Education Review Journal

A peer- reviewed open -access journal indexed in NepJol
ISSN 2976-1182(Print)
Published by Education Review Office, Nepal
Article History: Received on 15 April 2025; Accepted on 6 July 2025
DOI: https://doi.org/10.3126/erj.v2i01.86472

Large-Scale Student Assessment in Nepal: Current Practices and Future Needs

Lekh N. Paudel
Former Joint Secretary, Government of Nepal

Dr. Paudel (https://orcid.org/0009-0001-4021-253X) is a retired Joint Secretary of the Nepal Government. Correspondence concerning this article can be addressed to him at his email address: inp_001@hotmail.com.

Abstract

This article provides an overview of the concepts and practices of large-scale assessment in Nepal, reviewing relevant literature. It highlights the necessity of employing large-scale assessments to identify the performance of the education system through student achievement tests. The findings of this review indicate that while awareness towards the importance of large-scale student assessment is increasing among various stakeholders in Nepal, there remains significant potential for improvement in assessment practices. This includes ensuring conducive conditions, aligning assessments with the educational system, and enhancing overall assessment quality. Additionally, the article emphasizes the importance of being aware of the potential negative consequences of large-scale assessments, urging efforts to mitigate risks while maximizing their benefits for assessing and improving the education system.

Keywords: large-scale assessment, system-level assessment, assessment for accountability, assessment framework, factors influencing learning, national assessment of student achievement, Item Response Theory.

Introduction

Student assessment is a crucial part of the education system, which helps identify what students have learned through teaching and learning activities. It involves a systematic gathering, analysis, and application of evidence regarding student learning to enhance programs, schools, and overall student outcomes (Jankowski & Baker, 2019; Palomba &Banta, 1999). According to Clarke and Luna-Bazaldua (2021), student assessment is defined as "the process of gathering and evaluating information on what students know, understand, and can do." There is a growing global trend in the use of three assessment types: classroom assessment, public examinations, and system assessments, each serving distinct purposes. Classroom assessment primarily aims to enhance student learning by offering formative feedback. At the same time, public examinations focus on making summative decisions, such as assigning grades and categorizing students based on their performance. System-level assessment, often referred to as large-scale assessment, is designed to evaluate the performance of the educational system and provide feedback for policy interventions. Large-scale assessments ensure accountability in student learning and deliver system-level insight into educational performance. A balanced approach to these three assessment forms is essential for developing an effective education system.

This article, based on a literature review, describes the conceptual and theoretical frameworks of large-scale assessments and reviews the reports on the practices of large-scale assessment. It elaborates on the concept of large-scale assessment and explores its necessity. The article discusses the applications of large-scale assessments and traces their evolution and practices in Nepal. Additionally, it identifies strategies for enhancing the current practices of the national large-scale assessment in Nepal.

Large-Scale Student Assessment: Concepts and Needs

System-level assessments, often referred to as large-scale assessments, "provide insights into overall performance levels and trends within the education system, serving as a support for policy decision-making" (Clarke & Luna-Bazaldua, 2021). A national large-scale assessment of student achievement "evaluates the overall health of the entire education system, offering policy feedback to the government, as well as to teachers, parents, and students" (Poudel, 2016, p. 4). He further explains that large-scale assessments "do not analyse individual students and teachers, but instead compare various subgroups based on their achievement

scores." Globally, there are practices of large-scale assessments at the national, sub-national, regional, and international levels.

In addition to prominent international assessments, such as the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS), there are various regional assessments (UNESCO, 2019). Many national or sub-national education systems periodically conduct their large-scale assessments, with most based on samples; however, a few systems employ a census approach for their national or sub-national assessments.

The primary aim of system-level large-scale assessments is to ensure accountability by measuring performance across the entire system. More specifically, these assessments attempt to answer one or more of the following questions:

a) How well are students learning in this education system? Are they meeting specific learning standards? Are there particular strengths and weaknesses in students' knowledge and skills?
b) Do particular subgroups perform worse than others? Are there disparities, for example, between the performance of boys and girls, students in urban and rural locations, or students from different language groups?
c) What factors are associated with students' achievement? To what extent does achievement vary with the characteristics of the learning environment (for example, school resources or teacher preparation) or with students' home circumstances?
d) Do students' achievements change over time? What factors are linked to changes in student achievement over time? (Kellaghan & Greaney, 2008; Clarke & Luna-Bazaldua, 2021

Some of the initial questions mentioned can be addressed through public examinations, while others, particularly the last three, are challenging to resolve using these examinations. Nevertheless, we can use the results from large-scale assessments to answer all the questions posed. This underlines the necessity for system-level large-scale assessment within the school education framework.

Key distinctions between public examinations and large-scale assessments (Kellaghan & Greaney, 2008; 2020) further highlight the need for large-scale assessment. Public examinations aim to certify and select individual students, carrying significant consequences for students, teachers, and schools. In contrast, large-scale assessments focus on providing feedback to policymakers with low-

stakes impact for individual students, teachers, and schools. While all students at a specific grade level typically take public examinations, large-scale assessments often involve a representative sample of students instead. Public examination results are provided to individual students, whereas large-scale assessments yield aggregate results without attributing them to specific individuals. Additionally, public examinations do not allow for year-on-year result comparisons, while appropriately designed large-scale assessments facilitate such comparisons. Furthermore, public examinations lack questionnaires that assess factors affecting student learning information vital for policy decisions while large-scale assessments do include such tools.

These differences illustrate the distinct objectives and functions of system-level large-scale assessments compared to high-stakes public examinations, indicating the necessity for separate assessments to provide policymakers with feedback for enhancing the education system, as well as to certify and select students for future education, training, and employment. Researchers indicate that employing a single assessment for multiple purposes often leads to unsatisfactory outcomes, as it may overly prioritize one specific goal (Archer, 2017). The focus of public examination results on high-stakes decisions may indirectly support improvements in the education system; however, this is insufficient for effective policy feedback. Archer (2017) compared three assessment purposes supporting learning, accountability, and certification, progress, and transfer to the three sides of a triangle, emphasizing the importance of conducting and balancing assessments aimed at these varied objectives to ensure quality education. She argued that "an overemphasis on any one of the purposes of assessment will affect the other sides by diverting resources from one or both of the other essential assessment functions, thereby adversely influencing the quality of education" (Archer, 2017, p. 3). Achieving this balance necessitates three types of assessments: classroom assessments for learning support, public examinations for certification and classification, and large-scale assessments for accountability and system improvement.

Since the mid-1990s, the OECD and IEA have been conducting large-scale international assessments in various countries or educational systems. The OECD administers PISA, while the IEA oversees TIMSS and PIRLS. Additionally, there are several regional and national large-scale assessments. Nepal has been conducting national assessments (NASAs) but has not engaged in any international or regional assessments. A feasibility study of Nepal's involvement in international assessments, commissioned by the ERO in 2016, recommended focusing on

enhancing national large-scale assessments instead of pursuing participation in international assessments (CERSOD, 2016).

Use of Large-Scale Assessments

Large-scale assessment results can play a crucial role in enhancing the education system. For instance, insights gained from these assessments can shape education policies. They support establishing educational standards, prompting curricular reforms, directing resource allocation, setting and evaluating learning objectives, adjusting classroom practices and teacher training, and facilitating stronger connections between home and school to foster student learning (Clarke & Luna-Bazaldua, 2021). According to UNESCO (2018), we can use data from large-scale assessments for comparisons and benchmarking, refining the educational system through informed policy decisions, increasing access and equity, enhancing teaching and learning practices, implementing curriculum reforms, and employing strategies and indicators to monitor and assess educational processes.

The Education Agenda 2030, as part of the Sustainable Development Goals (SDGs), emphasizes the assessment of learning outcomes as a vital tool for ensuring quality education for all. It views large-scale assessments as vital for tracking learning and progress (UNESCO, 2019). According to UNESCO (2019), "LSLAs provide essential evidence to guide the setting of agendas, identify key issues, inform policy development, and support policy execution" (p. 19). Stakeholders at various levels can utilize the findings from large-scale assessments for a range of purposes. UNESCO (2019) provides examples of policy initiatives that these data can inform for different stakeholders.

Table 1

Examples of policy initiative that large-scale learning assessment data can inform, by level

System	Curriculum	School	Teacher	Home
• Allocating resources • Implementing programatic reforms • Outlining goals for curricular achievement	• Curriculum development • Curricular content and methods • Curriculum design	• Setting faculty priorities • Improving student support services • Enriching school environment	• Securing resources for professional development or improving pedagogical practices • Revising	• Informing programmes to encourage parental involvement

			courses and assignments	

Adapted from UNESCO, 2019.

The reviews referenced above highlight the significant role of large-scale assessment data and findings in shaping education policy, program design, implementation, and evaluation. The roles include informing policy development, recognizing priorities, enhancing curricula and teaching methods, supporting teacher development, improving education delivery, and promoting quality and equity in education. Additionally, it contributes to establishing benchmarks and monitoring educational progress.

Large-Scale National Assessment in Nepal

This section outlines the evolution of the National Assessment of Student Achievement (NASA) practices in Nepal. It highlights significant findings from NASA, illustrating how these insights have been utilized to guide policymaking.

Evolving the Practice of the National Assessment of Student Assessment

Recognizing the importance of three distinct functions of student assessment, Nepal launched large-scale system-level assessments in the mid-1990s, which became formalized with the establishment of the Education Review Office (ERO) in 2010. Since then, the ERO has been carrying out the National Assessment of Student Achievement (NASA) across various school grades by selecting a representative sample of schools and students. These periodic, large-scale assessments, with some variations in methods and processes, have been conducted for different grades. From 1997 to 2011, the Basic and Primary Education Project (BPEP) and the Department of Education (DOE) of the Government of Nepal hired consulting firms to conduct national assessments for grades 3, 5, and 8 (Poudel, 2017; Poudel & Bhattarai, 2018). Numerous assessments took place during this timeframe, including those by EDSC in 1997, 1999, 2001, 2003, 2008, and 2011; CERSOD in 2001. Additionally, BPEP performed a few assessments in 1995, 1997, and 1998 using relatively small sample sizes. However, the impact of these assessments was constrained by the lack of an institutional framework and a consistent system for national assessments, as they were primarily conducted by external agencies. Furthermore, comparability over the years was hindered due to the use of Classical Test Theory (CTT) with percentage scores (Poudel, 2016). Following the establishment of the ERO, from 2010 to 2015, it implemented two

rounds of national assessments for grades 3, 5, and 8 (ERO, 2013, 2015, 2015a, 2016a). Despite the application of standardized items, the selection of representative samples, and the inclusion of factors influencing student learning for the assessment, the dependence on classical test theory with mean percentage scores resulted in limited comparability over time. Additionally, there was an absence of a clearly defined assessment framework, standards, and benchmarks. To overcome limitations, ERO has enhanced its large-scale assessment practices since 2016 by establishing an assessment framework and implementing Item Response Theory (IRT) (ERO, 2018, 2019, 2021, 2022).

In 2014, Nepal carried out its first nationally representative Early Grade Reading Assessment (EGRA) (RTI, 2014). The objective of this study was to determine the reading proficiency of students in grades 2 and 3 nationwide, disaggregating the data by geographic location, socioeconomic status, language spoken at home, and various school-related factors that impact learning. The assessment presented only the percentage of correct answers in areas such as oral reading fluency (ORF), listening comprehension, letter sounds, and *matra*, without establishing benchmarks to interpret findings. To further evaluate early grade reading and mathematics, the Education Review Office (ERO) developed an assessment framework for grade 3 in 2018 that incorporated both reading and numeracy skills. Subsequently, the National Assessment of Reading and Numeracy (NARN) was conducted in 2020, utilizing representative samples from Nepal (ERO, 2020). This assessment also fulfills the reporting obligations for early-grade reading and numeracy as outlined in Sustainable Development Goal 4.1.

In Nepal, national assessments of student achievement have become institutionalized, resulting in significant improvements in educational practices. Nonetheless, there is a pressing need to broaden awareness about the roles, importance, and applications of sample-based large-scale assessments within the education system, as well as among key stakeholders. The influence of longstanding high-stakes examinations in Nepal remains deeply rooted, with some individuals still advocating for these public exams to take preference over large-scale assessments and classroom assessments. During discussions with various stakeholders, I noted that some people preferred utilizing public examination results for both the purposes of certification and feedback to the education system. Others support implementing both types of assessments based on specific objectives and needs. However, there is a growing consensus among education stakeholders including policymakers, teachers, and education managers on the necessity of

incorporating three types of assessments, large-scale assessments being one of them. Therefore, the pertinent question regarding large-scale assessments shifts from "why do we need them?" to "how can we enhance their effectiveness and utilize their results to improve the school education system in Nepal?

National Assessment of Student Assessment Findings and Their Uses

Poudel and Bhattarai (2018) summarized the objectives of the National Assessment of Student Achievement (NASA) conducted since 2011 into three main points: First, to assess students' performance levels in their respective grades and subjects; second, to examine variations in student achievement across different geographical regions, districts, development areas, language groups, and genders; and third, to identify the factors that impact students' learning. Since NASA's inception in 2011, the Educational Review Office (ERO) has established these three objectives for each assessment and has reported the results and findings corresponding to these goals. Each cycle of NASA has provided insights based on the results and has offered related recommendations. For example, I have included two summaries of NASA findings: the first summarizing the results from 2011, 2012, 2013, and 2015 assessments as outlined by Poudel (2017), and the second detailing the findings from the 2020 NASA for grade 8

Summary of NASA findings from NASAs 2011, 2012, 2013 and 2015

- Increasing learning achievement and reducing the persisting inequalities across districts, ecological zones, development regions, rural and urban locations.
- Promoting higher cognitive ability and improving reading ability.
- Balancing instructional activities to all content domains and areas.
- Managing various diversities including linguistic, cultural, socio-economic, gender through developing inclusive and child-friendly school and classroom environment.
- Reorganizing and revising teacher preparation and teacher development strategies,
- Making school accountable for student learning.

Summary of NASA finding of NASA 2020 for grade 8

- Learning from high performing schools and reducing the difference in achievement between public and private schools,
- Enrolling students of proper age and designing some alternative strategies for over aged students.
- Ensuring parents' and community participation in school, particularly in students' learning, and designing and carrying out parental education.
- Ensuring that children of school (Basic) age should not be engaged more than 2 hours a day in the household works.
- Eliminating child labour and ensuring that school (Basic) age children are not involved in any paid job.
- Designing some extra supports for those students whose parents and other family members are unable to support their study at home.
- Ensuring the access of a set of textbooks for each student on the first day of a new session and making sure that they have access to some additional reading materials.
- Eliminating the incidence of bullying at school.
- Improving the assessment system, focusing particularly on classroom-based assessment for formative purposes. (Adapted from Poudel, 2017, pp. 31-32)
- Lrge number of students are below grade level, and an alarming gap exists between the intended and achieved curriculum.
- Wide gaps in achievement between provinces.
- Huge disparity in achievement by type of school.
- Home language also brought a remarkable gap in achievement.
- There is a visible gap in the learning achievement between boys and girls.
- Students at the appropriate age performed better.
- The relationship between students' academic performance and socio-economic status is substantial, but its magnitude varies by subjects.
- The achievement of Janajati and Dalit children is lower than other ethnicities.
- Teacher regularity and availability of study resources have positive relationships with learning achievement.
- Decreasing patterns of achievement and consistency of NASA results.
- There is a positive relationship between students' academic performance and the use of their leisure time in school.

- Access to social media also has a positive effect on student achievement.
- Medium of instruction has impacted on achievement in community schools.
- Noticeable gaps in school governance. (Adapted from ERO, 2020.)

The two sets of summaries mentioned above indicate that the results and findings from NASA offer valuable insights for reforming policies and programs. In addition to presenting and detailing these findings, NASA has recommended various interventions aimed at enhancing student achievement. The effectiveness of NASA results in improving the educational system largely depends on the reform agenda adopted by both federal and local governments considering these findings. In Nepal, there are numerous instances where NASA findings have directly or indirectly influenced education policy, planning, and implementation. The following paragraphs provide examples of how NASA findings have been utilized in shaping education policy and practices in Nepal.

There is a growing awareness among schools, parents, and other educational stakeholders regarding the necessity of national assessments to evaluate educational outcomes. Relevant agencies and individuals view NASA results as "one of the valid and reliable sources of information for planning, programming, and reviewing educational programs and outcomes" (Poudel, 2017, p. 32). During the School Sector Development Plan period (SSDP, 2016-2021), the Ministry of Education formulated an action plan informed by NASA findings. For instance, 2013 action plan outlined six intervention strategies for school education in Nepal, based on NASA results. These strategies included enhancing the overall learning environment in schools, promoting activity-based learning and early grade reading, reforming the student assessment system, strengthening teacher preparation and professional development, improving professional support and supervision in classrooms, and institutionalizing the NASA system (Poudel & Bhattarai, 2018). Additionally, NASA findings played a role in initiating curriculum reforms in 2018 that focused on activity-based learning, integrated curricula for grades 1-3, the incorporation of soft skills across all grades, and placing significant emphasis on classroom-based assessments and formative feedback.

The current School Education Sector Plan (2022/23-2031/32) incorporates NASA findings when evaluating the status of education and setting future programs and strategies (MOEST, 2022). This plan includes a program of conducting NASA assessments periodically to identify student achievement levels and inform policy

and program development. The School Sector Development Plan (SSDP, 2016/17-2022/23) acknowledged the importance of large-scale assessments, stating that "NASA is a tool that decision-makers can use to target efforts and resources for maximizing gains in effectiveness and efficiency" (MOE, 2016, p. 24). This plan emphasized the need for utilizing NASA results during the planning phase to assess student learning outcomes and identify areas of strength and weakness, as well as recommending the use of NASA findings to inform teacher development and curricular reform. Moreover, education stakeholders, experts, and researchers have recognized assessment results as a crucial source of information for their studies.

Although there are notable influences of NASA findings on policy and program design in Nepal, many essential findings and recommendations from various NASAs regarding the improvement of educational quality and equity still require consideration during the development and implementation of education policies and programs.

Improving the Practice of Large-Scale Assessment in Nepal

One of the ways to assess the quality of a large-scale assessment system in an education system is to evaluate three quality drivers: enabling conditions, system alignment, and assessment quality (Clark, 2012). Using this framework, Poudel (2016) reviewed the quality of large-scale assessment practices in Nepal and suggested areas for further improvement. He assessed the enabling context for NASA by examining policies, leadership, public engagement, funding, institutional arrangements, and human resources; system alignment through learning/quality goals, curriculum, and pre-service and in-service teacher training; and assessment quality by evaluating the design, instruments, and processes of administration and analysis, as well as the effective use of results. He found that a lack of proper institutional arrangements and human resources contributed to poor enabling conditions. Similarly, he identified the need for further improvements in alignment with the education system and the quality of large-scale assessment in Nepal.

Since 2016, there has been an improvement in institutional arrangements, including the establishment of ERO as an agency for conducting large-scale national assessments. However, the issues of autonomy, organizational structure, and proper human resources have not been adequately addressed. Regarding system alignment, assessments are designed based on the curriculum; however, alignments with the pre-service and in-service teacher development processes are not visible. However, since 2016, ERO has begun developing an assessment framework, which

has contributed to improving alignment with the curriculum. Some improvements in assessment quality were made when ERO initiated the use of Item Response Theory (IRT) in 2017 and built the capacity of national human resources in assessment design, tool development, administration, and results analysis. However, the quality assurance aspect remains challenging due to the nature of ERO staff and the national capacity for assessment, particularly in designing, conducting, and analyzing assessment data using IRT. Similarly, as mentioned above, the effective use of assessment results to improve policies and education delivery has always been challenging.

As these three quality drivers are interrelated, the enabling conditions and quality of assessment influence their alignment and use. Simon et al. (2013) discussed four issues related to large-scale assessments that affect the quality of the assessment process, tools, and analysis. The effectiveness of assessment results largely depends on assessment design, development, and delivery; assessment of diverse populations; scoring, score reporting, and the use of scores; and psychometric modeling and statistical analysis.

This review of quality drivers related to large-scale assessments in Nepal indicates that the education system has recognized the need for and use of large-scale assessments. As a result, institutional arrangements and practices have been initiated since the establishment of ERO. Similarly, some noticeable improvements in alignment and quality have been made. However, there is a need to enhance several aspects of the large-scale assessment system to make it more effective in assessing the quality of the education system and informing policy and practice in education. The following are the key areas for improvement in developing an effective large-scale assessment system in Nepal:

Institutional Improvement. There is a need to provision adequate technical human resources and provide more autonomy to ERO. For this, current institutional mandates and human resource provisions need to be changed by offering more autonomy and employing expert human resources. In addition to regular human resources, there is a need to develop a pool of experts in assessment and psychometrics within the national system by collaborating with universities and research institutes. Similarly, we need to develop national and institutional capacity in test development, result analysis, and report preparation, particularly by strengthening ERO with technical capacity and technological infrastructure.

Ensure Technical Quality of Assessment. There is a need to improve assessment design by reviewing and enhancing the assessment framework,

including identifying content and cognitive domains, defining standards, and reviewing and redefining contextual variables that influence student learning. Similarly, we need to continuously improve the quality of instruments by ensuring an adequate level of reliability and validity, along with proper difficulty levels. Similarly, we need to review the sampling process, design of assessment booklets, contextual tools, administration of tests and contextual questionnaires, collect answer sheets, and improve the data entry process. Assessment results should be made comparable across years and groups by using appropriate items for the assessment, employing a practical approach to equating items, and refining the data analysis process. For this, ERO should strengthen the use of IRT in assessment design and result analysis.

Policy Linkage and Use of Results. To improve the use of assessment findings in policy and program design and implementation, ERO should disseminate assessment results adequately and collaborate with policymakers and implementers. The education system needs to establish an institutional arrangement for regular review and utilization of assessment findings to enhance policy and program design and implementation. Similarly, we need to decide appropriate intervals, grades, and subject areas for assessment, viewing its policy implications. The current three-year interval and four grades require revision. Based on our experiences and international practices, it would be more effective to conduct large-scale assessments every four years at only two specific grade levels in school education.

Conclusions

The Nepali education system has implemented three distinct types of assessment, each serving different purposes. Public examinations are designed to certify individual students and categorize them according to their achievement scores. Classroom assessments aim to enhance teaching practices, while large-scale assessments focus on improving the overall quality and equity of education by evaluating systemic indicators and factors that affect learning. This means that large-scale assessments generally do not provide insights into individual student performance, but rather reflect the overall achievement levels of the education system and specific groups within the population. This review highlights that the education system acknowledges the need for and application of large-scale assessments. The Educational Review Office (ERO) has been established as the specialised agency for large-scale assessments, equipped with specific institutional

frameworks. Current national large-scale assessment practices follow a standard procedure to ensure the quality standards of assessment are met.

However, it is necessary to bring it for improvement in institutional arrangements, autonomy, and human resource provisions for assessment. Furthermore, continuous evaluation and improvement of various aspects related to the quality and alignment of assessments are crucial. It is also essential to develop a system for effectively disseminating and utilizing assessment results and findings. Given international and national practices, resource requirements, and their applications, it would be justifiable to extend the interval for large-scale assessments from three years to four years and reduce the assessment points from four grades to two.

While well-implemented large-scale assessments can provide valuable insights for policymaking and program implementation, it is important to recognize the potential drawbacks associated with them. Four areas of concern regarding the negative impacts of large-scale assessments have been identified (UNESCO, 2019). The first issue is how these assessments can narrow the conceptualization of education and learning, confining achievement scores to specific content and cognitive domains. For instance, it is debatable whether critical social, civic, and personal learning can be accurately measured using traditional quantitative approaches (UNESCO, 2019). A second concern is the potential for a limited understanding of quality, as it may equate educational quality solely with test outcomes rather than recognizing learning as a multifaceted construct. The third limitation relates to a narrow focus on knowledge domains, disregarding the broader spectrum of skills, attitudes, and values. The fourth issue relates to a lack of emphasis on inclusivity and diversity, as assessments often cater to what are considered average students. If comparability is prioritized excessively, local validity may be overlooked, which is vital for understanding the circumstances of disadvantaged groups (UNESCO, 2019).

Large-scale assessment results are significant indicators of students' learning and the system's performance, providing valuable feedback on various aspects of the education system. When designed, implemented, and utilized effectively, large-scale learning assessments can offer numerous advantages to the education system. To maximize these benefits, it is essential to remain mindful of the associated risks and take steps to mitigate them through improved assessment design and processes.

References

Archer, E. (2017). *The assessment purpose triangle: Balancing the purposes of educational assessment.* Front. Educ. 2:41. https://doi: 10.3389/feduc. 2017.00041.

CERSOD (2016). *Feasibility study of Nepal's participation in international assessment.* Education Review Office (ERO).

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

Clarke, M., & Luna-Bazaldua, D. (2021). *Primer on large-scale assessments of educational achievement. National Assessments of Educational Achievement series.* Washington, DC: World Bank. https:// doi: 10.1596/978-1-4648-1659-8. License: Creative Commons Attribution CC BY 3.0 IGO.

Education Review Office. (2013). *Report of National Assessment of Student Achievement 2011, Grade* 8. ERO, GON.

Education Review Office. (2015). *Report of National Assessment of Student Achievement 2012, Grades 3 and 5.* ERO/GON.

Education Review Office. (2015a). *Report of National Assessment of Student Achievement 2013, Grade 8.* ERO, GON.

Education Review Office. (2016*). Assessment framework: National Assessment of Student Achievement for Grade 8.* ERO, GON.

Education Review Office. (2016a). *Report of National Assessment of Student Achievement 2016, Grade 8.* ERO, GON.

Education Review Office. (2018). *Report of National Assessment of Student Achievement 2017, Grade 8*. ERO, GON.

Education Review Office. (2019). *Report of National Assessment of Student Achievement 2018, Grade 5*.ERO/GON.

Education Review Office. (2020). *National Assessment for Reading and Numeracy,* 2020, *Grade 3*. ERO/GON.

Education Review Office. (2021). *Main Report of National Assessment of Student Achievement (NASA), 2019.* ERO/GON.

Education Review Office. (2022). *Report on National Assessment of Student Achievement 2020 in Mathematics, Science, Nepali and English for Grade 8*.

Jankowski, N. A., & Gianina R. B. (2019). Movement Afoot: Fostering Discourse on Assessment Scholarship. In *Trends in assessment: ideas, opportunities,*

and issues for higher education, edited by Stephen P. Hundley and Susan Kahn. Pp. 19-32. Stylus, STERLING.

Kellaghan, T., & Greaney, V. (2008). *Assessing National Achievement Levels in Education.* The World Bank.

Kellaghan, T., & Greaney, V. (2020). *Public Examinations Examined.* World Bank. https:// doi: 10.1596/978-1-4648-1418-1.

Ministry of Education. (2016). *School Sector Development Plan (SSDP, 2016/17-2022/23).* MOE.

Ministry of Education, Science, and Technology. (2022). *School Education Sector Plan (SESP, 2022/23- 2031/32).* MOEST.

Palomba, C. & Banta, T.W. (1999). Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. Jossey-Bass, Inc., San Francisco.

Poudel, L. N. (2016). Reviewing the Practice of National Assessment of Student Achievement in Nepal. *Nepalese Journal of Educational Assessment, 1(1), pp. 1–16.*

Poudel, L. N. (2017). A Review of the Results of National Assessments of Student Achievement in Nepal. *Nepalese Journal of Educational Assessment, 2(1), pp.19-36.*

Poudel, L. N., & Bhattarai, G. P. (2018). *Integrating the Findings from the National Assessment of Student Achievement into the Policy Process: An Experience from Nepal.* Using Assessment Data in Education Policy and Practice: Examples from the Asia Pacific. UNESCO, Bangkok/NEQMAP/ACER.

Research Triangle Institute (2014). Nepal-Early Grade Reading Assessment (EGRA). MOE, GON.

Simon, M., Ercikan, K. & Rousseau, M. (2013). *Improving large-scale assessment in education: theory, issues and practice.* Routledge.

UNESCO. (2018). *The Impact of Large-scale Learning Assessments.*UNESCO.

UNESCO. (2019). *The Promise of Large-scale Learning Assessments: Acknowledging Limits to Unlock Opportunities.* UNESCO.