18	0.11
Baglung	790	-1.44	0.59	0.00	564	-1.34	0.57	0.07	0.11
Sindhuli	858	-1.65	0.78	0.00	562	-1.74	0.63	0.16	0.09
Salyan	677	-1.92	0.73	0.00	525	-1.84	0.67	0.00	0.07
Baitadi	784	-1.42	0.72	0.03	552	-1.47	0.52	0.08	0.05
Manang	16	-1.53	0.74	0.25	9	-1.48	0.46	0.62	0.05
Saptari	720	-2.06	0.76	0.03	527	-2.11	0.79	0.18	0.05
Mahottari	664	-1.97	0.83	0.04	465	-1.94	0.74	0.11	0.03

1. Districts are ordered on the basis of the difference between the grade 3 and 5.
2. Effect size for the difference between boys and girls.
3. Difference between grade 3 and 5 in standard units.

The difference in achievement due to the district is statistically significant ($p < 0.001$) in both grades. The variation explained in achievement due to the district is $\eta^2 = 0.197$ in grade 3 and $\eta^2 = 0.237$ in grade 5, which means, that the district explains around 20% and around 24% of the variation in grade 3 and 5 respectively in the data. Effect size ($f = 0.50$ in grade 3 and $f = 0.56$ in grade 5) indicates remarkably high difference between the lowest performing and highest performing district.

Though, in general, girls perform better than boys in Nepali language at district level, in some districts girls lag behind. In comparison with the grade 3, in grade 5 the girls are behind the boys in more districts. These districts are, for example, Bardiya and Darchula from the Mid-Western region in grade 3 and only Saptari from the Eastern region. Similarly, in grade 5, the girls lag behind the boys in Saptari from the Eastern region; in Mahottari from the Central region; in Manang from the Western region; Jumla and Salyan from the Mid-Western region; and Darchula, Kailali and Achham from the Far-Western region. It is found that, out of the 28 sample districts, Saptari and Mahottari are the lowest performing districts.

In some districts the gap between the learning outcomes at lower and upper grade is prominent (like Solukhumbu and Bhaktapur which are also high performers) whereas, in some other districts, the difference is moderate. Overall, the effect size shows high discrepancies between the high and the low performing districts. Though girls are better at the national level, but while analysing at the district level, in some districts boys outperform the girls. Example of such districts are: Bardiya, Darchula and Saptari in grade 3 and Darchula, Jumla, Salyan, Kailali, Achham, Saptari and Mahottari in grade 5.

School Location and Student Achievement

Previous assessment studies had shown the urban and rural disparity in achievement. The NASA 2012 also shows the similar results. The achievements of the students in rural and urban schools are presented in table 9.

Table 9. Achievement score in rural and urban areas

Location	Grade 3				Grade 5				Difference
	N	Mean	SD	CV	N	Mean	SD	CV	
Rural	14,112	-1.70	0.76	223,7	10,266	-1.53	0.69	221,7	0.17
Urban	3,990	-1.19	0.65	183,1	2,708	-0.94	0.64	146,9	0.25
Total	**18,102**	**-1.59**	**0.77**	**206,5**	**13,131**	**-1.41**	**0.73**	**193,1**	**0.18**

The difference between grade 3 and 5 student's achievement levels is more prominent in the urban area than in rural. Again, students' performance gap between urban and rural is found more in grade 5 than in grade 3. The difference is statistically significant at $p < 0.001$. The effect size is medium high in grade 3 (Cohen's $d = 0.69$) and high in grade 5 (Cohen's $d = 0.85$). As the institutional schools are located mainly in the urban location, they tend to raise the achievement

score of the urban areas. In the test, out of total, nearly 90 percent community schools with grade 3 and 5 were from the rural areas. However, among the institutional schools 56 percent in grade 3 and 64 percent in grade 5 are found to be located in urban areas. The high difference between rural and urban schools is due to the fact that the institutional schools are more often located in the urban areas. In any case, the difference is not good sign from the equality point of view particularly for the rural schools.

Language at Home and Student Achievement

In the grade 3 test, 41 percent of students have the mother tongue different from Nepali and it was 36 percent in 5th grade. The mother tongue reflects in many ways the difference in learning achievements, especially in lower grades. From the equality perspective, students' mother tongue should not be an obstacle in learning. But the reality is somehow different in schools, hence, inequality exists. First, the language is analysed dividing into two groups Nepali and non-Nepali speakers (table 10). Then it is analysed on the basis of 12 categorizes of home languages as reported in the background questionnaire (table 10 and figure 5).

Table 10. Achievement score of Nepali and non-Nepali speakers

Language	Grade 3			Grade 5			Difference
	N	Mean	SD	N	Mean	SD	
Nepali	11,610	-1.5	0.7	8,983	-1.3	0.7	0.2
Other	7,891	-1.8	0.8	4,987	-1.6	0.7	0.2
Total	**19,501**	**-1.6**	**0.8**	**13,970**	**-1.4**	**0.7**	**0.2**

Nepali speakers have higher achievement score than the non-Nepali speakers. When the Nepali speakers are in the upper grade, their language skill advances more than that of the non-Nepali speakers. Cohen's *d* also shows high effect sizes in both grades (0.38 in grade 3 and 0.47 in grade 5) and the differences are wider in grade 5. The difference between the grades is seen to be the same (0.2 standard units) with Nepali speakers and the others. This means that the some speakers other than Nepali do not reach the level to the Nepali speakers within the two years as their beginning level was lower and the same level of gap continues to remain in grade 5.

The table 11 and figure 5 do not include small language groups (less than 10 students) as well as not reported ones. From the perspective of students' home language, the scenario looks quite interesting. Magar and Awadhi/Bhojpuri speakers are better performer in both grades. However, it is noteworthy that the majority of the Magar speakers come from the Valley region – 83% in grade 3 and 89% in grade 5. The Nepali speakers are in 3rd position in grade 5 whereas in grade 3 they are in 4th position just after the Tamang. All the language groups are performing better within their respective language groups in grade 5 than in grade 3 except Rai and Tharu students. However, Urdu speakers are in the same ability level in both grades. The difference between the Limbu speakers in grade 3 and 5 is remarkable i.e., 0.8 standard units. The differences between

students in the highest and lowest performing language groups are statistically significant. The Partial Eta shows that the home language explains 9% ($\eta^2 = 0.086$) of variance in grade 5 and 5% ($\eta^2 = 0.051$) in grade 3 and the effect size in grade 5 is moderate or high (Cohen's $f = 0.31$), where as in grade 3 the effect size is medium (Cohen's $f = 0.23$). It shows the latent ability of the highest performing and lowest performing group is notable, mainly in grade 5.

Table 11. Home language and achievement score

Home language[1]	Grade 3				Home language	Grade 5			
	N	Mean	SD	CV		N	Mean	SD	CV
Magar	122	-1.0	0.47	47.0	Magar	181	-0.9	0.65	73.9
Awadhi/Maithili	35	-1.4	0.56	39.0	Awadhi/ Maithili	38	-1.3	0.59	47.1
Tamang	375	-1.5	0.52	35.9	Nepali	8,983	-1.3	0.70	54.4
Nepali	11,610	-1.5	0.73	49.3	Limbu	15	-1.4	0.57	40.7
Urdu	782	-1.5	0.66	43.3	Tamang	361	-1.4	0.54	37.4
Rai	143	-1.6	0.61	37.5	Urdu	631	-1.5	0.61	40.7
Sherpa	11	-1.7	0.40	23.8	Gurung	20	-1.5	0.93	60.5
Tharu	817	-1.8	0.72	39.4	Newar	424	-1.6	0.61	38.9
Newari	629	-1.8	0.66	36.0	Sherpa	29	-1.6	0.61	37.0
Gurung	83	-2.2	1.02	46.9	Rai	45	-1.7	0.59	35.0
Limbu	4	-2.2	0.65	29.9	Tharu	727	-2.0	0.70	36.0
Other	4,890	-1.8	0.87	47.4	Other	2,084	-1.7	0.76	44.9
Total	**19,501**	**-1.6**	**0.78**	**48.7**	**Total**	**13,538**	**-1.4**	**0.73**	**51.9**

1) The languages are ordered on the basis of the Mean.

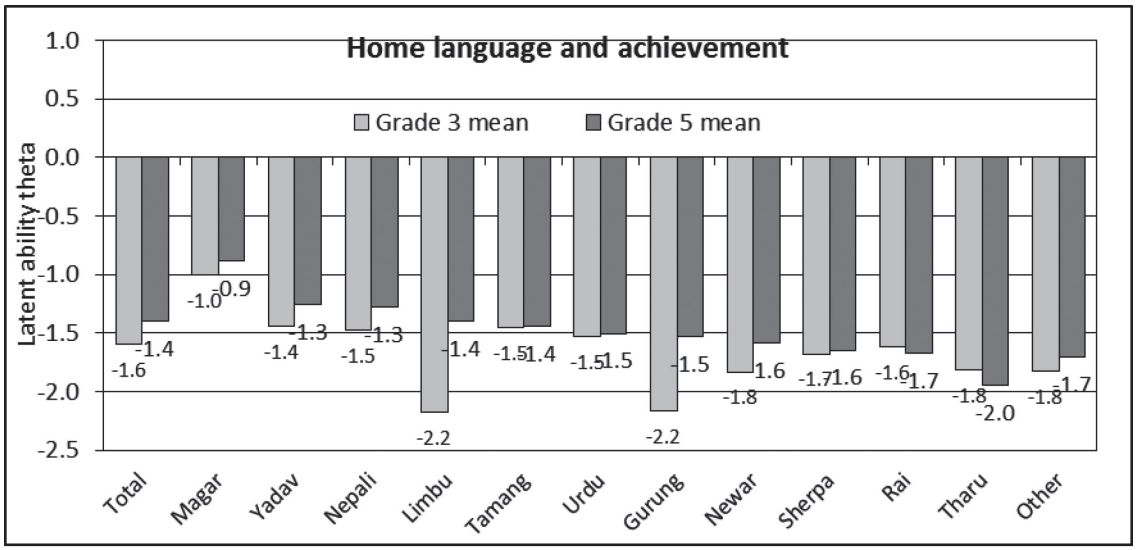

Figure 5. Home language and achievement

While analysing the home language based on the Development regions (table 12a and 12b), Tharu speakers are the low performer in the Eastern and Central regions in both grades. Further, Tharu speakers' ability is comparatively lower in grade 5 in all the regions. Except in Central region, Newari speakers are the low performers in grade 3 whereas in grade 5 in Eastern and Far-Western they are also found to be poorer. Although Magar speakers are the higher performer at the national level, in Central region they are recorded as the low performer. Similarly, Tamang speakers are also better in grade 5 compared to grade 3.

Table 12a. Development region wise home language and achievement score in grade 3

Dev. Region	Nepali	Magar	Tharu	Tamang	Newari	Urdu	Awadhi/ Bhojpuri	Rai	Gurung	Limbu	Sherpa
Eastern	-1.8	-1.9	-1.9	-1.6	-2.0	-1.8		-1.6	-2.4	-3.2	-1.7
Central	-1.6	-1.7	-1.8	-1.5	-1.4	-1.6	-1.7	-2.1	-2.3	-1.8	
Western	-1.4	-1.0	-0.8		-1.7	-0.9	-1.3	-1.1	-2.5	-1.9	
Mid-Western	-1.8	-2.5	-0.9		-2.1	-1.0		-2.9			
Far-Western	-1.7	-1.4	-1.6		-1.7				-2.2	-1.9	
Valley	-1.0	-0.9	-0.6	-0.9	-0.9	-1.1	-1.4	-0.7	-1.3		-1.6
Total	-1.5	-1.0	-1.8	-1.5	-1.8	-1.5	-1.4	-1.6	-2.2	-2.2	-1.7
Grade 3 =N	**11,610**	**122**	**817**	**375**	**629**	**782**	**35**	**143**	**83**	**4**	**11**

Table 12b. Development region wise home language and achievement score in grade 5

Dev. Region	Nepali	Magar	Tharu	Tamang	Newar	Urdu	Awadhi/ Bhojpuri	Rai	Gurung	Limbu	Sherpa
Eastern	-1.5	-2.2	-2.0	-1.8	-2.0	-1.4		-1.6	-1.2	-1.7	
Central	-1.4	-1.7	-2.0	-1.4	-1.6	-1.6	-1.3	-1.7	-2.5	-2.2	-1.7
Western	-1.2	-0.6	-1.2	-2.2	-1.5	-1.3	-1.2	-3.4	-3.3	-1.1	-1.5
Mid-Western	-1.6	-1.1	-1.1	-2.0	-1.7				-1.8		
Far-Western	-1.5	-1.0	-1.9		-1.5	-1.5					-1.9
Valley	-0.8	-0.7	-0.6	-0.6	-0.7	-1.3	-1.3		-0.4	-0.1	-0.3
Total	-1.3	-0.9	-2.0	-1.4	-1.6	-1.5	-1.3	-1.7	-1.5	-1.4	-1.6
Grade 5 =N	**8983**	**181**	**727**	**361**	**424**	**631**	**38.0**	**45**	**20**	**29**	**15**

Note: S*tudents less than 10 in the group are excluded from the analysis*

On the basis of table 12a and 12b Nepali speakers are seen comparatively better in grade 5 also by Ecological zones. Within the Nepali speakers, in Mountain zone, the students are poorer in grade 3. Tharus in Hills are better than in Tarai and their ability level is lower at grade 5 than at grade 3. There is no difference in latent ability of Urdu speakers in grade 3 and 5 except in Valley. This means either that the 3rd grades are very good or that the 5th graders are not gaining much.

Table 13a. Ecological zone wise home language and achievement score in grade 3

Eco Zone	Nepali	Magar	Tharu	Tamang	Newar	Urdu	Awadhi/ Bhojpuri	Rai	Gurung	Sherpa	Limbu
Mountain	-1.7	-1.4				-1.4	-1.9	-1.2	-1.6		
Hill	-1.6	-1.4	-1.7	-1.7	-2.4	-1.6	-1.1	-1.7	-2.7	-1.7	-2.3
Tarai	-1.6	-1.8	-1.8	-1.5	-1.8	-1.9	-1.7	-2.1	-2.1		-1.9
Valley	-1.0	-0.9	-0.6	-0.9	-0.9	-1.1	-1.4	-0.7	-1.3	-1.6	
Total	-1.5	-1.0	-1.8	-1.5	-1.8	-1.5	-1.4	-1.6	-2.2	-1.7	-2.2
N	**11,610**	**122**	**817**	**375**	**629**	**782**	**35**	**143**	**83**	**11**	**4**

Table 13b. Ecological zone wise home language and achievement score in grade 5

Eco Zone	Nepali	Magar	Tharu	Tamang	Newari	Urdu	Awadhi/ Bhojpuri	Rai	Gurung	Sherpa	Limbu
Mountain	-1.4	-0.5	-1.8	-1.5	-1.7	-1.4	-1.8	-0.6	-1.7		
Hill	-1.5	-1.8	-1.4	-1.2	-1.8	-1.6	-1.1	-1.7	-1.5	-1.7	-1.0
Tarai	-1.3	-1.8	-2.0	-1.5	-1.6	-1.9	-1.3	-2.4	-3.3		-1.6
Valley	-0.8	-0.7	-0.6	-0.6	-0.7	-1.3	-1.3		-0.4	-0.1	-0.3
Total	-1.3	-0.9	-2.0	-1.4	-1.6	-1.5	-1.3	-1.7	-1.5	-1.6	-1.4
N	**8,983**	**181**	**727**	**361**	**424**	**631**	**38**	**45**	**20**	**29**	**15**

Note : The highest frequencies are highlighted.

The language effect was found to be higher in both grades when divided into two groups 'Nepali' and 'Non-Nepali' speakers. The dataset shows the myth and reality of language in different direction. The myth is that Nepali speakers are better in Nepali language skills. But in reality, the results shows that the Magar and Awadhi/Maithili speakers are far better in Nepali. Nevertheless, their participation in test is very low in comparison with the Nepali speakers (Nepali speakers 60–66% and Magar 0.6–1.3% and Awadhi/Maithili 0.2–0.3%). It is also notable that the majority of Magars are from the Valley (83 – 89%). Further, Tharus from Tarai zone and Eastern and Central regions performed the lowest among the language groups. After Tharu, Newari speakers are noted as the low ability groups. It is noteworthy to understand that very few Newari speakers are from the Valley which is mainly populated by Newars. It hints that geographic factor explains more than language factor in students ability though there are many other factors associated with.

Ethnicity/Caste and Student Achievement

After home language, ethnicity/caste is another important factor associated with the achievement in Nepali language. Historically, the Brahmans and Chhetris had their higher level of educational attainment and other groups like Dalits and certain groups of Janajatis had their low participation in education. However, over the last decade the educational participation of those minority groups have increased dramatically though their learning quality is yet to ensure in order to create the equitable knowledge society in Nepal. The results concerning the ethnicity/castes and achievement are condensed in table 14.

Table 14. Achievement in the ethnicity/caste groups

Caste/Ethnicity	Grade 3			Grade 5			Difference
	N	Mean	SD	N	Mean	SD	
Brahman	2,026	-1.2	0.6	1,668	-1.0	0.7	0.2
Chhetri	3,537	-1.5	0.7	3,234	-1.3	0.7	0.2
Janjati	5,345	-1.5	0.8	3,913	-1.3	0.7	0.2
Madhesi	1,138	-1.7	0.7	906	-1.6	0.7	0.1
Dalit	2,152	-1.7	0.7	1,590	-1.6	0.6	0.1
Others	2,402	-1.7	0.7	1,968	-1.7	0.8	0.0
Total	**16,600**	**-1.5**	**0.7**	**13,279**	**-1.4**	**0.7**	**0.1**

The latent ability of all the caste groups differs from grade 3 to 5 from 0.1 to 0.2 in theta score, except the 'Others' group. Brahmans are high performers in both grades. Madhesi, Dalit and Others groups are below the national average in both grades. The overall difference between the groups is statistically significant ($p < 0.001$) and the effect size is medium in grade 3 ($f = 0.22$) and in grade 5 ($f = 0.28$). While dividing students according to their ethnicity/caste background, the division explains 5% and 7% of the student variation in grade 3 ($\eta^2 = 0.045$) and 5 ($\eta^2 = 0.072$) respectively.

The gender role varies depending on the caste groups in Nepal. Thus, it is also important to observe the ethnicity/caste and gender-wise performance (figure 6).

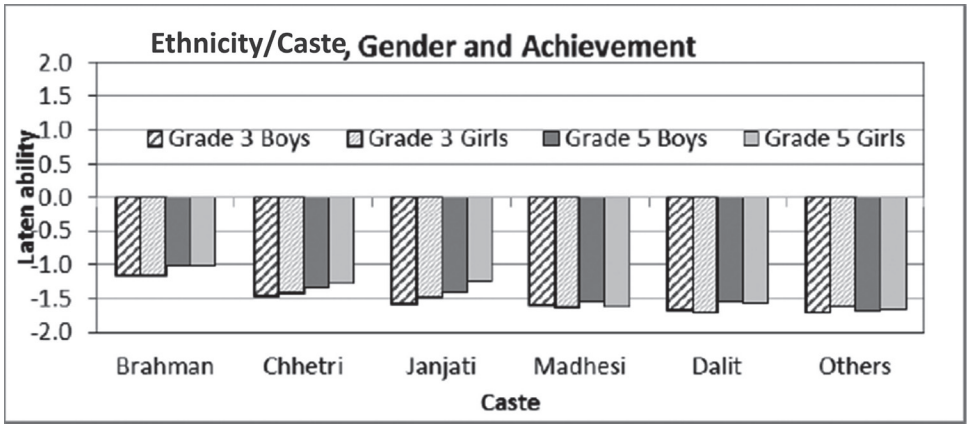

Figure 6. Ethinicity/caste, gender and achievement scores

Difference between girls and boys is not noticeable in Brahman and Dalit caste groups whereas it is notable in other groups at least in one of the grades. After Brahman, Janjati and Chhetri are outperforming the other groups, especially Janajati girls in grade 5 are outlying (0.2 theta score difference between girl and boys). Only in Madhesi group girls' ability is below the boys mainly in grade 5. The effect size of gender-wise, ethinicity/caste group in achievement is medium in grade 3 (both $f = 0.22$). Although effect size is medium in grade 5, they are slightly higher than grade 3 (girls, $f = 0.3$ and boys $f = 0.27$). The variation is also higher in grade 5 than in grade 3. In grade 3, ethinicity/caste explains nearly 5% in both girls ($\eta^2 = 0.047$) and boys achievement ($\eta^2 = 0.045$). In grade 5, ethinicity/caste explains 8% ($\eta^2 = 0.081$) in girls achievement and 7% in boys ($\eta^2 = 0.07$) in achievement.

Brahmans are ahead among the other groups whereas Madhesi and Dalit students'are found poorer in Nepali language ability. The caste-wise effect size of high performer and low performer is medium in both grades. However, the variation can be explained better in grade 5 (7%) than in grade 3 (5%). Gender difference is not seen in high achiever Brahman and low achiever Dalit. Except in Madhesi group, girls' language ability is better than boys in all caste and ethnic groups.

Comparison with the Objective Standards – CEFR Levels

The above comparison of the latent ability of students is based on *norm-referenced testing*: the test scores, presented in Standardized Normal scale, form a norm in which the different groups are compared. In languages, it is possible also to use *criterion-based testing*. It uses external criteria based on the standard setting procedures for language proficiency. Some of the well-

known standards are the CEFR, TOEFL, Cambridge Examinations, or IELTS. The CEFR was selected for the basis of the standard setting in Nepal because the procedures and standards are well-described in the literature (for example, in Takala, 2009; Kaftandjieva, 2004; Van der Schoot, 2009; FNBE, 2004) and the levels are transformable into other standards (http://en.wikipedia.org/wiki/Common_EuropeanFramework_of_Reference_for_Languages).

In this assessment – as well in 2011 – an adaptation of CEFR, more precise than the original, is used. The adaptation was prepared in the Finnish National Board of Education (FNBE) for assessing the language proficiency in school. The original scale is categorized into five groups: A1 level, elementary proficiency level (Limited communication in the most familiar situation); A2 level, first stage of basic proficiency (Basic needs for immediate social interaction and brief narration), B1 level, functional basic proficiency (Dealing with everyday life); B2 level, first stage of independent proficiency (Managing regular interaction with native speaker); C1 level, First stage of fluent proficiency (Managing in a variety of demanding language use situations). In the adaptation of FNBE, every level is further classified into two to three sub-levels (ERO, 2015, Chapter 2, Table 10)

Comparison of the Reading Proficiency in Nepali

On the basis of the proficiency levels of 19,501 students of grade 3, the average reader of grade 3 is at the CEFR level of A2.1 (27%) in reading. However, among 13,971 students of grade 5, the average 5th graders students are mainly located at two levels: A2.1 (27%) and A2.2 (26%) (fig. 7).

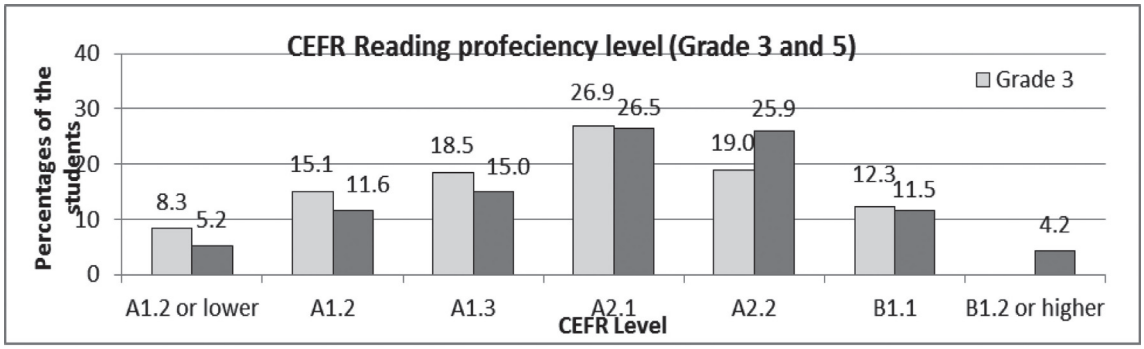

Figure 7. Distribution of reading proficiency in Nepali

Though at A2.1 level equal portions of students gain more or less the same language skill, figure 7 illustrates how the language proficiencies differ between grade 3 and 5 students. The description of an average reader of grade 3 and 5 in Nepal A2.1 level is as follows: "[(S)he] *can understand simple texts containing the most common vocabulary (personal letters, brief news items, everyday user instructions); can understand the main points and some details of a few*

paragraphs of text and can locate and compare specific information and can draw very simple inferences based on context." This A2.1 level is the first stage of the proficiency level. Besides, around 26% of 5th graders "*Can understand the main points and some details of messages consisting of a few paragraphs in fairly demanding everyday contexts (advertisements, letters, menus, timetables) and factual texts (user instructions, brief news items); can acquire easily predictable new information about familiar topics from a few paragraphs of clearly structured text and can infer meanings of unfamiliar words based on their form and context*." Overall, the A2.1 and A2.2 levels mean that the 3rd and 5th graders are able to do basic social interaction and brief narration in their daily lives.

Comparatively, more of the 3rd graders are still below the A2.1 level (42%) of language proficiency in reading than the 5th graders (32%), whereas the scenario is just reversed above A2.1 level – 31% in grade 3 and 42% in grade 5. Similarly, some 4% of 5th graders have B1.2 or higher reading proficiency than the average. It supports the fact that when students are promoted to upper grade, they also further advance their language efficiency accordingly. In reading proficiency in both grades, almost the same percentage of students, i.e., round 27%, has reached CEFR level A2.1 in language proficiency. In grade 3, most of the students remained below A2.1 level whereas it is just opposite in grade 5, most of the students have A2.2 or higher reading skills.

Comparatively major portion of Gurung (28%) and Limbu speakers (25%) have only the lowest level < A1.2 of reading proficiency in grade 3 (table 15a). The Gurung speakers (15%) are low in grade 5 as well (table 15b). However, the proportion of grade 5 students is less than grade 3. In grade 5, Sherpa (14%) and Tharu (13%) speakers have also low level of reading language proficiency. This < A1.2 level indicates that they can recognize only familiar words or phrases. When looking at the proficiency in totality of A1 level (i.e., combining A1.2 to A1.3), Limbu (75%), Gurung (62%), and Newari (59%) speaking students of grade 3 have limited communication level with elementary proficiency. Nonetheless, the same pattern is not seen in grade 5. It means that, in grade 5, some other language group students like Tharu (60%), Sherpa (59%), and Rai (53%) have low reading proficiency.

Table 15a. Reading proficiency levels in different language groups (%) in grade 3

CEFR Level	Nepali (N=11610)	Magar (N=122)	Tharu (N=817)	Tamang (N=375)	Newari (N=629)	Urdu (N=782)	Maithili (N=35)	Rai (N=143)	Gurung (83)	Sherpa (N=11)	Limbu (N=4)	Other (N=4890)
< A1.2	5.9	-	9.9	3.2	12.1	5.8	2.9	5.6	27.7	-	25.0	13.9
A1.2	13.5	1.6	17.1	7.5	20.3	14.8	14.3	14.0	15.7	9.1		18.9
A1.3	17.7	10.7	20.9	16.8	25.4	19.1	20.0	25.9	18.1	36.4	50.0	18.9
A2.1	27.3	27.0	29.0	35.2	27.0	29.5	20.0	24.5	26.5	27.3	25.0	24.5
A2.2	21.3	38.5	16.6	23.2	11.6	20.8	28.6	19.6	6.0	27.3	-	14.1
B1.1 or above	14.3	22.1	6.4	14.1	3.5	10.0	14.3	10.5	6.0	-	-	9.6

Table 15b. Reading proficiency levels in different language groups (%) in grade 5

CEFR Level	Nepali (N=8983)	Magar (N=181)	Tharu (N=727)	Tamang (N=361)	Newari (N=424)	Urdu (N=631)	Maithili (N=38)	Rai (N=45)	Gurung (N=20)	Sherpa (N=29)	Limbu (N=15)	Other (N=2084)
< A1.2	3.5	2.2	13.2	1.7	4.0	5.2	-	2.2	15.0	13.8	-	11.3
A1.2	10.0	3.3	23.5	10.5	15.1	12.2	10.5	22.2	5.0	17.2	-	14.5
A1.3	14.1	3.3	23.4	12.7	20.3	13.6	7.9	28.9	5.0	27.6	33.3	16.9
A2.1	26.5	13.8	23.5	27.1	28.8	30.4	34.2	35.6	35.0	31.0	6.7	26.3
A2.2	27.8	37.0	10.6	32.4	22.4	24.9	26.3	4.4	35.0	3.4	26.7	23.6
B1.1	13.1	24.9	3.3	12.7	7.3	10.6	15.8	6.7	-	6.9	33.3	6.1
B1.2 or above	5.1	15.5	2.5	2.8	2.1	3.0	5.3	-	5.0	-	-	1.2

Majority of students have A2.1 and A2.2 level of language proficiency except few language groups and are proportionately placed mainly in A1 sub-levels. The range in between 25% to 65% of grade 3 and 33% to 70% of grade 5 students of each language groups have first stage of basic Nepali language proficiency of A2 level (A2.1 and A2.2). They can do basic social interaction and narrations in Nepali language.

Most of the Magar speakers are far better reader (B1.1 and B1.2 level) in both grades followed by Maithili and Nepali language groups. However, Tamang speakers of grade 3 have also B1.1 level of reading proficiency. It indicates that students with B1 level of language proficiency can easily handle the everyday life interaction. Still both grade students need to advance the language skills to gain the independent proficiency level to speak and interact with the native speakers.

Except Magars in higher level and Gurungs in lower level of proficiency, different pattern in reading proficiency between grade 3 and 5 is noticed in the other language groups. Overall, Magars are far better than the other language groups in Reading. The situation of Gurungs indicates necessity of more support in their reading proficiency. Most of the 3rd graders have A2.1 level of reading skills and 5th graders are placed into two levels A2.1 to A2.2. Still their proficiency need to be upgraded from A2 level to enable them understand the demanding paragraphs needed in, for example, reading a newspaper in their everyday life.

In grade 3, around 9% of Dalit,'Others' caste and ethnic students are located in the Basic level of proficiency, i.e., at the CEFR level A1.This is comparatively a high percentage of students among all caste and ethnic groups (table 16a). However, in grade 5, the proportion of 'Other' and Madhesi groups are high at the level indicating lower than A1.2 proficiency (11% and 8% respectively) (table 16b). In grade 3, except the Brahmin students whose mode proficiency level is A2.2, the mode proficiency level in all castes/ethnic groups is A2.1. In grade 5, together with Brahman students, Chhetri and Janajati students also have slightly higher portion of students

at A2.2 level. Impressively, some students from all the caste and ethnic groups have acquired B1.2 or higher level of language proficiency. Overall, the Brahman students are ahead of all the remaining groups in Nepali language proficiency in both grades.

Table 16a. Reading proficiency levels in different castes (%) in grade 3

CEFR Level	Brahman (N=2026)	Chhetri (N=3537)	Janajati (N=5345)	Madhesi (N=1138)	Dalit (N=2152)	Others (N=2402)
< A1.2	2.2	4.7	7.5	8	9.2	8.9
A1.2	7.3	11.5	15.4	15.9	17.6	17.1
A1.3	12.6	18.7	17.8	18.1	21.9	20.9
A2.1	26.1	28.9	26.2	30.4	26.4	26.3
A2.2	28.5	22.2	19.7	17.9	15	17.9
B1.1 or above	23.3	14	13.4	9.7	9.9	8.9

Table 16b. Reading proficiency levels in different castes (%) in grade 5

CEFR Level	Brahman (N=1668)	Chhetri (N=3234)	Janajati (N=3913)	Madhesi (N=906)	Dalit (N=1590)	Others (N=1968)
< A1.2	1.5	3.4	3.8	7.7	4.8	10.5
A1.2	6.4	8.5	11.3	14.8	14.7	15.8
A1.3	8.6	14.8	13.9	16.1	18.9	19.9
A2.1	22.2	28.8	26.3	24.7	29.6	26.1
A2.2	31.9	29.5	26.7	23.7	24.7	18.4
B1.1	21.3	11.2	13.0	8.7	5.9	6.7
B1.2 or above	8.1	3.9	5.0	4.2	1.4	2.7

Brahman students are far better in both grades than the students from the other caste and ethnic groups. Around 9% to 10% of students from Others, Dalit and Madhesi caste and ethnic group can hardly read unfamiliar words and short messages in grade 3.

Writing Proficiency in Nepali

The average writer of 3rd and 5th graders is at the CEFR level A2.1 – 22% in grade 3 and 23% in grade 5 (figure 8), though there are quite a number of very good writers as well.

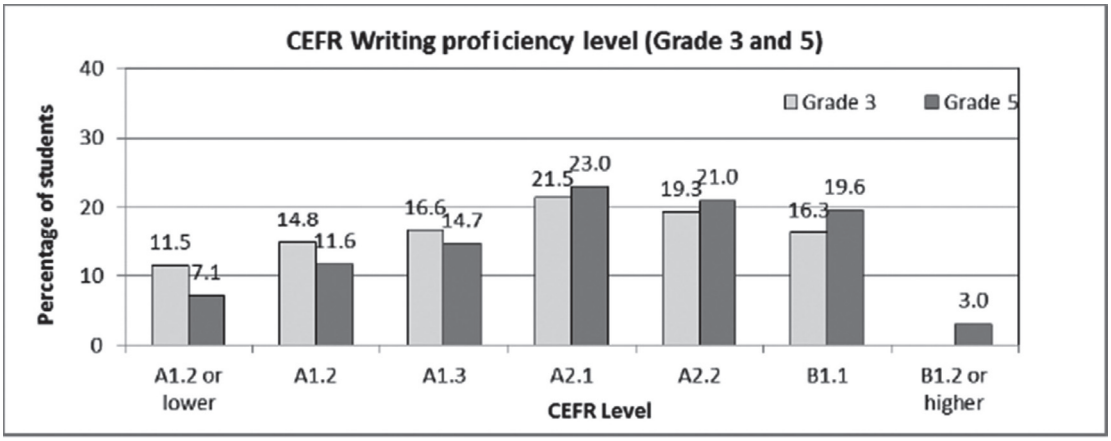

Figure 8. Distribution of Writing proficiency in grade 3 and 5

From the CEFR level A2.1, the proportion of the 5th grader is higher than 3rd graders (see figures 7 and 8). Like in reading proficiency, the proportion of 3rd graders at the level below A2.1 is higher than 5th grader in writing proficiency. Hence, the description of an average writer in Nepali, both in grade 3 and 5 , is as follows: (S)he] *can manage in the most routine everyday situations in writing* and *can write brief, simple messages (personal letters, notes), which are related to everyday needs, and simple, enumerated descriptions of very familiar topics (real or imaginary people, events, personal or family plans)* (ERO, 2015). The level is more or less similar as that of the reading proficiency. With this level an average 3rd and 5th grader writers cannot write everyday experience in their own. The writing skills in both grades is seen at the higher level than the reading skills and the proportion of students in higher levels is also higher in writing than reading (compare figures 7 and 8). Some 3% of students have B1.2 or higher level of writing skill in grade 5; their skills are very high in relation to their age.

In comparison to the reading proficiency, most of the students have better writing proficiency. Though in both grades most of the writers are located in A2.1 level, the proportion in level A2.1 and above is higher in grade 5 than grade 3. Unfortunately, this level of proficiency of writers is not sufficient to write even the everyday experience in their own. Impressively, some 3% of writers have remarkably high level (B 1.2 or higher) of writing skills in 5th grade. Table 17a and 17b shows the writing proficiency in Nepali language of grade 3 and 5 respectively in different language groups.

Table 17a. Distribution (%) of Writing proficiency in Nepali in grade 3

CEFR Level	Nepali (N=11610)	Magar (N=122)	Tharu (N=817)	Tamang (N=375)	Newari (N=629)	Urdu (N=782)	Maithili (N=35)	Rai (N=143)	Gurung (N=83)	Sherpa (N=11)	Limbu (N=4)	Other (N=4890)
< A1.2	8.5	1.6	14.0	2.4	12.6	8.7	2.9	9.1	31.3	-	25.0	19.3
A1.2	12.6		22.2	12.0	22.9	14.7	17.1	19.6	13.3	18.2	-	18.0
A1.3	15.5	4.9	22.9	19.7	20.7	14.1	17.1	22.4	15.7	45.5	75.0	17.9
A2.1	21.3	13.1	23.5	31.2	26.4	25.3	14.3	28.7	21.7	27.3	-	19.6
A2.2	21.8	35.2	12.2	24.0	11.8	22.8	31.4	10.5	15.7			14.5
B1.1 or above	20.2	45.1	5.3	10.7	5.7	14.5	17.1	9.8	2.4	9.1		10.7

Table 17b. Distribution (%) of Writing proficiency in Nepali in grade 5

CEFR Level	Nepali (N=8983)	Magar (N=181)	Tharu (N=727)	Tamang (N=361)	Newar (N=424)	Urdu (N=631)	Maithili (N=38)	Rai (N=45)	Gurung (N=20)	Sherpa (N=29)	Limbu (N=15)	Other (N=2084)
<A1.2	4.3	1.7	21.6	3.3	9.0	7.0	-	15.6	20.0	6.9	-	14.6
A1.2	9.1	6.1	27.6	9.4	17.2	14.1	7.9	11.1	-	10.3	13.3	15.7
A1.3	12.4	3.9	23.2	19.9	18.9	15.4	13.2	20.0	20.0	20.7	26.7	20.2
A2.1	22.7	11.0	14.9	35.5	26.4	29.0	31.6	24.4	25.0	20.7	33.3	23.4
A2.2	23.8	21.5	6.7	18.8	17.7	19.2	13.2	15.6	5.0	31.0	6.7	14.6
B1.1	23.6	47.0	5.0	11.6	10.4	14.3	31.6	13.3	30.0	10.3	20.0	10.7
B1.2 or above	4.1	8.8	1.0	1.4	0.5	1.1	2.6	-	-	-	-	0.7

Among the language groups, the Limbu students have A1.3 or lower level of writing proficiency in grade 3; while combining all the A1 levels (<A1.2 to A1.3), Limbu students(100 %) have the lowest writing proficiency followed by Sherpa (64%), Gurung (60%), and Tharu students (59%) (table 17a). In the grade 5, Limbu students are notably better. (table 17b). Contrary to grade 3, in grade 5, Tharu language group stands at very low level in writing:about 22% of the students did not reach even level A1.2. Tharu students are followed by Gurung (20%) and Rai students (16%). This level indicates the critical situation of those writers who are even not able to write simple letter and describe familiar topics reflecting everyday contexts.

Like the reading proficiency, Magar speakers in grade 3 are ahead also in writing with the mode of B1.1 level of proficiency(45%) followed by Nepali (20% at the level B1.1) and Maithili (17%) speakers. In grade 5 also, 47% of the Magar students are at the B1.1 level of writing proficiency which is the highest among the language groups. Magar students are followed by Maithili (31% at the level B1.1), Gurung (30%) and Nepali speakers (24%). The extreme case is noted among Gurung speakers in grade 5: while quite many of the Gurung students, 20%, are at the lowest possible level (< Al.2), at the same time 30% of the students are at quite a

high writing level (B1.1). The B1.1 level indicates that students can produce intelligible texts and write cohesively. From table 18b it is notable that there are some very high level (B1.2 or above) writers in grade 5 who can express very eloquently their thoughts, feelings and construct structured texts. In this category too, Magars (9%) are ahead of the other groups (4.1 or less). Like in reading, majority of Magar students are ahead also in writing proficiency. On the contrary, Limbu, Gurung, and Tharu students are at the lower level among the language groups.

Table 18a. Writing proficiency levels in different ethnicity/castes (%) in grade 3

CEFR Level	Brahman (N=2026)	Chhetri (N=3537)	Janjati (N=5345)	Madhesi (N=1138)	Dalit (N=2152)	Others (N=2402)
< A1.2	3.2	7.6	10.1	10.9	11.7	13.5
A1.2	7.3	11.8	14.3	18.4	17.8	15.7
A1.3	9.8	16.7	15.5	18.7	21.2	17.5
A2.1	19.1	23.0	22.2	23.5	22.0	22.1
A2.2	28.5	21.2	19.9	14.8	18.4	18.7
B1.1 or above	32.1	19.7	18.1	13.8	8.9	12.4

Table 18b. Writing proficiency levels in different ethnicity/castes (%) in grade 5

CEFR Level	Brahman (N=1668)	Chhetri (N=3234)	Janjati (N=3913)	Madhesi (N=906)	Dalit (N=1590)	Others (N=1968)
< A1.2	2.0	3.9	5.1	11.7	7.6	14.0
A1.2	5.5	9.5	10.8	15.8	14.2	15.5
A1.3	7.6	15.4	13.4	17.5	18.9	17.8
A2.1	17.6	24.7	23.1	23.2	26.9	22.9
A2.2	27.0	23.1	21.2	16.8	19.5	16.2
B1.1	32.0	20.6	22.9	12.9	11.9	12.5
B1.2 or above	8.3	2.8	3.5	2.1	0.9	1.0

In writing proficiency in grade 3, students from the Janjati, Madhesi, Dalit and 'Other' caste/ethnic groups are at the lower level compared to the students from the Brahman and Chhetri castes (table 18a). However, in grade 5, 'Other' and Madhesi groups have remarkably lower writing skills than the other groups: 12–14% of the students are at the lower level than A1.2. Proportionately, more Brahman students lie at the levels B1.1 and B1.2 (40%) than students from the other caste/ethnic groups (13–26%). After Brahman students come the Janajati (26%), Chhetri (23%), and Madhesi students (15%) respectively. Brahmin students have higher level of writing skills than other caste/ethnic groups. Of the students at the level 'lower than A1.2' in grade 5, the majority comes from 'Other' and Madhesi students. The disparity in language proficiency shows the inequality among the language and caste and ethnic groups.

Summary

- The differences in proficiencies between boys and girls are moderate, especially in grade 3 which is a positive sign towards equality between boys and girls. It is a noteworthy tendency that the girls are slightly out-performing the boys.

- The institutional school students outperform the community school as a whole and the gap of learning ability becomes wider in the upper grade. However, in both grades, the differences between girls and boys are pronounced more in the institutional schools than community schools. From the equality point of view, it is a positive aspect in the community schools that, although the ability level is not as high as compared to the institutional schools, the gap between boys and girls is narrow.

- For both grades, there is not much differences in different content areas. In comparison to the international standard of grade 4 students, the proficiency level is much lower among Nepali students. Girls perform better than boys in all the content areas in both grades, however, the gap widens more in grade 5. Lots of efforts need to put to improve the achievement level to reach the average international standard.

- The variation between the Ecological zones in language proficiency is remarkable. The Valley students outperform the others in all the areas. After Valley, Mountain zone is the second highest among the Ecological zones. The differences expands slightly within the years. Less or no gender gap is noted among zones.

- The differences between the Development regions are remarkable. In the Valley, which is incomparable with other zones, students outperform the students in all other regions. Eastern region is recorded as the lowest. Less or no gender gap is noted among regions.

- In some districts the gap between the learning outcomes of lower and upper grade is noticeable whereas, in some other districts, the difference is mild. Overall, the effect size shows a high differences between the high and the low performer districts. Though girls are better in the national level, when analyzing at the district level, in some districts boys outperformed the girls; example of these districts are Bardiya, Darchula and Saptari in grade 3 and Darchula, Jumla, Salyan, Kailali, Achham, Saptari and Mahottari in grade 5.

- The medium difference between rural and urban schools is seen due to fact that the institutional schools are located more often in the urban areas. In any case, the difference is not good from the view point of the equality for particular location.

- Magar and Awadhi-Bhojpuri speakers are found far better than Nepali speakers. Nevertheless, their participation in test is very low in comparison with the Nepali speakers. It is also notable that the majority of Magars are from the Valley (83–89%). Further, Tharus from Tarai zone and Eastern and Central regions performed the lowest among the language groups. After

Tharu, Newari speakers are noted as the low ability groups. It is noteworthy to understand that very few Newari speaking students are from the Valley which is mainly populated by them. It indicates that geographical factor explains more than language factor in students ability though there are many other factors associated with.

- Brahmans are ahead of all the group in language ability where as Madhesi and Dalits are seen to be poorer. The caste-wise effect size of high performer and low performer is medium in both grades. However, the variation can be explained better in grade 5 (7%) than in grade 3 (5%). Gender difference is neither seen in high achiever Brahman nor in low achiever Dalit. In all caste and ethnic groups, girls language ability is better than boys except in Madhesi groups.

Conclusion

The comparison of the latent abilities of grade 3 and 5 students shows the positive indication towards gender parity in Nepali language achievements, especially in grade 3. Unlike other tested two subjects –Mathematics and English, in Nepali, the girls are slightly out-performing the boys. It is noteworthy to understand that girl's language skill is better across the OECD countries as well (Francis & Skelton, 2009:3). Still, lots of efforts need to put to raise the language proficiency to the level at the international average. The difference in achievement between institutional and community school students urges for major intervention on the pedagogical process in the community school. If this issue is not addressed immediately, a high risk of widening the gap will continue to persist in the future to come and in the upper grades too.

The Ecological zone, Development region, location and district wise variations in language proficiency are remarkable. The Valley students outperform all the others in all areas. After Valley, Mountain zone is the second highest among the Ecological zones. Eastern region is the lowest achiever. The district level variances show the more precise results of students' performance levels. In some districts, the gap between learning abilities of lower and upper grade is noticeable (like Solukhumbu and Bhaktapur which are also high performers) whereas, in some districts, the difference is moderate. These differences show the regional imbalance in the development of the education system. Though girls are better at the national level, when looking at the district level, in some districts boys outperform the girls; example of these districts are Bardiya, Darchula and Saptari in grade 3 and Darchula, Jumla, Salyan, Kailali, Achham, Saptari and Mahottari in grade 5. The high difference between rural and urban schools indicates the probability of high ratio of institutional schools in the urban areas.

The language effect was found to be higher in both grades. The myth that the Nepali speakers are better in Nepali language skills does not match with the result of the grade 3 and 5. This can be seen in the fact that the Magar and Awadhi/Maithili speakers are far better in Nepali than the Nepali speakers, however, their participation in test is very low in comparison with the Nepali speakers (Nepali speakers 60–66% and Magar speakers 0.6–1.3% and Maithili speakers

0.2–0.3%). It is also notable that the majority of Magars are from the Valley (83–89%). Further, Tharus from Tarai zone and Eastern and Central regions performed the lowest of all the language groups. After Tharu, Newari speakers are noted as the low ability groups. It is noteworthy to understand that very few Newari speaking students in the sample is from the Valley even though it is populated mainly with Newars. It indicates that geographic factor explains more than language factor in students ability though there are many other factor associated with it. Anyway, the difference is not good indication from the equality view point for the rural schools.

Brahmans are ahead of all the group whereas Madhesi and Dalit students' Nepali language ability is seen to be poorer. The caste-wise effect size of high performer and low performer is medium in both grades. However, the variation explains more in grade 5 (7%) than grade 3 (5%). Except in Madhesi group, in all caste and ethnic group girls' language ability is better than boys.

Summing up from the CEFR comparison, generally, the international trend shows that the receptive skills (reading) are usually at the higher level than the productive skills (writing). However, this trend is not well fit in all the proficiency levels in the context of Nepal and especially when the diversity issues matter (e.g., various language and caste/ethnic groups). The findings indicate that though there are better readers than writers up to A2.2 level of language proficiency, above the A2.2 level (B1.1 or higher) the percentage of the better writers is higher than the readers. Most of the 3rd graders have A2.1 level of reading skills and major portion of 5th graders lie to two levels A2.1 to A2.2. However, majority of students' proficiencies need to be upgraded from A2 level to make them able to understand the demanding paragraphs of everyday life.

Inequality is prominent among the various language, caste and ethnic groups when it comes to reach the adequate level of language proficiency. The results are difficult to generalize among the language groups as the outperformer group is other than Nepali speakers, i.e. Magar. Most of the Gurung speakers are at the lowest level of language skill (A1 level) in reading and writing in both grades. It is seen that, except some language groups, the students, whose language is different other than Nepali, need lots of support and attention to improve their language skill. The students at the lowest level of proficiency level (< A1.2) with limited knowledge of words and phrases require education delivery also in mother language simultaneously at least to improve vocabulary and communication skills so that they can transfer language skill in further grades easily. Nevertheless, the challenge is that there exist wide varieties of languages and dialects groups scattered all over the country making difficult to identify their needs and address them accordingly. At present, the demand of Nepali and English language is high in the society, which is one of the challenging issues regarding the medium of language for instruction in general and particularly in using mother tongue as the medium of instruction.

The results of the caste and ethnicity implies certain groups are at the advantage position than others. Brahmans' reading and writing proficiencies are much better than others. The Dalits and 'Other' unidentified groups are at risk in language skills. Still around 9% to 10% of students from 'Others', Dalit, Madhesi caste and ethnic groups can hardly read unfamiliar words and short

messages as they have below A1.2 level of language skill, especially in grade 3. It means that lots of efforts need to put to provide equal opportunities to all background students to enable them reach the equal level of reading and language proficiency and to achieve the equality among all caste and ethnic groups.

To sum up in short, there are 23% of students who lie at the < A1.2 and A1.2 level of reading proficiency and 26% of them lie at that level in writing proficiency in grade 3. Upgrading their language efficiency up to the maximum level of A1 level, that is A1.3 has been an urgent need of the day to meet the objectives of the national curriculum for the grade 3.[3] Likewise, reading proficiency of 32% students and writing proficiency of 33% students of grade 5 need to be upgraded at least up to the A2.1 level so that they can easily meet the objectives of grade 5 curriculum.[4]

References

BPEP. (1995). *Basic and Primary Education Project Achievements Study*. Kathmandu: Author.

---- (1997). *The Effect of New Curriculum on the Achievement of Grade 4 Student.* Kathmandu: Author.

---- (1998). *The Effect of New Curriculum on the Achievement of Grade 5 Student*. Kathmandu: Author.

CERID. (1998). *Performance Levels of Grade V Students.* Kathmandu: Author.

CERSOD. (2001). *District Level Assessment. 2001.* Sanothimi: Department of Education.

Cohen, J. (1988). *Statistical Power for the Behavioural Science (2nd Edition).* Hillsdale NJ; Lawrence Erlbaum.

Curriculum Development Center (CDC). (2006). *The Primary level Curriculum: Grade 1-3*. Sanothimi: Author.

---- (2008). *The Primary Level Curriculum: Grade 4-5.* Sanothimi: Author.

EDSC. (1997). *National Assessment of Grade 3 Students*.. Kathmandu: BPEP.

3 *The 3rd grade students should be able to read simple vocabulary and sentences, messages, information, posters calendar, child literature; and can write simple and short sentences from their everyday experiences (CDC, 2006 [2063 BS], pp. 7-8).*

4 *The 5th grade students should be able to read simple messages, descriptions, letter, application, news, poster, calendar, simple child literature and students can explain pictures and objects by using grammar and vocabulary properly; can produce creative writing, independent writing and so on (CDC, 2008 [2065 BS], p. 7).*

---- (1999). *National Assessment of Grade 5 Students.* Kathmandu: BPEP.

---- (2001). *National Assessment of Grade 3 Students.* Sanothimi: DOE.

---- (2003). *National Assessment of Grade 5 Students.* Sanothimi: DOE.

---- (2003). *National Assessment of Grade 5 Students*. Sanothimi: DOE.

---- (2008). *National Assessment of Grade 5 Students.* Sanothimi: DOE.

---- (2008). *National Assessment of Grade 8 Students.* Sanothimi : DOE.

Education Review Office (ERO). (2013).w*here Are We Now? Student Achievement in Mathematics, Nepali and Social Studies in the Year 2011*. Kathmandu: Author

---- (2015). *National Assessment of Student Achievement (NASA) 2012 (Grade 3 and 5).* Sanothimi: Author

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

Fulbright Consultant (FBC) & Cooperative Hands in Restoration (CHIRAG). (2008). *A Final Report on National Assessment of Grade-V Students*. Kathmandu: DOE.

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

Moss, G., Francis, B. & Skelton, C. (2009).*Gender and Education Myth Busters - Addressing Gender and Achievement: Myths and Realities.* Nottingham: Department for Children, School and Families (DCSF) Publication.

OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing.

Teachers' Effect on Students' Learning Achievement

Rishi Ram Rijal, PhD, Associate Professor, Faculty of Education, TU

Abstract

This article presents the teacher effect on students' learning achievement brought about by the teachers' age, qualification, training, experience as well as their job types like permanent, temporary, etc. The data for this study were generated from 361 Mathematics and 343 Nepali teachers of 28 districts representing all ecological as well as administrative regions. The tools were test items categorized in three different sets and standardized with the help of professional experts, and calibrated with the pre-test. The main finding shows that the Mathematics and Nepali teachers teaching their respective subject disciplines do not have so much effect on the learning of their students. Even the academic courses, trainings and experience have not been that much effective.

Keywords: teacher effect, learning achievement, training, work experience

Introduction

After getting immersed into teaching learning activities for a certain period of time, the students have to demonstrate that they have acquired the knowledge and skills as specified by the curriculum prescribed for their level and grade, and perform some particular tasks such as writing an essay, carrying out an experiment, interpreting solution to a problem, etc. Acquisition of such knowledge, skills and values is known as the students' learning achievements. Darling Hammond (1993), an educationist, argues that these are the 'meaningful performances in the real world' setting that can closely capture the richness of what students understand about and how they can apply these knowledge and skills particularly. So the teachers in all classrooms are expected to be knowledgeable, skilled and accountable practitioners, who can raise the standard of achievements of all students by stimulating their interest in learning (Day, 2005). They are also expected to promote school-parent relationship and address the issues of classroom management, disciplines and students' need and expectations.

Teacher effect plays very significant role on the students' learning achievements if the teachers know how to arouse the interest of the students in the field of study; if they are the masters in the field and are in touch with the latest development in their subjects and if they are themselves the fellow travellers in the exiting pursuit of knowledge (Radhakrishnan as cited Vishala, 2006). Thus teachers' effect is the cause to bring out the best in students through the art of teaching that lies in understanding how people learn and what to do about it.

The ultimate goal of teaching any subject or skill is facilitating the learners to achieve higher performance in the respective field. It is known that talking of teachers to transmit knowledge to their students is not effective teaching. Similarly, students' attempt of hearing or simply attending in the classroom is not effective learning of the students. Teaching learning is a two way process of enabling the students to develop knowledge, skills and values in them. However, teaching in many places and cases has not been as effective as expected. There might be various factors affecting the effectiveness of teaching-learning and hindering students to gain high achievement in the performance. Among these various factors like school effect, student effect, situation effect, policy effect etc., teacher effect deserves high significance because none of the other elements can function well till the teacher element remains weak and ineffective which paralyze the whole education system. Teachers also vary a great deal in terms of the subject disciplines, age, qualification, training and experience, school type as well as their job type. In this context, it is very urgent to have the notice for the authority concerned regarding what kinds of effects the variables related to age, qualification, training, and experience along with the job type of teachers leave in the students' learning achievement. So, this research tries to seek the answer to this research question: What effect of teachers can be seen in the students' learning achievement with the variables of teachers' age, qualification, training, experience, and job type?

Methodology

The data analysed in this article were borrowed from National Assessment of Student Achievement (NASA) record 2013 which were generated from the samples of 361 Mathematics and 343 Nepali teachers of 28 districts representing the various ecological and administrative regions along with the Kathmandu valley as a separate region of Nepal. The sampled teachers were teaching at grade 8 ranging their age from below 25 to more than 45 and educational qualification from less than Proficiency Certificate Level (PCL)/grade 12 to Master level; job type from permanent to temporary; teaching experience from less than five years to more than 20 years; and their training and qualification from academic degree alone to ten month training plus Teacher Professional Development (TPD) along with their minimum required qualification. The data were generated through administering the test items prepared by the subject teachers and subject specialists with the help of professional experts of designing question items and were piloted to determine their difficulty level and confirm their standardization. The question items were designed as far as possible by covering the contents and skills as determined in their curriculum. For securing the genuineness of the students' achievement, those question items were presented in three different sets of equal standard and supplied with appropriate instruction for their administration. The gathered data were tabulated for analysis using Optical Mark Reading (OMR) sheet and interpreted with the results of One Parametric Logic Model (OPLM). Before analysing the data, two types of relevant literatures were reviewed: theoretical and empirical.

Theoretical Orientation

Teacher effect or teacher effectiveness can be understood by defining what the effective teachers know and do, by understanding the behaviours that effective teachers incorporate into their daily professional practice. In view of Cullingford (1995), the professional practice involves a deep understanding of subject matter, learning theories and student differences, planning, classrooms instructional strategies, assessment of student understanding and proficiency with learning achievement. It also includes a teacher's ability to reflect, collaborate with colleagues and continue ongoing professional development.

A variety of instructional planning activities, teaching strategies and materials are common in the repertoire of effective teachers. In Barry's (2007) view," an effective teacher has high expectations for students' learning, provides clear and focused instruction, closely monitors students' learning progress, manages remedial teaching for students who are seen week in their learning progress, and uses incentives and rewards to promote learning" (p. 4). They are highly efficient in their daily routines, enforce high standard for classroom behaviours, and maintain excellent personal interaction with the students.

Danielson (2009, p.5) has outlined the measures of relating to effective teaching organization into four domains: planning and preparation, classroom environment, instruction and professional responsibilities. But Marzano's (2007) model of teaching effectiveness has sequenced the different components as: (i) establishing learning goals, (ii) students interact with new knowledge, (iii) students practice to deepen understanding, (iv) engaging students, (v) effective classrooms management, (vi) effective student-teacher relationship, (vii) communicating high expectations for students, and (viii) effective standard based formative and summative assessment practices.

There is the challenge as to how to ensure that these practices are in every classroom and every teacher's repertoire of professional practice. According to Cullingford (1995), the challenge can be addressed by creating school culture, which aligns the practices of teacher hiring, expanding career opportunities for teachers' professional development and performance evaluation in a continuum of professional practice that uses the principles and behaviours of teaching effectiveness as its foundation.

Canadian Education Association (2009, p. 4) has established five core principles for effective teaching (as listed below) which provide a foundation for an effective teaching.

i. It begins with the thoughtful and intentional design of learning that engages students intellectually and academically.

ii. The work that students are asked to undertake is worthy of their time and attention, is personally relevant, and deeply connected to the world in which they live.

iii. Assessment practices are clearly focused on improving students' learning and guiding teaching decisions and actions.

iv. Teachers foster a variety of interdependent relationships in classroom that promote learning and create a strong culture around learning.

v. Teachers improve their practice in the company of peers.

Effective teachers have their own approaches to planning, designing, and implementing instructions and assessments. Their focus is on 'student learning' to inform their own teachers. So they know who their students are, their learning styles, their strengths and defects as learners. They are masters of subject matter; but more importantly, effective teachers always focus on their students' learning. Thus, effective teachers have a direct influence on enhancing students' learning. They not only make students feel good about their school and learning, but their work also results in increased students' achievement.

Professional qualities of teachers are associated with higher levels of student achievement. These qualities of effective teachers include formal teacher preparation training, varied experience and high expectation for themselves and their students (Roberts, 1997). They are caring, fair, and respectful; and they dedicate extra time to instructional preparation and reflection (Tickle, 2000). They are skilful enough to maximize instructional time via effective classroom management and organization. They enhance their own institution by varying instructional strategies, activities and assignments. So, they can present contents to students in a meaningful way that fosters understanding. They possess the quality of monitoring students' learning by utilizing pre- and post-assessment, providing timely and informative feedback and teaching materials to the students who cannot achieve mastery. This is how they can demonstrate effectiveness with the full range of students' abilities in their classrooms, regardless of the students' academic diversities. The students' learning achievement, according to Cohen, Manion & Morrison (2007, p.397), can be measured by:

i. Focusing on learning outcomes that require complex cognitive skills and students' performance;

ii. Selecting or developing tasks that represent both the contents and the skills that are central to important learning outcomes;

iii. Minimizing the dependence of the task performance on skills that are irrelevant to the intended purpose of the assessment task;

iv. Providing the necessary scaffolding for the students to be able to understand the task and what is expected;

v. Constructing task direction so that the students' task is clearly indicated; and

vi. Clearly communicating performance expectation in terms of the criteria by which the performance will be judged.

Teachers are the school's greatest asset. According to Day (1999), they stand at the interface of the transmission of knowledge, skills and values. They will be able to fulfil their educational purposes only if they are both well prepared for the profession and able to maintain and employ their contribution for the better achievement of all students. Professionally developed teachers are the effective teachers who can integrate their efforts to raise the standards of teaching, learning and achievement.

There are different factors that affect students' level of achievement. These factors include leadership commitment, school ethos, student effect, school culture, physical facilities, along with teacher effect. The number of study days, teacher–student ratio, number of students in the classroom, effective use of time and attention of students, teachers' and students' absence from the school, students' attitude towards the subjects and the teachers, parents' socio-economic and literary conditions, teaching methods adopted by the teachers and the schools, support to the teachers, etc. are also the affecting factors in the students' learning achievement (ERO, 2013). However, the most important one lies in the effectiveness of school teachers, without which all the other factors like leadership efforts, physical facilities, students' regularity, advanced curriculum, etc. become lame like the ineffective driver to a long journey by bus.

In Nepalese context, teachers are appointed to teach at different levels with the criteria of minimum qualification. For primary level teachers, at least the candidates must have School Leaving Certificate (SLC) level qualification, for the lower secondary level they have to pass grade 12 or Proficiency Certificate Level (PCL) or equivalent to it, and for secondary level they have to pass Bachelor's level or higher. So the age of the teachers also varies based on their qualification as well as the level. The primary teacher can enter into school teaching job just after their SLC, which can be 17 or 18 years as well. In many cases of Nepal, because of the shortage of qualified and capable mathematics teachers, just SLC passed teachers sometimes have to teach up to grade eight. In such a condition, they can run their classes but we cannot expect so much effectiveness from them to enable their students to secure high achievement in performance.

As the candidate for teaching job requires minimum qualification such as Proficiency Certificate, Bachelor and Master, their qualification also varies a great deal. Though there is some provision of compulsory teaching license, because of the government's flexible policy in this respect teachers are now appointed from different faculties and institutes, for example: Education, Humanities, Management, Science and Technology, etc. Academic qualification of the teachers is said to represent the level of subject knowledge, pedagogical skills and the familiarity of curriculum, psychology and evaluation procedures. Generally it is expected that the higher the qualification of the teachers, the better the students' learning outcomes. The academic qualification

is linked with some pre- and in-service training.

Trainings refer to activities directly focused on a teacher's present responsibilities and are typically aimed at short-term or immediate goals (Richards and Farrel, 2005, p.3). They are especially organized as the preparation for orienting into a first teaching position. In other words, they are given as the preparation to take a new teaching assignment or responsibility. They further add that teacher training involves understanding basic concepts and principles as a pre-requisite for applying them to teaching.

Woodward (as cited in Head & Taylor, 1997) in this regard, points out that training is compulsory, competency based, short term and one off in nature. It is skill or technique and knowledge based dealing with the external agenda, which is compulsory for entry to the profession. It is also top down, product or certificate oriented as the means of getting the teaching job or fulfilling the definiteness in teachers' skills.

Most training courses do not prepare teachers for the kinds of situations that involve on-the-spot-responses. Teachers develop skill in dealing with these situations through actual classroom experience, and are far more likely to explain their actions in terms of what was happening at the moment of decision and of feeling that they had about it, than to refer back to the received wisdom of their past training (Cullingford, 1995, p. 23). The training status of the teachers in Nepal seems that some of the teachers have only the academic degree like Bachelor in Education (B. Ed.), Master of Education (M. Ed.), Master of Arts (M. A.), Bachelor in Arts (B. A), and so on. Some others have received ten month training, and some of them even have either academic course plus Teacher Professional Development (TPD) or ten month training plus TPD as well.

Not all the knowledge that teachers bring to their teaching has been learned in formal training. Much of it accumulates from experience. The different teaching or training situations that they have been in, from childhood onwards, will all have left their mark on the kind of teachers they are today, and on their subjective picture of what good teaching and learning are. So, the teaching style is shaped not only by the training the teachers have received but also by contact with people like students, colleagues, family members and friends, by working with different kinds of materials and resources, by all the different kinds of experience that shape their life and make them the person that they are.

In community schools of Nepal, there are different types of teachers by their job status. Some teachers are categorized as 'permanent', who have faced the tests administered by Teacher Service Commission and got permanency in their job and enjoyed different facilities like promotion, different types of leaves, pensions, etc. as the public servant of different sectors of government offices. But in the cases of other types of teachers like 'temporary', these facilities are not provided. The difference in the provision of facilities can have negative effects in the perception and attitude of teachers.

Review of Empirical Research

Several research works have been carried out on teacher's effect or teacher effectiveness as well as the effective teaching practices in and out of the country. In the Nepalese context, while conducting large-scale assessments, called NASA, for grade 3, 5 and 8 by the ERO to assess the students learning achievement, it has been found that there are great differences in achievements between the students, school, districts and development regions, types of schools; cognitive skills, socio-economic backgrounds of the students and ethnicity, language and culture (ERO, 2015). In comparison of different studies (For example, 2008, 2011), the results in mathematics have slightly declined whereas they have been increased in the case of Nepali (ERO, 2015). It indicates that the students learning achievements have been highly influenced by the performance of effective teachers.

In the context of International studies in 'What did you do in the school today?' Canadian Education Association (CEA) has maintained that it has taken the initiative to have deigned to capture areas and inspire new ideas about enhancing the learning experiences of adolescents in classroom and schools. The research it carried in May 2009 states that preparing teachers for the 21st century requires a close look at what it means to teach and learn in increasingly networked, technology-rich , digital classroom. It further adds that schools and teachers need to thoughtfully and intentionally design learning environment and tasks in which teachers can explore issues that are relevant and develop pedagogies effective for a 'knowledge era'. This view is consistent with the statement made by Gilbert (2005) who states: 'former conception of knowledge, minds and learning no longer serves a world where we know is less important, what we are able to do with knowledge in different context and where our capacity of learning for outweighs the importance of our abilities' (p. 10).

As mentioned in Tucker, Stronge and Sanders' (nd.) research has been pivotal in re-asserting the importance of the individual teachers on student learning. One aspect of this research has been the addictive and cumulative effect on teachers' effectiveness on the students' achievement. Elaborating the research finding, Tucker and colleagues (nd., p. 5) state… 'The results of this study will document that the most important factor affecting the students' learning is the teacher'. In addition, the result shows wide variation in effectiveness among teachers. The implication of his finding is that many things can be done to improve education by improving teacher effectiveness rather than any other single factor. Thus, the effectiveness in teachers does not come automatically along with their initial teacher preparation courses. The expertise in them stems up through such attributes like their age, qualification, training, experience as well as the type of their job and role status in the school community. I have developed the following chart to show the connection of teachers' effect on students' learning achievement.

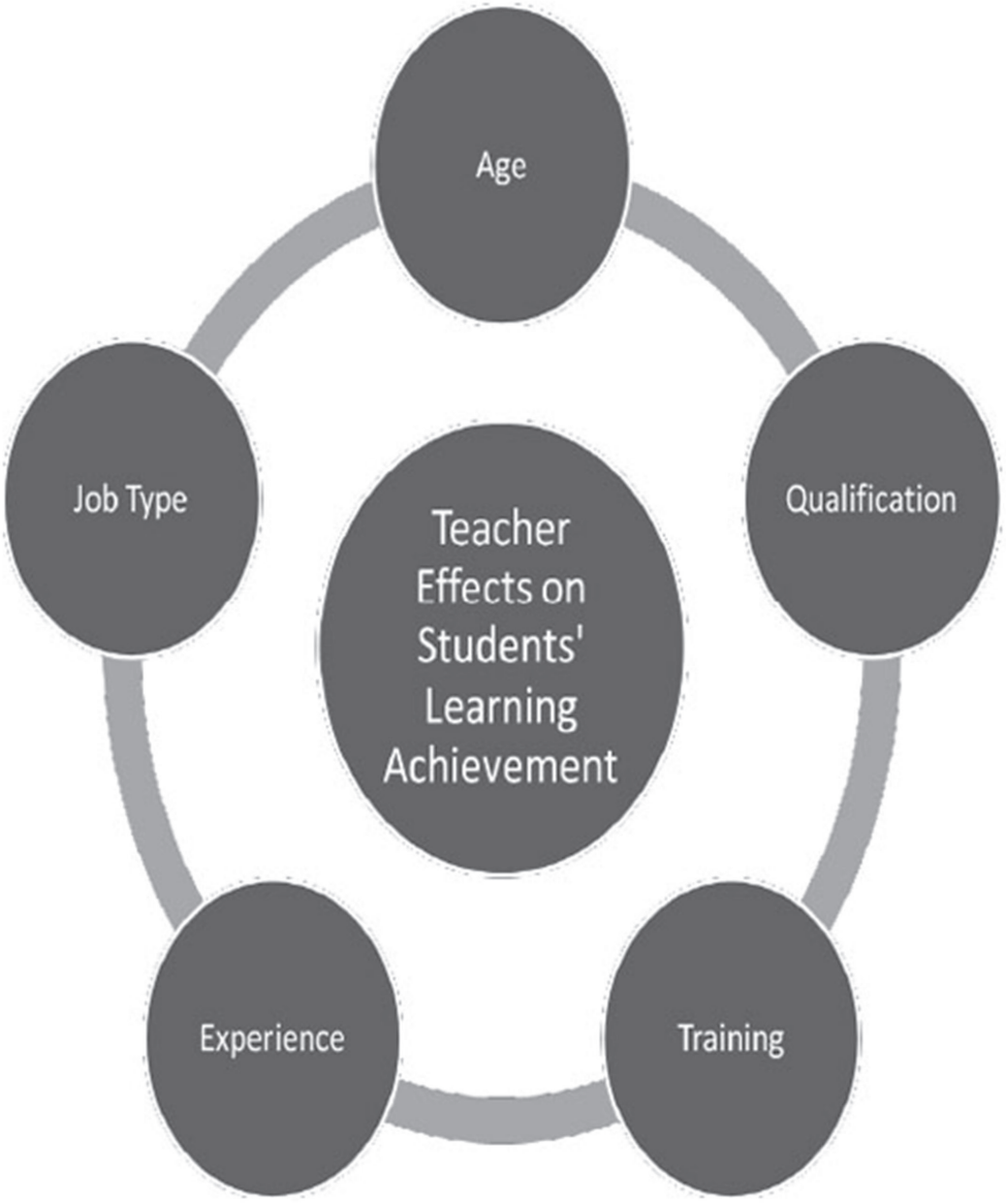

Figure 1. Teacher effects on students' learning achievement

Analysis and Interpretation of Data

The data through survey questionnaire have been analyzed by using the mean score and presented in the tabular and figurative forms putting the data on Mathematics teachers and describing their seen trends along with the interpretation based on the above described theoretical background. The data have been presented under the five personal attributes such as age, qualification, training, experience and job type of the teachers and the students' learning achievement.

Teachers' age and students' achievement

The teachers' ages have been categorized into four groups ranging approximately ten years from below 25 to more than 45. The mean achievement of the different age groups of teachers can be observed in this table.

Table 1. Age of the teachers and students' learning achievement

Age group	Mathematics		Nepali	
	Teacher N	Mean	Teacher N	Mean
Less than 25	99	38	35	53
25 to 35	153	37	130	49
35 to 45	84	28	103	50
More than 45	25	36	75	50

Source: NASA 2013 database

The information in the table depicts that the mean score of mathematics and Nepali teachers between their ages of 35 to 45 years is the lowest one but it is slightly higher in the teachers of below 25 years compared to the other two groups. This result reminds us a famous saying which states that, for some people, eighteen years of experience is one year's experience repeated seventeen times (Tsui, 2003 p. 13).

In the case of Nepali teachers, the data depicts that the mean score of the teachers' effectiveness does not show so much variation although it is higher in the age group of below 25 years by one point. The ANOVA test result also indicates that the age attribute does not have any significant effect on the teachers' performance to enable their students to secure high achievement. On the whole, the highest mean score in Mathematics is only 38 in comparison to the average school; this score is not high enough to show the positive effect of teachers on students' learning achievement. The mean score of the Nepali teachers also does not show very high score (53) which is only average, and it does not indicate so high teacher effect on students' learning achievement.

Qualification of teachers and students' learning achievement

The data on qualification of the teachers have been categorized into four groups specifying as 'less than PCL', 'PCL or equivalent', 'Bachelor' and 'Master'. Though the expected qualification for teaching Mathematics and Nepali is PCL or equivalent for teaching Mathematics and Nepali at lower secondary grades, some of the teachers were underqualified and some others had additional qualification like Master's degree in the respected disciplines.

Table 2. Qualification of teachers and students' learning achievement

Qualification	Mathematics		Nepali	
	Teacher N	Mean	Teacher N	Mean
Less than PCL	2	19	5	47
PCL or equivalent	108	32	78	44
Bachelor	170	34	149	51
Master	82	43	108	54

Source: NASA 2013 database

The data shown in the table shows that the teacher effect of under PCL teachers is very low (only 19 point) but it is somehow higher (43 point) of the Mathematics teachers holding Master Degree. The result of ANOVA indicates that it is statistically significant (that is 0.05). However, in the case of Nepali teachers the mean score of the teachers holding less than PCL degree is slightly higher than the mean score of the teachers holding PCL or equivalent degree. This does not show any positive co-relation between the academic qualification and the positive effect on students' learning outcomes. On the whole, as in both subjects – Mathematics and Nepali, the mean scores are not more than 43 and 54 respectively, so, we cannot claim that there is much significant effect of teachers on students' learning achievement. The statistically significant result of ANOVA test can also be seen in the figure 2 presented below.

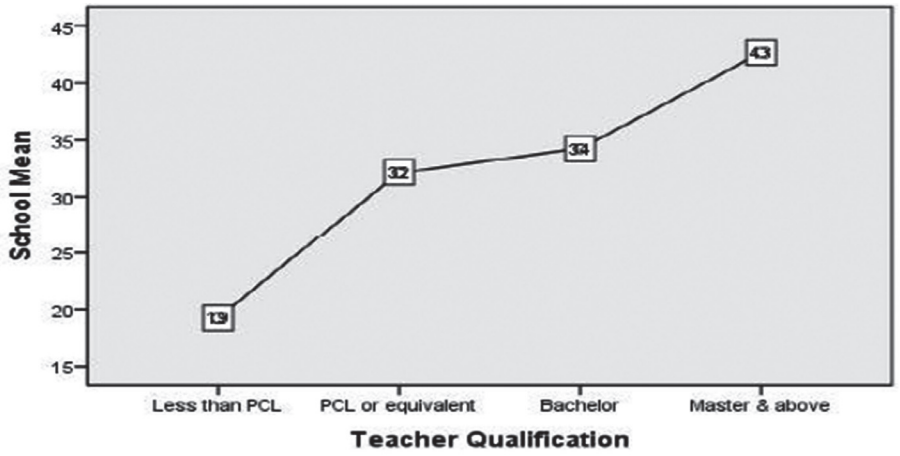

Figure 2. Teachers' qualification and mean score

The status of teacher training and students' learning achievement

Teacher training is a means of improving teachers' performances and increasing students' learning achievement. It has been provided to the teachers of various levels by the government and its line agencies by investing a big sum of money and efforts. They have been aided by different multinational organizations and companies. The training status has been categorized here in terms of academic degree like B. Ed., M.A., academic degree plus ten month training like one year B.Ed.; Teacher Professional Development training organized by NCED and different training centres and training hubs. These TPD programs were in total divided into three phases and completed in a phase-wise manner within five years' time following the given procedures of face-to-face mode, school induction and feedback giving and certification, and the ten month training plus TPD program.

Table 3. Teachers' job type and students' learning achievement

Training	Mathematics		Nepali	
	Teacher N	Mean	Teacher N	Mean
Academic degree only	153	35	152	52
Only ten month	26	30	28	44
TPD	34	33	36	41
Academic + TPD	54	29	51	45
Ten month + TPD	13	25	26	48

Source: NASA 2013 database

The presented data in the table reveals the fact that in the case of Mathematics teachers, except the TPD training holders, the trend seems to be regressive one because the mean score goes down from academic degree (35) to ten months + TPD training holders (25). This data indicates that the trainings organized for the teachers have not left any positive influence to bring change in teachers' professional behaviours.

In the case of Nepali teachers, the highest mean score is 52 point which belonged to the teachers holding academic degree only; and the lowest mean score (41) belonged to the teachers holding TPD training only. This data indicates that neither the TPD training (which was regarded as a so-called demand driven training) nor other pre-service (e.g. ten month) training or other in-service trainings have contributed to leave positive effect on students' learning achievement. This result is very consistent with the Fulbright Report (2006) that the effectiveness of teacher training is always questionable in the Nepalese context.

Teaching experience and students' learning achievement

Teaching experience, one of the influential teacher professional learning means, here has been presented into five distinct categories dividing the groups by assigning five years interval ranging from less than five years' experience to more than 20 years' experience.

Table 4. Teaching experience and students' learning achievement

Experience Year	Mathematics		Nepali	
	Teacher N	Mean	Teacher N	Mean
Less than 5	85	38	94	49
5 to 10	39	38	63	52
10 to 15	18	40	29	52
15 to 20	18	33	35	52
More than 20	11	37	23	49

Source: NASA 2013 database

The presented and calculated data with mean score from the table shows that the highest mean score the mathematics teachers secured is only 40 and the lowest is 33. The highest mean belonged to the teachers holding 10 to 15 years' experience. This data indicates that during this phase teachers have to face the state of diversification (Huberman, 1993, p.13) in which phase some of the teachers are more active but some teachers are the sufferers of the assessment and turn to conservatism. Here, Mathematics teachers have become more active rather than more conservative ones. However, the teachers holding 15 to 20 years' experience (33 mean score) seem to be more conservative because their mean score has become regressive one.

In the case of Nepali teachers, the mean score is nearly constant, that is 49, 52, 52, 52 and 49 respectively. The experience of these teachers does not have varied influence in the teacher effect on students' learning achievement. On the whole, here too, the mean points do not show much effectiveness in their teaching behaviors which could bring positive effect on the students' learning outcomes.

Teachers' job type and students' learning achievement

In the community schools of Nepal, the teachers with various status work in teaching job together. These teachers have been categorized as permanent, temporary including temporary position *(Rahat*) and per capita based (PCF), and others. Under the category of 'others', teachers have also been working as School funded, Volunteer, Substitute, etc.

Table 5. Job type of the teacher and student achievement

Job type	Mathematics		Nepali	
	Teacher N	Mean	Teacher N	Mean
Permanent	57	30	96	50
Temporary	90	36	64	53
Rahat quota	64	27	77	43
PCF/other	130	38	92	51
	341		329	

Source: NASA 2013 database

The information presented in the table shows that the *Rahat* teachers in both subject disciplines were weak because the mean point for both subjects (Mathematics and Nepali) has remained 27 and 43 respectively. In both subjects, the PCF and other teachers were found having slightly higher influence on their students' learning outcomes. In comparison to permanent teachers of both subjects, the temporary teachers had higher influences on their students' learning. The fact presented in the table indicates that in the appointment of *Rahat* teachers, there is intervention of the local people; and the teachers working on temporary basis are more sincere in their assigned duties.

Findings and Discussion

Based on the analysis and interpretation of the data through survey questionnaire, the main finding of this study is that teachers' effect on students' learning achievements does not show so much positive trend because the mean score does not exceed 53 in all attributes(age, qualification, training, experience and job type) . Instead, it has gone down up to 19 in the case of mathematics teachers holding less than PCL qualification.

i. As mathematics teachers between the age of 35 and 45 years have the lowest mean (28 only), they do not have any effect on students' learning mathematics.

ii. Though there seems to be corresponding teacher effect of academic qualification ranging from less than PCL (with the mean score of 19), PCL or equivalent (32), Bachelor (34) and Master (43), these scores do not represent significance on the whole.

iii. The teacher who were employed with per capita funding like part-time and other teachers have slightly higher mean score (38 and 51) in both Mathematics and Nepali teachers in comparison to permanent teachers. In case of temporary teachers, the effect of local intervention can be clearly felt because the mean score in both subjects is lower in comparison to other groups of teachers.

iv. Regarding the training status and students' learning achievement, the TPD training, which is popularly known as demand driven training, has the weaker effect on teachers because the mean of mathematics is 25, 29 and 33 in the case of ten month plus TPD, Academic plus TPD and TPD respectively.

The lowest mean score showing the Mathematics teacher effect on students' learning achievement belongs to the teacher holding 25 to 20 years' experience. This is the effect of diversification stage as mentioned by Dreyfus & Dreyfus (1987).

The above results show that Nepali as well as Mathematics teachers teaching these subject disciplines at grade eight do not have so much effect on their students' learning achievement. There might be several factors to prevent them from the likely effect, but the most important might be the lack of the ability of setting high expectations to inspire, motivate and challenge students, problems in promoting good progress and outcomes in terms of students' learning achievement, and lack of well-structured planning and teaching of lessons. They are also unable to adjust teaching to respond to the strengths, weaknesses of all students, make productive use of evaluation effectively to ensure a good and safe learning environment. These teachers cannot expect from their students what they themselves cannot do. The implication of this proposition is that unless the teachers are involved themselves in continuous life-long learning, they cannot lead their students to gain higher achievement in their performance.

Unless teacher trainings can be thought as the component of teacher development and till it remains top-down, decontextualized, one-size-fits-all and prescriptive, the effect of teachers on their students' learning achievement cannot be expected. So school based and demand driven trainings should be organized following the bottom-up approach, after analysing the teachers' problems, needs, demands and interest through informal talking with them, listening to their grumbles, observing their classroom teaching and analysing their teaching portfolios. Teaching portfolio can be one of the most reliable means of analyzing their progress and weakness though many teachers and administrators stand against it. The reason they pose against it is the fact

that they feel it brings extra burden in them while collecting the evidences of their progress of the years. Similarly, there must be the formal provision of teacher induction in the beginning of their entry in the career with experienced teacher's mentoring and recommendation for their permanency in the teaching profession. The strong and effective monitoring and critical feedback should also not be undermined for the improvement of students' achievement.

References

Barry, R. A. (2010). *Teaching Effectiveness and Why It Matters*. Retrieved from <http> chalkboardproject.org/wp.../teacher-effectiveness-and-why-it-matters.pdf

Cohen, L., Manion, L., & Morrison, K. (2007). *A Guide to Teaching Practice.* London, New York: Routledge Falmer.

Cullingford, C. (1995). *The effective teacher*. London: Cassell.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria: ASCD

Danielson, C. (2009). *Talk about Teaching: Leading professional conversations.* Corwin: Thousand Oaks.

Darling-Hammod, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation*. Phil Delta Kappan,* 74, 10, 752-61.

Day, C. (1999). *Developing teachers: The challenges of lifelong learning*. London and Philadelphia: Falmer Press.

Day, C. (2005). *A passion for teaching*. London/New York: Routledge Falmer.

Dreyfus, H. L. & Dreyfus, S. E. (1987). *Mind over machine*. New York: Free Press.

ERO. (2013). *Where are we now? Student achievement in Mathematics, Nepali and Social Studies in 2011.* Kathmandu: ERO.

ERO. (2013). *Where we are now? Student achievement in Mathematics, Nepali and Social Studies in 2011 Report. Kathmandu:* Authors.

ERO. (2015). *Report of National Assessment of Student Achievement 2013 (Grade 8: Mathematics, Nepali and Science).* Sanothimi: ERO.

Gilbert, J. (2005). *Catching the knowledge wave? The knowledge society and the future of education*. Willington, N.J: NZCER Press.

Head, K. & Taylor, P. (1997). *Readings in teacher development*. Oxford: Heinemann English language teaching.

Marzano, R. (2007). *The art and science: A comprehensive framework for effective instruction*. Alexandria: ASCD.

Richards, J. C. & Farrel, S. C. (2005). *Professional development for language teachers*. London: Cambridge University Press.

Roberts, J. (1997). *Language teacher education*. London, New York, Sydney, Aukland: Arnold.

Tickle, L. (2000). *Teacher induction: The way ahead*. Buckingham: Open University Press.

Tsui, A. (2003). *Understanding expertise in teaching*. New York: Cambridge University Press.

Tucker, P. D. & Stronge, J. H. (nd). *Linking teacher education and student Learning*. Retrieved from <http://www.ascd.org/publications/books/104136/chapters/The-Power-of-an-Effective-Teacher-and-Why-We-Should-Assess-It.aspx>

Vishala, S. M. (2006). *Guidance and counselling*. New Delhi: S. Chand and Company Ltd.

Implication of Institutional Performance in Student Learning Achievement

Kul Prasad Khanal, Under Secretary, Education Review Office

Abstract

Accountability is measured against performance standards based on specified job descriptions of individuals and organization. Reports published by Education Review Office on institutional performance audit provide educational institutions with the feedback for improvements. On this ground, institutional performance geared to improved learning achievement seems to be one of the central concerns of education system. To this end, the purpose of this article is to introduce the concept of performance audit considering its linkage with student achievement. Similarly, the article highlights some international practices and relevant theories in the light of Nepalese practice of performance audit. The aim is to establish institutional performance and student assessment as concepts complementary to each other rather than being mutually exclusive notions. To reach the conclusion, secondary data from performance audit and NASA reports backed up with documents study of related literature have been used. The data shows that institutions having high mean score in management performance demonstrated remarkably low status in student assessment –lacking a consistent pattern of relationship between the two variables. Inter-institutional linkages between District Education Office and Resource Centre are also not consistent. Finally, it has been concluded that stakeholders' sensitization on performance audit through a coherent scheme of capacity development coupled with case studies of extreme results are imperative.

Keywords: performance audit, student assessment, performance indicators

Introduction

Job descriptions and responsibilities are the basic performance indicators leading to institutional accountability at large. Performance is reflected in the satisfaction of clients; and more specifically, in the learning achievement of students in case of education system. Measurement of performance against specified criteria is understood as auditing. School Sector Reform Plan (SSRP) of Nepal (Ministry of Education [MOE], 2009) and other policy documents envisioned an efficient and effective delivery of education services through demonstrated performance of individual and the education system. As per the mandates given, Education Review Office (ERO) has taken initiatives in auditing the performance of schools as well as institutions under the MOE. In this context, this article discusses the conceptual understanding of performance auditing with a review of some international practices and the theoretical background; seeks implication of performance audit after making linkages of institutional performance audit and student learning achievement with reference to the recent reports prepared by ERO. The analysis is based on secondary data collected from recently published reports. Performance audit has two-fold

methodologies. At first, the institutional auditing was accomplished nationwide with purposive stratified sampling procedure using openly selected external auditors. Methods used in data presentation and analysis were document study, content analysis and simple comparisons using means and correlations. Theoretical context setting and conceptual understanding of the theme was based on document study of related literature. Secondly, in this article, the performance level of District Education Offices are based on the results of performance audit carried out by ERO and the student learning achievement of districts are based on the national assessments of student achievement conducted by ERO. Data taken from the reports were arranged and analysed making a linkage between performance audit scores and that of National Assessment of Student Assessment (NASA) results in different subjects across grades and years. Effort has been made to draw pedagogical, managerial and governance-related implications amidst two types of reports. Performance audit data was taken from those eight districts that participated in 2014 audit and linkage was identified between institutional performance and assessment results. Content analysis, simple comparison using means and correlations are the procedures of analysis and interpretation. Educational institutions like District Education Offices (DEOs) and Resource Centres (RCs), schools are used as unit of analysis and have been considered as the representative samples for showing linkages and comparisons.

Conceptual Understanding of Performance Audit

The concept of performance audit dates back to 1916 when United States of America established the 'Bureau of Efficiency' to tackle waste in the US federal government (Lee, 2006 as cited in Talbot, 2010). According to Arter (2003), audit in early days was used to record cargo on a ship by listening to the crew who called out the items and quantities making sure that taxes on that cargo would be properly recorded. From the beginning, auditors were concerned with control and compliance (Arter, 2003). Performance audit, thus, involves assessing whether government policies, programs and institutions are well managed and running economically, efficiently and effectively (Lonsdale, Wilkins, & Ling, 2011). Arter (2003) defines an audit as "a systematic, independent and documented process for obtaining audit evidence and evaluating it objectively to determine the extent to which audit criteria are fulfilled" (p. 18). Likewise, a technical audit has been described as a 'systematic and independent process for obtaining evidence and evaluating this objectively to determine the extent to which needs or expectations are fulfilled (http://www.nvc.gov.np).

Performance audit has been understood as an independent examination of a programme and procedures of an organization to assess whether it is achieving economy, efficiency and effectiveness in the employment of available resources (http://www.audit.nsw.gov.au). In most countries, the external audit bodies carry out performance audits of government activities. It is an independent examination of the efficiency and effectiveness of government undertakings, programs or organizations, with due regard to economy leading to improvements. As such, the purpose of audit is to examine how well the planned or designed criteria are met by the organizations within the timeline and budget and to understand how well the organization is doing

in accomplishing its mission and goals. Byrne (n.d.) argues for evidence based critical analysis with regard to performance.

Discussing on audit process, Arter (2003) categorizes compliance and performance audit by purpose. A compliance audit looks for conformance to a set of rules giving assurance that activities have been performed properly. Audits may also be classified by system perspective as product, process, and system audits (ibid.). A product audit is quite similar to an inspection where the completed item or task is examined to see the required characteristics. Products come from processes. A process then is a set of inter-related and interacting activities, which transforms inputs into outputs. A system is a group of processes all working together to achieve a common goal. Likewise, Ng (2002) relates performance audit to the issues of accountability and thus presents three categories of performance: economy, efficiency and effectiveness. These categories, in turn, create a "five-tier ladder of accountability for determining the performance of management. These are: (i) accountability for probity and legality; (ii) process accountability; (iii) performance accountability; (iv) program accountability; and (v) policy accountability". (p. 111).

Quality audit always requires quality auditors. Auditors need to possess certain qualities and competencies for auditing process. Discussing the qualities of an auditor, Arter (2003) identifies four basic rules of audit: (i) Audits provide information for decisions; (ii) Auditors are qualified to perform their tasks; (iii) Measurements are taken against defined requirements; and (iv) Conclusions are based on fact. In similar vein, Education Review Guidelines (MOE, 2067 BS) highlights the selection criteria of the auditor, stating that an auditor should (i) have a master's degree in education; (ii) be an ex-professional of education with an experience of at least under-secretary post; (iii) be a professor/associate professor of education having at least five years of working experience in a university. In tune with this, Arter (2003) suggests the need for mechanical, intellectual, and emotional skills for auditors. Emotional skills include individual and group relations, confidence, empathy and patience. The mechanical skills deal with data gathering and analysis, such as sampling, tracing cause-and-effect relationship, etc. Likewise, intellectual skills are used to communicate with others through writing, planning, speaking and organizing skills (ibid.). Auditing is seen as a skill, which could be developed through rigorous practice. Selection of auditors based on specified criteria along with intensive orientation course for auditors is indicative of ERO's commitment to ensuring basic competencies of external auditors in Nepal.

Theoretical Perspective on Performance Audit

Positivism, post-positive, and constructivism are major paradigm perspectives that provide theoretical approach to ontology, epistemology, and methods of academic discourse (Thietart et al., 2001). Ontology refers to the nature of reality. Epistemology refers to how we know about that reality; and methods refer to the technical issues of how we attempt to observe phenomena. Viewing from positivist tradition (Talbot, 2010), public organizations have the real and actual

levels of performance that exist independent of any observation. From this perspective, the best way to know about organizational performance is to disaggregate individual elements and study them, and establish the relations between them through carefully defined constructs using specific measures through well-designed statistical analysis. The emphasis is heavily based on reductionist and quantitative methods. Much of the early 'organizational effectiveness' literature of the 1960s and 1970s was very much of this nature. The constructionist tradition came into being as a reaction to the perceived problems of positivism in general and in relation to human systems in particular. Constructionists argue that humans are incapable of understanding 'reality'. Thus, constructionists believe that there is not a 'real' thing for public organizations. The performance which is studied belongs to different sets of actors, understandings, and constructions of performance. The emphasis in study is therefore qualitative and interpretive – the focus is on understanding what various groups and individuals think about performance and how it is socially constructed by themselves (ibid). According to Thietart et al. (2001) the positivist paradigm is dominant in organizational science. However, they maintain, "there has always been a conflict between positivism and interpretivism, which defends the particularity of human sciences in general and organizational science in particular. Constructivism, meanwhile, is becoming increasingly influential among researchers working in organizational science". (p. 14). Within this conceptual frame, this article utilizes the objectively gathered data using performance-based indicators realized in the real working situation. Objective process of knowledge generation, thus, fashions post-positivist paradigm which encompasses realist ontology (nature of knowledge being real as it exists in the real world of institutional setting), logical objectivist epistemology (indicator-based objective information collected and verified by supporting documents and observation) and naturalistic methodology (information collected from the real context).

Anthropological perspectives view that public values are essential to understand good or bad performance of an organization. Such approaches and values are derived from various political, social, cultural and economic theories (Brown, 1991 as cited in Talbot, 2010). Economic theories have always been concerned with the public as well as private institutions, and they relate to knowledge of the public domain. Theory of political economy is also concerned with political power sharing structure and economic processes around organizational context. By power is usually meant for the ability to coerce or influence compliance from citizens of a specific community or to mobilize support for specific collective action decisions (Talbot, 2010). This leads to the discussion that there are some specific theories with direct relevance to organizational performance. For example, institutional theory argues that belief structure is an important element in explaining human behaviour; "structures persist over time, even when individuals change; structures cause greater regularity in human behaviour than would otherwise be the case; and positive feedback between individuals and the institutions they inhabit over time reinforce institutions and patterned behaviour" (Talbot, 2010. p. 70).

Likewise, resource dependent and resource-based theories focused on the idea that organizations are successful to the extent that they can gather resources from their external

environment. Resource dependence, or exchange, has special relevance to the public sector for several reasons. Public agencies always exist within an external performance regime-authorizing environment (Moore 1995 as cited in Talbot, 2010), are subject to externally imposed mandates (Bryson 1995 as cited in Talbot, 2010), and most usually rely on their key resource – in terms of funding – on external providers, particularly governments. Evolutionary and ecological perspective in organization maintains that organizational performance and the factors that influence it are based on a methodology, which goes something like: (a) identify a group of successful companies/organizations; (b) see what they have in common in terms of internal factors like strategy, leadership, human resource policies, culture etc. (or even what is different between them and a control group of less successful companies); (c) conclude that the commonalities are what bring about enhanced performance (ibid.). Similarly, resource-based theories concentrate much more on the internal capabilities and resources of the organization (Barney and Clark, 2007 as cited in Talbot, 2010). Both "resource-dependence and resource-based theories have the obvious advantage that they focus on real assets that exist 'out there' and are not merely social constructs. Both approaches are useful in thinking about public and on organizational performance" (ibid., p. 72).

Performance status of the educational institutions of Nepal can be interpreted in a varied way using such theories and models. From political economy perspective, an organization is viewed from three lenses – educational, economic and political (Hirosato & Kitamura, 2009). The educational lens focuses on improved learning achievement; economic lens is related to the efficient use of resources; and political lens lies in participation of various stakeholders in educational activities, with an interest on the issue of decentralization, which is closely related to the issues of power and authority (Hirosato & Kitamura, 2009; Wayland, 2006). In similar vein, Acemoglu & Robinson (2013) argue that "there is a game of politics in the society and who wins depends on the distribution of political power in society" (p.79). This shows that decisions about resource allocation and usage at school are influenced by three interacting dimensions that give way to "informal goverance and political partronage network " (UK Department for International Development [DFID], 2009, p. 5), particulaly in current transitional political arrangements of the country thereby contributing to influence the orgizational performance and efficiency. This is how the institutional performance could be understood from various theoretical perspectives.

International Practices on Performance Audit

Performance audit has developed over several decades in many parts of the world with the aim of assessing how government organizations have performed and have used the resources provided to them (Lonsdale, Wilkins, & Ling, 2011). It grew initially in Europe, Australia and North America during the 1970s and 1980s. An overview of the performance audit mandates and provisions in some of the countries are summarized in the following table.

Table 1. Performance audit mandates in selected countries

Year	*Country*	*Legislation*
1921	United States of America	General Accounting Office established with the broad mandate to investigate 'all matters relating to the receipt, disbursement, and application of public funds' and 'to make recommendations looking to greater economy and efficiency in public expenditures'. Subsequent acts have clarified and expanded the mandate.
1948	Austria	Federal Law enables to examine the economy, efficiency and effectiveness of the operations of corporate public bodies, local authorities and provincial governments, and the economy and efficiency of state economic enterprises.
1967	France	Legislation provided for the SAI, the Courdes Comptes, to examine aspects of the economy, efficiency and effectiveness of public money.
1976	Netherlands	Government Accounts Act broadened out remit to performance audits to determine the performance of government, organization and management services. It was extended in 1992 to allow for examination of policy. The Algemene Rekenkamer investigates whether the central government revenue and expenditure are received and spent correctly, and whether central government policy is implemented as intended.
1977	Canada	The Auditor General Act 1977 provides the original legal basis for the Auditor General to carry out performance audits. It was amended in 1995 to include responsibilities related to environmental matters.
1983	United Kingdom	National Audit Act formalized the NAO's ability to examine the economy, efficiency and effectiveness of government spending.
1993	Ireland	The Controller and Auditor General (Amendment) Act allows the C&AG to carry out examinations of the extent to which acquisition, use and disposal of resources have been carried out economically and efficiently, but does not directly look at effectiveness.
1997	Australia	Auditor General Act authorizes the Auditor General to conduct a performance audit of an entity, a Commonwealth authority or company, other than a Government Business Enterprise performance audits focused on themes of governance and project management; border security and national security; community support and well- being; environment; industry, science and education; and transformation of entities.
1998	Belgium	Court of Audit undertakes audits at the federal level as well as at the regional and provincial level.Provides for the audit of the sound use of public funds and to examine economy, efficiency and effectiveness. Performance audits covers topics such as museums of fine arts and history, the implementation of the Kyoto protocol, the use of scientific knowledge in healthcare policy, public–private partnerships, support for people with difficulties integrating into the labor market, staff planning in government, rational use of energy in buildings, the functioning of the office for employment and professional training, and educational and administrative support given to elementary and secondary schools.

Source: Lonsdale, Wilkins, & Ling (2011, p. 5).

Examples of international practice demonstrate the focus of performance audit on economy, efficiency and effectiveness. Additionally, environment, national security, industry, education, public fund and rational use of resources are also the key concerns. Performance audits have been evolved taking on new forms and purpose in different parts of the world. For example, in Western Australia and Tasmania, performance auditing encompasses 'the range of audit and review activities from annual attest work on financial statements and performance indicators through to the preparation of direct reports on performance examinations' (Nichol, 2007 a cited in Lonsdale, Wilkins and Ling, 2011). In Canada, the Office of the Auditor General states that its performance audits 'examine the government's management practices, controls and reporting systems with a focus on results. In the United Kingdom, performance audit is designed to gather evidence to conclude on whether 'value for money' has been achieved, a term it defines as 'the optimal use of resources to achieve the intended outcomes.' (p. 7). In Sweden, performance audit includes reports on: controls on cross – compliance in the Government's shares; cultural grants – effective control and good conditions for innovation; program to provide immigrants with basic language skills; higher education quality; and the quality of private care for the elderly (Riksrevisionen, 2009 as cited in Lonsdale, Wilkins, & Ling, 2011). The National Audit Office in the United Kingdom produces around 60 major reports a year, mostly 'value for money' reports. It has statutory powers to examine the economy, efficiency and effectiveness with which central government and a range of other bodies – including universities, further education colleges, hospitals and private contractors working for government – use their funds.

Nepal's Situation of Institutional Performance Audit

In Nepal, Office of the Auditor General (OAG) with due consideration to regularity, economy, efficiency and effectiveness as deemed necessary in the acquisition and utilization of public resources, has been carrying out performance audit of selected public institutions every year since it gained the legal mandate of conducting performance audit in 2037 BS (OAG, 2015). In view of the mandate, OAG has included Information and Technology (IT) audit and Environment audit within the scope of performance audit since 2067 and 2068 BS respectively with key objectives of (i) disseminating information on management of public resources; (ii) improving the performance status of public institutions and; (iii) promoting public accountability of the institutions (ibid.). In similar vein, performance audit of educational institutions and agencies under the Ministry of Education has been initiated as provisioned in the School Sector Reform Plan (SSRP 2009-15). The SSRP envisioned an autonomous and independent auditing agency contributing to quality education (MOE, 2009). With this policy backup, Education Review Office (ERO) has been instituted in 2010 with the mandate of carrying out independent assessment in education. It has two-fold mandates of conducting student assessment to provide systemic feedback to the system for improving quality and equity in education, and promoting accountability of functionaries of MOE in effective service delivery through performance audit of schools and other agencies of the ministry of education. MOE through Education Policy Committee (EPC) is supposed to take necessary steps for improvement based on the assessment and audit reports submitted by ERO. To fulfill this mandate, ERO's activities have been guided by the Education Review Office

Guideline 2010 and Education Review Office Operation Guideline 2012 (MOE, 2069BS ; MOE, 2067BS). In this reference, the following table shows the total number of institutions covered in performance audit so far.

Table 2. Number of institutions covered in performance audit

Year	Schools	Resource Centres	District Education Offices	Regional Education Directorates	Central Level Agencies
2012	30	-	5	-	2
2013	48	-	5	-	-
2014	42	8	8	-	-
2015	45	15	15	1	-

As the table portrays, performance audit was delimited in selected schools and agencies based on the resource available. Audit is done using data collection tools developed by incorporating the indicators of performance based on specified job description of the respective schools and agencies under the Ministry of Education. Scope of audit thus includes both performance and compliance aspects of auditing. Auditors are externally nominated based on open selection criteria (MOE, 2067 BS). With a view to establishing auditing process as a regular function of the system, agencies are supposed to self-evaluate their performance using the same tools prepared for audit purpose. This self-evaluation further verified by the respective supervisors becomes a basis for external audit process. Ultimately, ERO prepares annual report of audit that shows performance status of institutions in terms of performance indicators put categorically as fully implemented, partially implemented and unimplemented indicators. Report thus produced is forwarded to the Education Policy Committee (EPC) in Ministry of Education for policy initiative to be taken and also sent back to respective agencies/schools as performance feedback aiming to develop and execute performance improvement action plan. Following this, School Supervisors, Resource Persons and personnel from Regional Education Directorate (RED) are supposed to be responsible for providing the monitoring/supervisory and follow up support to the respective institutions so as to ensure the smooth implementation of the action plan.

Institutional Performance Audit and Student Learning Achievement

Improved student learning achievement in schools reflects the performance effectiveness of the educational institutions. ERO's attempt through its performance audit and student assessment reports has been geared to this end. Selected DEOs' performance was assessed creating indicator based thematic categories – namely, policy compliance, school supervision and monitoring, activities related to district education committee, internal management, educational management, learning achievement indicators, innovative works and client satisfaction survey. Mean score on performance audit has been calculated from the given total score on indicators under each theme. Assuming that students' performance is the key measure of DEO's performance, attempts have

been made to explore a linkage between two sets of results. To be specific, the following table demonstrates the mean score of NASA results in relation to the mean score of DEOs covered in performance audit.

Table 3. District-wise mean of NASA and performance audit

Districts involved in Performance Audit	NASA Mean Score											Audit Score
	Grade 3		Grade 5			Grade 8						
	2012		2012			2011			2013			2013
	Nepali	Maths	Nepali	Maths	Eng	Nepali	Maths	Social	Nepali	Maths	Sci	
Dhanusa									39	28	24	25
Dolakha	61	52	57	56	56							68
Lalitpur	78	72	73	60	77	64	52	51	65	54	51	73
Nuwakot									54	33	38	31
Chitwan	63	53	63	51	51							55
Nawalparasi						43	45	47				41
Manang	64	67	58	51	39							28
Surkhet									43	17	31	82

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

The results of DEO's performance audit and NASA presented together show the variation in terms of subjects, districts, grades and years. The following figure shows this linkage separately in terms of average mean of the subjects in different years and districts.

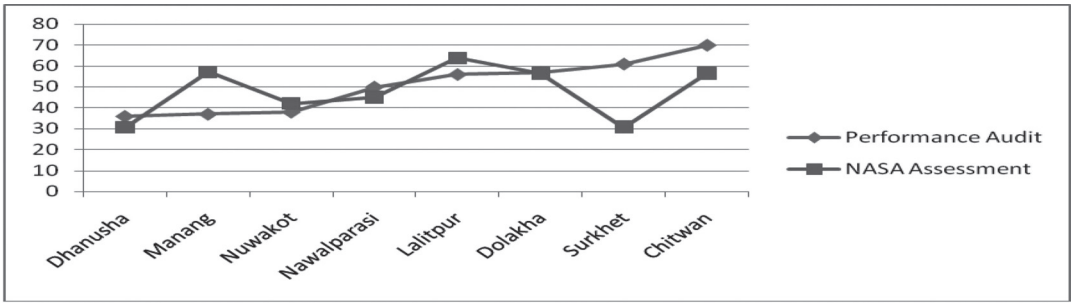

Figure 1. Comparison of NASA assessment results and DEO's performance audit

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

The figure demonstrates individual district's status on performance audit and learning achievement. Some districts like Surkhet, Lalitpur and Dolakha have a higher profile in performance audit while Manang, Nawalparasi, Nuwakot and Dhanusa are high in learning achievement. Surprisingly, high performing district Surkhet demonstrates low learning achievement.

Conversely, it looks quite opposite in case of Manang. With reference to performance audit status, NASA assessment mean score of the district keeps closer to each other except for Manang and Surkhet. This shows some kind of irregular pattern in two types of means. Correlation coefficient of 0.29 between DEO's performance and NASA assessment shows the weak relation between two variables; and effect size of 0.09 also indicates that DEO's performance has no significant effect on the learning achievement in the school. Likewise, subject-wise comparison below gives similar picture of the situation.

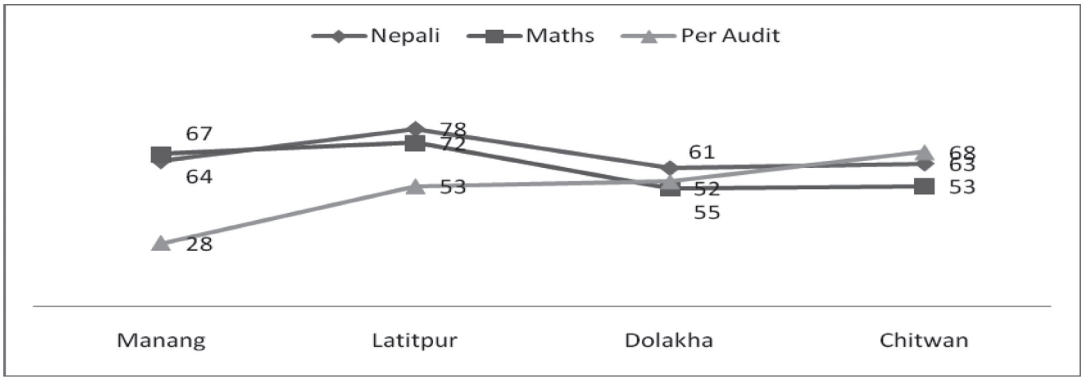

Figure 2. Status of learning achievement and performance audit in grade 3 (2012)

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

Comparing in terms of three subjects in grade three, we find DEO Manang and Lalitpur weaker in performance audit, whereas Dolakha to some extent keeps balance. Chitwan's position holds that there seems to be no significant relation between performance and learning achievement score. A negative correlation between performance audit and Nepali mean and between performance audit and Maths with zero effect size indicates that DEOs' performance has no effect on learning achievement. In the same fashion, the following figure presents similar situation about grade 5.

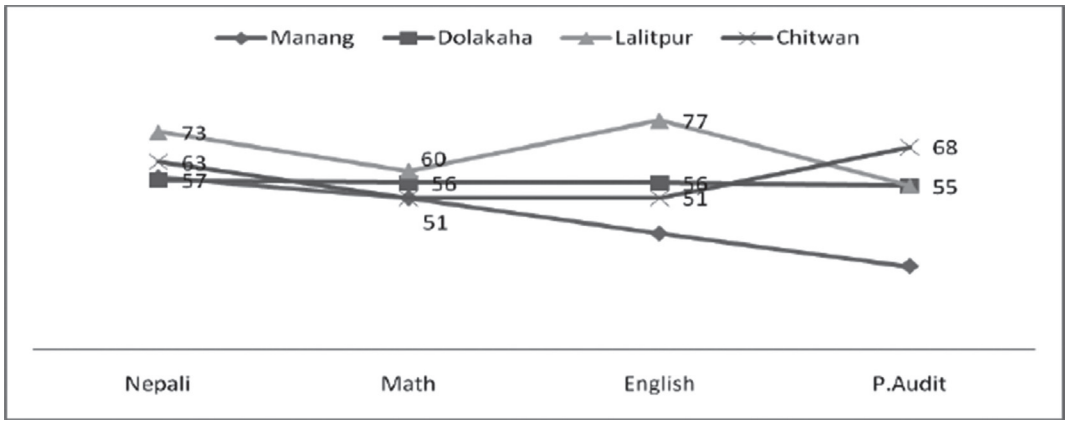

Figure 3. Status of learning achievement in Maths and performance audit in grade 5 (2012)

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

Dolakha shows comparatively a balanced status in performance audit and learning achievement, whereas Latitpur demonstrates high performance but low learning achievement, Chitwan and Manang are weak in performance and comparatively stronger in learning achievement. The correlation of performance audit with Nepali, Maths and English is 0.36, 0.22 and 0.49 with effect size of 0.13, 0.5 and 0.24 respectively. It means performance of DEO has moderate positive effects on student learning achievement in schools. At the same time, the following chart depicts the status in terms of grade 8.

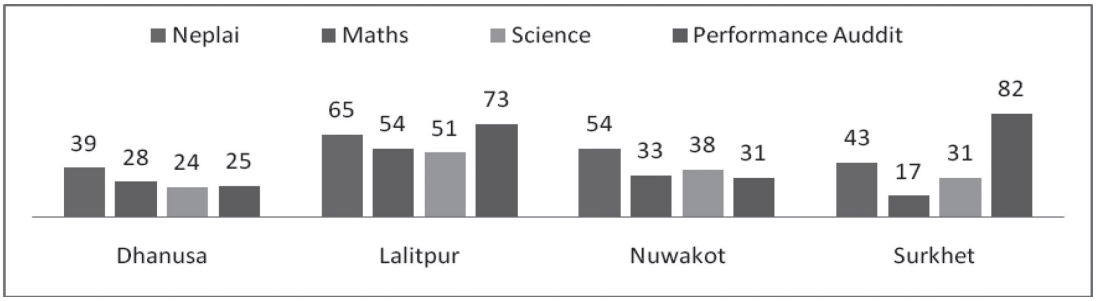

Figure 4. Status of learning achievement and performance audit in grade 8 (2013)

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

Comparing the performance audit status of DEOs with learning achievement in grade 8, we can see some kind of relation between two variables in case of Dhanusa and Nuwakot; whereas in Lalitpur and Surkhet performance audit result has not been found supportive to the learning achievement of students. The moderate and even weak correlation coefficient of 0.31,

0.07, 0.45 and effect size of 0.10, 0.01, 0.20 between performance audit and Nepali, Mathematics and Science respectively provides evidence of it. In the same manner, the following figure shows that performance status of Resource Centre also follows the similar pattern.

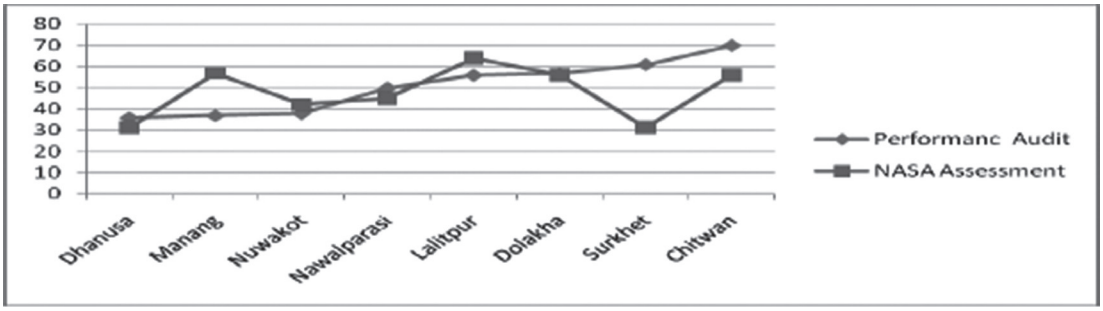

Figure 5. Status of learning achievement and performance audit of resource centre

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

As the figure and correlation coefficient of 0.29 shows, except for Manang and Surkhet, performance of DEOs was seen to some extent supportive to learning achievement. Moreover, the effect size of correlation 0.08 means that 8% of the variation in mean achievement scores among the different districts can be predicted from the relationship between mean performance audit and NASA achievement score. (Conversely, 92% variation in mean NASA scores cannot be explained.)

Likewise, the performance of Resource Center (RC) and District Education Office is also seen as follows.

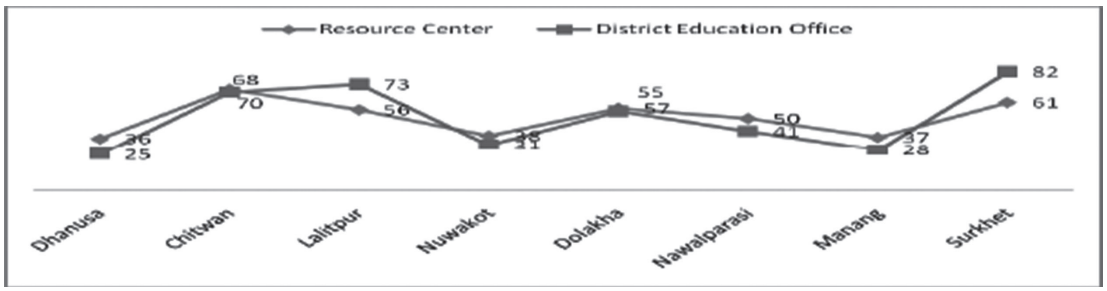

Figure 6. Performance audit status of Resource Center and District Education Office

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

The above table reports that in case of Resource Center and District Education Office, reasonable gap exists in the performance status, where Lalitpur and Surkhet demonstrate

comparatively wider gaps than other districts. Nevertheless, it shows some kind of uniform trend of relationship in the performance status of the two institutions. This implies that, to some extent, RCs' and DEOs' performance has corresponding influence on each other.

The following figure shows the relationship of schools' performance audit status with the national mean of learning achievement in different subjects. To show the linkage between two variables, the subject-wise mean obtained in different grades and years are presented as given in the parenthesis.

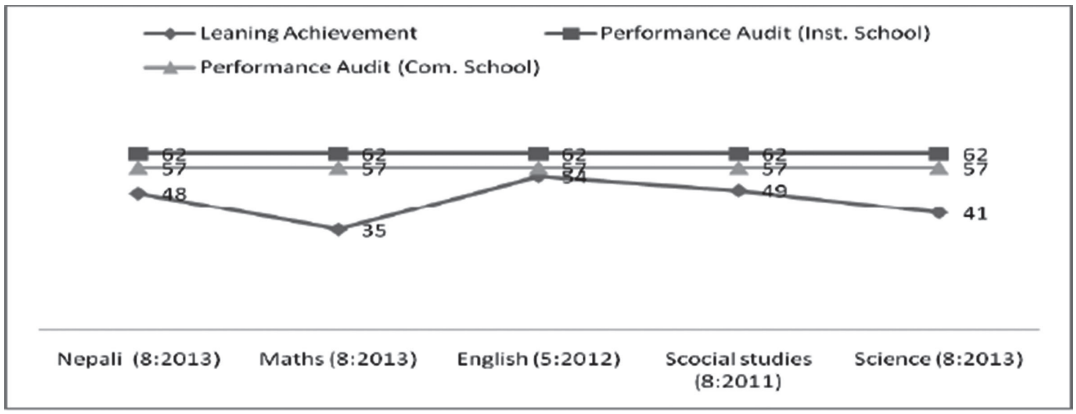

Figure 7. Status of schools' performance and national leaning achievement

Source: ERO (2071 BS); ERO (2070 BS); ERO (2013); ERO (2015).

Figure 7 demonstrates that institutional schools are better than the community schools in overall performance. Viewing from the vantage point of national mean of learning achievement in different subjects, English was seen favored more by the school management than other subjects. On the other hand, the wide gap seen between performance mean and mathematic mean indicates that schools are less supportive to mathematics than other subjects. Similarly, the performance status of schools in terms of various thematic indicators is depicted below.

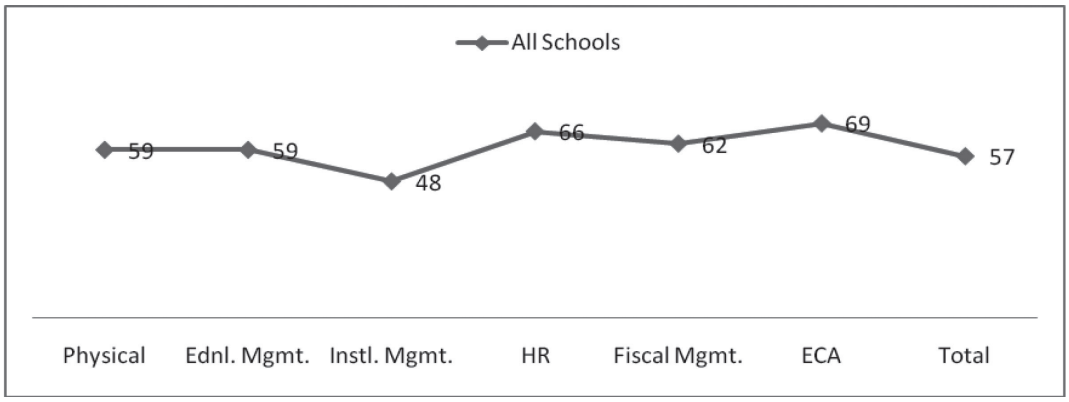

Figure 8. Performance status of schools in terms of thematic indicators

Source: ERO (2071, BS).

All schools' mean score on performance audit demonstrates that level of performance is moderate. Schools were found giving much importance to extracurricular activities and less priority was given to activities regarding instructional management. This result coincides with previously mentioned low achievement of students in different subjects of NASA as well.

So far, preceding paragraphs have explored the extent of linkage between two variables, viz. performance audit and student achievement. Status thus explored through national and international practices gives way to drawing important implications and learning as well as areas of focus of further study and concern for improvements. Paragraphs that follow provide further explanation on this matter.

Implication, Discussion and Learning

Performance audit provides evidence-based analysis of the tasks to be accomplished by schools and other agencies. It assists the management and those charged with governance in using the information to improve program implementation, reducing costs, facilitate decision making, initiate corrective action, and ultimately contribute to public accountability (www.projectauditors.com). The structure of Ministry of Education is vertically aligned from top to bottom. This structure needs to develop vision, mission, goal, strategies, objectives, activities and values aligned with the macro structure. Performance audit should be instrumental to develop this super structure as an integrated whole. Moreover, critical to the success of the organization is the work of managing and motivating the work force that is charged with reaching the goals. Performance audit supports the strategic plan, which in turn supports the organizational vision, mission and values (www.organizationaleffectivenessgroup.com).

Theories of management relate to the issue of performance in public agencies in several ways. These theories can be important as a perspective in understanding how performance policies of government and others might seek to change the rules of game. Attempts made through performance audit, which aim at improving performance leading to the rule of the game to change itself. Mintzberg et al. (1998 as cited in Talbot, 2010) opine that evolutionary and ecological theory emphasizes on environmental pressures, ecological niches, and organizational life cycles and extinctions. This gives the idea of integrating theoretical base to developing and designing the audit program. Complex systems are not readily predictable, because even minute variations in starting conditions can cause large, non-linear variations in outcomes (p. 73).

Discussing on various theoretical models of educational management, Bush (1995) maintains that theories are useful for influencing performance practice when they suggest new ways in which events and situations can be perceived for organizational transformation. According to Bush (1995), "political models assume that leaders are active participants in the process of bargaining and negotiation which characterizes decision making in organizations and thus mediate between groups in order to develop acceptable policy outcomes." (p. 146). On the other hand, in ambiguity models the leader "sets the framework for decision making but avoids direct involvement in the policy making process" (ibid.). Similarly, according to cultural model, "leaders are expected to communicate the organization's core values and beliefs." (ibid.) In current changed socio-political context of the country, political, ambiguity and cultural theoretical models will have a great implication for institutional transformation through improved performance and accountability.

Discrepancies between theory and practice can cause dilemma in achieving goals; it is seen from the Eastern classic of Bhagawat-Gita where Arjuna, the actor of the truth, diverts from his duty of war forgetting his previous promises made in favour of the truth and justice (Das, 2012). In this way, theoretical framework provides the leader or manager with room for reflecting his/her performance behaviour. Similar to this concept, Paulo Freire (Freire, 1993) with reference to praxis thesis states two dimensions of the word (reflection and action) focusing on "sacrifice of action is verbalism and sacrifice of reflection is activism" (p. 68). This implies that performance audit plan and interventions without theoretical orientation and transforming actions are likely to lead to anomalies and chaos.

Performance audit should 'concentrate on the issues that are important to society and in which there are clear risks of shortcomings in efficiency' (SNAO, 2008 as cited in Lonsdale, Wilkins, & Ling, 2011, p. 8). This implies that the focus of audit should be on a presumed problem and concern of actual government activities that are either being implemented or not implemented. Improvement policy and operation plan need to be based on such results.

International practices emphasize audit work from external bodies with wider coverage of public responsibilities of the organization with focus on the aspects of economy, effectiveness and efficiency. Rigorous selection of external auditors based on ethical consideration with localized

knowledge is really a challenging job. Nevertheless, specific operation guideline and committed facilitation from the trained personnel can ensure right selection of the auditors. In keeping with this, audit tools have to cover almost all areas of the task specified for an organization, particularly giving emphasis on the major areas.

In comparison, narrow gap between English mean and schools' performance coincides with parents and schools' inclination to English medium instruction these days. However, wide gap between Mathematics and schools' performance indicates the danger of over emphasizing English at the expense of other subjects. Therefore, proper balance of time, resource and effort has to be maintained in instructional planning and management in schools.

As demonstrated by the linkage above, some districts are extremely high in performance audit but low in learning achievement. It might be because of error related to sampling, instrumentation, administration, analysis and reporting process. This demands detailed review of the procedures and further exploration of the issues and causes behind this. Extremely high and extremely low performance backed up with inconsistent linkages of means between performance audit and NASA achievement across subjects and grades indicates areas of attention for improvements. The districts with high performing status and remarkably low student achievement call for strategies to minimize the gap and reduce discrepancies seen. Strategies to initiate may be concentrated on capacity building or intensive orientation of local actors such as auditors, focal persons, head teachers, school management committee members, resource persons, head teachers and teachers. Correspondingly, development and distribution of printed or web-based program manual and resource materials, online services, coverage on regular radio and TV program can support the dissemination and sensitization of performance audit and assessment program. Equally, local stakeholders like journalists, political workers, community members and parents also need to be aware of the intent and procedure of the auditing and assessment system so that collective efforts to ensure accountability in school could be strengthened. Identically, it is desirable to carry out the case studies of extremely low/high and moderately performing districts and to draw implications for improvements.

Weak correlation and effect size seen between RCs' and DEOs' performance and students' achievement is evident that RCs'/DEOs' performance management is just a little predictor of students' learning achievement in schools. This is a matter of serious concern, as these institutions invariably have to contribute to schools' performance reflected in the improved learning achievement. Reports published by Department of Education (DOE, 2011) supports the findings that "resource centre is found detached from the community and centers are not functioning, as Resource Centres are working only as data collection centres" (p. 68). The report recommends to specifying the roles, responsibilities and duties of the resource person along with a separate cell in DOE to look after all matters of RCs (Ibid.). This fact is further confirmed by the findings of GFA Consulting Group (2015) that reports: "Supervision is one of the weakest links in the school management system in Nepal. There is negligible supervision and monitoring visits by the RP and school supervisors"(p. xi). This obviously implies that either the roles of RC and DEO have to

be revisited and capacitated in the light of intended school performance or more resources with continued follow up support need to be channelled to the schools. At this point, this gives way to further areas of the study as well.

Conclusion

Compliance to existing rules and regulations aligned with the specified job descriptions, program and budgeting, scope of service delivery and assigned responsibilities to individual authority (ultimately leading to improved learning achievement) constitutes the foundation of performance audit of educational organization. Moreover, performance audit data are the reliable measures of institutional accountability as well. Performance audit carried out internationally in a varied way find external auditors and the auditing reports as the essence of the auditing task. Relevant normative and complexity theories of management are seen to provide the conceptual and analytical back up for understanding the critical areas of institutions during and after the performance audit. Presumably, DEO's performance and student learning achievement have to go together. However, mismatched linkages indicate the need for improvement in areas of institutional management and capacity building of human resources involved, together with re-allocation of resources backed up with continuous supervisory support to the schools and revisiting of the roles and responsibilities of RCs/DEOs. Identically, heterogeneous status of performance audit and learning achievement demonstrated by selective examples demonstrate persisting inconsistencies in educational institutions leading to non-compliance and under-performance. Self-assessment of the institutions followed by verification by respective supervisory agency, auditing by external auditors, dissemination of reports locally, and integrated longitudinal study initiated by ERO are expected to help ensure the accountability of educational institutions. Findings drawn out of selected data utilized in this article provide one level of understanding of the situation. However, further analysis of comprehensive data related to remaining schools and agencies covered in performance audit in different times may yield different results.

References

Acemoglu, D. & Robinson, J. A. (2013). *Why nations fail: The origins of power, prosperity and poverty.* London : Profile Books.

Arter, D. R. (2003). *Quality audits for improved performance*. Milwaukee: American Society for Quality.

Brynes, W. J. (n.d.). *Board audit*. Retrieved from http://www.managementandthearts.com/pdf/OrgAuditandBoardSurvey.pdf.

Bush, T. (1995). *Theories of educational management*. London: Paul Chapman Pub. Ltd.

Das, G. (2012). *The difficulty of being good*. New Delhi: Penguin Books.

Department of Education. (2011). *Role of resource centre for improving quality education in schools*.Sanothimi: Author.

Education Review Office. (2013). *Where are we now? Student achievement in Mathematics, Nepali and Social Studies in 2011*. Sanothimi: Author.

---------- (2015). *Education review office: An introduction*. Sanothimi: Education Review Office.

--------- (2070 BS). *National assessment of student achievement 2070 (grade 8): A brief report*. Sanothimi: Author.

--------- (2071 BS). *Performance audit of schools and institutions: An integrated report 2070/71*. Sanothimi: Author.

---------(2015). *National Assessment of student assessment: A glimpse of results*. Sanothimi: Author.

Freire, P. (1993). *Pedagogy of the oppressed.* London: Penguin Books.

GFA Consulting Group. (2015). *Joint Evaluation of Nepal's School Sector Reform Plan 2009-16.* Kathmandu: Author.

Hirosato, Y. & Kitamura, Y. (Eds.). (2009). *The political economy of educational reforms and capacity development in Southeast Asia.* Nagoya: Springer.

Lonsdale, J., Wilkins, P., & Ling, T. (Eds.). (2011). *Performance auditing: Contributing to accountability in democratic government*. Massachusetts: Edward Elgar Pub., Inc.

Ministry of Education. (2069 BS). *Education Review Office Operation Guideline*. Kathmandu: Ministry of Education.

---------- (2009). *School sector reform plan 2009-15*. Kathmandu: Author.

---------- (2067 BS). *Education Review Office Guideline, 2067*. Kathmandu: Author.

Ng. A. Y. (2002). Public accountability and performance auditing in government. *International Journal of Auditing, 6,* 109-118.

Office of the Auditor General. (2071). *A compilation of performance and special audit reports 2071*. Kathmandu: OAG.

Talbot, C. (2010). *Theories of performance: Organizational and service improvement in the public domain*. New York: Oxford University Press.

Thietart, R. A. et al. (2001). *Doing management research.* London: Sage Publication Ltd.

UK Department for International Development. (2002). *Condducting conflict assessment: Guiding notes.* London: UK Department for Int Development.

Wayland, F. (2006). *Elements of political economy*. Pensacola: Vance Publications.

Request to the Writers for Academic Articles

The Education Review Office (ERO) invites academic articles from writers for the forthcoming issue of Nepalese Journal of Educational Assessment (NEJEA) on the area of educational assessment, which includes themes, principles, theories, practices, issues and innovations on student assessment and performance audit of educational institutions. Along with electronic copy of article writer (s) should provide with a brief CV.

Articles should:

- **be written in English language.**
- **consist minimum of 4000 to maximum of 6000 words typed using Times New Roman Font.**
- **be well organized in standard form along with proper citation and referencing in established format based on APA referencing guidelines.**
- **have a short and precise abstract of 100 to 150 words.**
- **be an original not published elsewhere earlier.**
- **be either conceptual or empirical, however, we encourage for the articles having a strong linkage between theory and/or empirical research with the implications for policy and/or practice on educational assessment.**

Editorial Board reserves the right to accept, reject or publish with editing the articles received.

Government of Nepal
Ministry of Education
Education Review Office
Sanothimi, Bhaktapur
Nepal
Phone: +977-1-6633556, +977-1-6633557
Fax: +977-1-6633556
Email: nejea.ero@gmail.com
Website: www.ero.gov.np

NEJEA
ERO