unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

WG/GAML/11/2.5.d

DRAFT REPORT ON THE UIS BENCHMARKING STUDY IN NEPAL

Global Alliance to Monitor Learning

Draft Report on the UIS Benchmarking Study in Nepal

Dr. Jeff Davis (International Consultant)

Dr. Lekha Nath Poudel (National Consultant)

13 February 2025

Acronyms

ACER	Australian Council for Educational Research
BPEP	Basic and Primary Education Project
CAS	Continuous Assessment System
CB-EGRA	Classroom-Based Early Grade Reading Assessment
CDC	Curriculum Development Centre
CERID	Center for Educational Research, Innovation, and Development
CERSOD	Center for Educational Research and Social Development
CTT	Classical Test Theory
CWPM	Correct Words Per Minute
DG	Director General
DOE	Department of Education
EDSC	Education and Development Service Centre
EGMA	Early Grade Mathematics Assessment
EGN	Early Grade Numeracy
EGR	Early Grade Reading
EGRA	Early Grade Reading Assessment
ERO	Education Review Office
FEDUC	Foundation for Educational Change
GON	Government of Nepal
GPF	Global Proficiency Framework
IRT	Item Response Theory
ITAD	Information, Training, and Development
MOE	Ministry of Education
MOEST	Ministry of Education, Science, and Technology
NARN	National Assessment of Reading and Numeracy

NCF	National Curriculum Framework
NEGRP	National Early Grade Reading Program
NEQMAP	Network for Educational Quality and Monitoring in Asia-Pacific
NESP	National Education System Plan
NORC	National Opinion Research Center
ORF	Oral Reading Fluency
PASEC	Program for the Analysis of Educational Systems of CONFEMEN
PILNA	Pacific Islands Literacy and Numeracy Assessment
PIRLS	Progress in International Reading and Learning Study
PISA	Program for International Student Achievement
PLD	Proficiency Level Descriptor
RC	Reading Comprehension
RTI	RTI International
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SC	Steering Committee
SDG	Sustainable Development Goal
SLC	School Leaving Certificate
SME	Subject Matter Expert
TERCE	Third Regional Comparative and Explanatory Study
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute for Statistics
USAID	United States Agency for International Development
WFP	World Food Program

1. International Context

The objective of the UIS benchmarking study in Nepal is to produce a comprehensive case study on the benchmark setting process for precursor skills in reading. The study focuses on benchmarking for reporting on Sustainable Development Goal (SDG) 4.1.1, which is a powerful international indicator of student learning outcomes at three levels: a) in Grade 2 or 3; b) at the end of primary education; and c) at the end of lower secondary education. The indicator is used for reporting on the proportion of children and young people achieving at least minimum proficiency in reading and mathematics at each of the three levels, by sex.

Countries must have national student assessment data for reporting on 4.1.1, and the data must be recent (see below). The UNESCO Institute for Statistics (UIS) lists possible cognitive tests for use with 4.1.1: EGRA/EGMA, PASEC, PILNA, PIRLS, PISA, SACMEQ, TERCE, and TIMSS. Of these, EGRA/EGMA can be implemented at the national level and the others are implemented at the regional or international level while providing data at the national level. Countries may also administer their own national assessments not on the UIS list.

Once a country has assessment data, they need to set or adopt benchmarks, apply the benchmarks to the assessment data, and report to UIS on the percentage of students achieving minimum proficiency. Of the three SDG 4.1.1 levels, the least amount of data is available for 4.1.1a. According to the current UIS Data Gaps dashboard, only 35 out of 205 countries (17.1%) have reported on 4.1.1a. The low rate is likely due to countries not having the assessment data and benchmarks, although in some cases they may not have reported their results to UIS.

Over the past several years, UIS, Gates Foundation, U.S. Agency for International Development (USAID), and Australian Council for Educational Research (ACER) have supported pilot efforts by psychometricians and education specialists to improve the benchmarking process. This has included developing benchmarking methods, training country-level specialists, and applying the methods to assessment data for calculating the percentage of students achieving minimum proficiency.

An important issue has been how to ensure reasonably fair reporting across countries so that the assessment results are comparable. As mentioned, countries use different assessments to measure

learning outcomes. They also use different benchmarking methods, if they have benchmarks at all. This leads to non-standardized reporting. It is in contrast to measuring and reporting on indicators such as enrollment rates or student-teacher ratios that have standardized methodologies.

Two methods have been proposed by international psychometricians to compensate for the non-standardization of student assessments and increase the comparability of reporting on learning outcomes across countries. One method involves statistical linking, which requires equating assessments using common items or common people across tools, with minimum proficiency benchmarks adjusted based on the equating. This method has practical issues, whether with embedding common items in tools or administering the same assessments to a sufficient sample of common people. It also needs trained psychometricians to lead the equating process. Another method involves non-statistical linking, which requires setting minimum proficiency benchmarks by applying common descriptions. The descriptions can be in the form of the Global Proficiency Framework (GPF), the development of which was supported by UIS and others. The non-statistical method has fewer practicality issues since education specialists in countries make judgments about the placement of benchmarks on their assessment tools in relation to the common descriptions. It does not need trained psychometricians to lead the judgmental process. With either statistical or non-statistical linking, benchmarks can be set on different assessments that reflect the relative difficulty of those assessments.

Substantial resources have been devoted to developing and piloting benchmarking methods using statistical and non-statistical methods. However, even after solving the technical issues, the most expensive and time-consuming part of assessing and reporting on student learning remains collecting assessment data at appropriate grade levels on a regular basis. Benchmarking takes a short amount of time after data collection and before reporting.

Even though benchmarking is a relatively small part of the whole assessment process, it is important to determine best practices for setting the benchmarks. The first question is whether the benchmarks are set using an internationally accepted method. A second question is whether the benchmark setting follows the method with fidelity and efficiency. A third question is whether the benchmarks are acceptable for national and international reporting. A fourth question is whether countries only set a single benchmark for (minimum) proficiency or whether they set multiple

benchmarks for performance categories. A fifth question is whether the benchmarks are appropriate for the selected grade levels.

2. Structure of the Report and Method Used

Following the UIS scope of work, the Nepal case study on setting benchmarks at the early grades in reading attempts to answer these questions. It includes ten sections:

1. International Context: Provides the overall structure and outline of the report, including the international context and existing track record in measurement and the use of measurement for improving foundational learning performance.
2. Structure of the Report and Method Used: Describes the overall structure of the report and the method and process used in preparing this report.
3. Background: Details the history and context of the drive towards foundational learning in Nepal, including the influence of the SDG process, involvement of donors, NGOs, and units within the Ministry of Education.
4. Early Grade Reading Benchmark Setting Practices: Presents the overview of the initial early grade reading benchmark setting activities and the activities on revising the early grade reading benchmark in Nepal.
5. Decision-Making Process on Benchmark Setting: Describes how the decision was made to numerically benchmark skills, including links to programs for reading improvement and curricular statements.
6. Expert and Participant Selection for Benchmark Setting: Explains how experts and participants were chosen for the benchmarking process.
7. Benchmarking Process Methodology: Describes the technical aspects of the benchmarking methods used, including statistical techniques and the use of international examples or evidence. It also includes details of the methodology used in the benchmarking process, including the number of sessions, logistics, and participation of local experts.
8. Benchmarking Results: Presents the benchmarking numerical levels for key skills and discusses their psychometric properties and substantive meaning in terms of reading

science.

9. Uses of the Benchmarks: Explains how the benchmarks have been used or are planned to be used, including target setting and the degree of official adoption.
10. Benchmarking Lessons Learned: Provides lessons learned and advice for other countries and the global community, including UIS.

The case study uses data from a literature review and interviews with relevant specialists. The literature review includes several documents on the global assessment context and Nepal's assessment practices, including general assessment, early grade reading assessment (EGRA), and EGR benchmarking. The reference section has a list of the documents reviewed. The interviews were with specialists from government agencies and INGOs/NGOs working in EGR. The annex has a list of the people interviewed with descriptions of their positions and experiences as of the interview dates. Additional information on the process was collected from interviews for this study with personnel from the Ministry of Education, Science, and Technology (MOEST), Education Review Office (ERO), Curriculum Development Center (CDC), National Early Grade Reading Program (NEGRP), Early Grade Reading Program (EGRP), Plan International, and Foundation for Education Change (FEDUC), along with other specialists who participated in setting benchmarks in both 2017 and 2021**.**

3. Background

Early Grade Reading Program

Nepali language is a compulsory subject with a focus on the four skills of reading, writing, speaking, and listening. However, teaching-learning activities and assessment strategies traditionally concentrate on writing. This was a factor in poor reading achievement on the national EGRA conducted in 2014 (see below). An analysis showed that one of the key issues in EGR was a lack of comprehensive and balanced approach to the teaching of reading, with problems in areas such as 1) systematic teaching of decoding, 2) adequate emphasis on comprehension, 3) availability of appropriate materials for developing fluency; and 4) regular assessment of skills (MOE, 2014). To address these issues, the Government of Nepal (GON), with the support from USAID, designed a National Early Grade Reading Program (NEGRP) program beginning in 2014-

2015. NEGRP was designed to improve the reading performance of students in grades 1-3 and build the GON's capacity for the program.

The GON implemented the NEGRP through the USAID-funded Early Grade Reading Program (EGRP) in 22 districts, i.e. 6 districts starting in 2016, 10 districts starting in 2017, and 6 additional districts when the program was extended in 2020 for two years. It had six components: 1) instructional design; 2) materials development and distribution; 3) teacher professional development and instructional support; 4) continuous assessment and remedial support; 5) community development; and 6) research, monitoring, and evaluation. (NORC, 2020).

The end-line evaluation of NEGRP showed some improvement in reading achievement in program-supported schools (NORC, 2020). Based on the evaluation, the CDC adopted materials, instructional, and assessment strategies developed during program implementation in its revised curriculum for grades 1-3 in 2019 (CDC, 2019). In this new program, about 19% percent of classroom instructional time is allocated for Nepali and 15% percent for English. Schools may also teach a local language though this is not compulsory (CDC, 2019).

Realizing the need for further strengthening the NEGRP and early grade numeracy (EGN), the GON and USAID designed and implemented a new USAID-funded Early Grade Learning (EGL) program in 48 districts starting in late 2023. A continuous support system was needed for improvement in EGR and EGN, particularly with enhancing teachers' instructional capacity and assessment practices. However, this program was suspended in early 2025.

Large-Scale Assessment

There have been four main categories of large-scale, sample-based student assessment implementation at the national level since 1997. Each successive assessment showed improvements in technical processes, including in content, psychometrics, and benchmarking.

The first category involved national assessments commissioned to external agencies. From 1997 to 2011, the Basic and Primary Education Project (BPEP) and the Department of Education (DOE) of the GON commissioned large-scale national assessments of student achievement by consulting firms for grades 3, 5, and 8 (ERO, 2013; Poudel, 2017; Poudel & Bhattarai, 2018). Several assessments were conducted during this period, such as those by the Education Development and

Service Center (EDSC) in 1997, 1999, 2001, 2003, 2008, and 2011; the Center for Educational Research, Innovation, and Development (CERID) in 1999; and the Center for Educational Research and Social Development (CERSOD) in 2001. However, the usefulness of the assessments was limited due to the absence of a national assessment framework and system. Moreover, the use of Classical Test Theory (CTT) with percentage scores limited comparability of results across the years (Poudel, 2016).

The second category included national assessments conducted by the GON's Education Review Office (ERO) after its establishment in 2010 until 2015. During this period, ERO conducted two rounds of national assessments at grades 3, 5, and 8 (ERO, 2013, 2015a, 2015b). Although the use of standardized items, representative samples, and contextual surveys improved the quality and rigor of these assessments, the lack of a well-defined national assessment framework and continued reliance on CTT with percentage scores resulted in limited comparability across the years. Furthermore, there was a lack of standards and benchmarks.

The third category included national assessments at grades 5 and 8 starting in 2016 (ERO, 2016). The first assessment using a well-defined national framework and Item Response Theory (IRT) took place in 2017, both of which the ERO has continued to use in subsequent assessments (ERO, 2018, 2019, 2021, 2022). IRT has allowed the ERO to compare results across the years. In addition, the ERO produced scale scores with benchmarks to categorize the scores into four levels, with descriptors on what students typically know and do at those levels.

The fourth category included national assessments that focused on EGR and EGN starting in 2014. In the first nationally representative sample-based EGRA, the reading levels of students in grades 2 and 3 were identified using individually administered tools, with disaggregated data by geographical location, socio-economic status, and language use at home (RTI, 2014). There were also surveys on school-related factors that influence learning. However, the assessment only reported the percentage of correct responses in the sub-content areas of oral reading fluency (ORF), reading comprehension (RC), letter sounds, listening comprehension, and Matra. It did not use an assessment framework, composite scores, benchmarks, or IRT. In the second nationally representative sample-based grade 3 reading and numeracy assessment, ERO developed an assessment framework for both subjects in 2020. It is called the National Assessment of Reading

and Numeracy (NARN), which included revised tasks from EGRA and Early Grade Math Assessment (EGMA) (ERO, 2020). The reading assessment covered ORF, RC, and other sub-content areas. Scores were reported using IRT and benchmarks, including a combined benchmark for ORF and RC proficiency. This assessment fulfilled the majority of the reporting requirements for EGR and early grade numeracy (EGN) under SDG 4.1.1a.

In addition to these two national assessments, EGRP commissioned external agencies to conduct three EGRAs to establish the baseline, midline, and endline levels of students' reading in the 22 program districts (NORC, 2017, 2018; RTI, 2022). Their objective was to evaluate EGRP interventions aimed at improving reading in grades 2 and 3. Thus, these EGRAs had a more limited geographical scope compared to the 2014 EGRA and the 2020 NARN. In addition to EGRP, some INGOs conducted EGRAs to evaluate the results of interventions in their project areas (Such as World Education, 2017; Room to Read, 2019). While none of these sub-national EGRAs were appropriate for national reporting, some used benchmarks.

4. Early Grade Reading Benchmark Setting Practices

The CDC led an initial benchmark setting process for EGR in 2017 with technical support for NEGRP. They selected a team comprised of external experts from EGRP and subject matter experts (SMEs) from CDC (CDC, 2017).

Unfortunately, the report showed that there was a lack of benchmarking understanding and capacity among the team members, along with limited data from EGR assessments in Nepal. Although the report stated that the process involved a literature review, consultations, training, benchmarking workshops, and teacher feedback, it did not provide evidence that clearly linked the activities described to the benchmarks. A combined benchmark with ORF of 45 correct words per minute (CWPM) and RC of 80% correct responses was presented in a table with a lack of contextual information or an explanation of an adequately standardized or evidence-based benchmarking process (CDC, 2017).

After it was endorsed by the EGR Steering Committee (SC) at the MOEST, two main concerns arose among experts and teachers regarding the combined benchmark. First, applying the same benchmark to grades 1, 2, and 3 students posed significant challenges due to use of the same texts

across different grades for the students with different learning levels. Second, having a single benchmark was not appropriate for students with varying levels, including those from a diverse population with first and second language learners. It would be more effective to establish sets of benchmarks to accommodate ongoing progress in learning to reading. In this regard, the combined benchmark with ORF of 45 CWPM and RC of 80% correct responses could be perceived as an advanced level of learning at grade 3 rather than as a national benchmark for all students in grades 1, 2, and 3.

The technical reports for the 2018 EGRP EGRAs, the 2019 Room to Read EGRA, the 2020 NARN, and the 2022 EGRP EGRA reported on the percentages of students achieving the combined benchmark. In addition, the 2020 NARN reported on four ORF levels for grade 3 by setting cut-off points between adjacent levels (ERO, 2020):

1) No reader: A student who cannot read a single word correctly in one minute.

2) Initial reader: A student who can read up to 1–15 words correctly in one minute.

3) Emergent reader: A student who can read between 16 to 44 words correctly in one minute.

4) Fluent reader: A student who can read over 45 words correctly in one minute.

Following the 2020 NARN, the ERO worked with NEGRP to address concerns from the 2017 benchmark setting and set a new benchmark for early grade reading. The subsequent sections of this report present the benchmarking process and results of the 2021 EGR benchmark setting.

5. Decision-Making Process on Benchmark Setting

ERO collaborated with the EGR SC to establish and implement a new benchmark setting process in 2021 (ERO, 2022a). The EGR SC selected an eight-member task team, led by the Director General (DG) of ERO, for planning, coordinating, and implementing the revised EGR benchmark setting process. The ERO led the decision-making progress on benchmark setting, with implementation by the task team.

6. Expert and Participant Selection for Benchmark Setting

The task team appointed a national expert to oversee the process, review literature, conduct workshops, and prepare reports. They selected 10 subject matter experts (SMEs) for the benchmark setting, with two experts from each of five categories: classroom teachers, curriculum and materials development experts, assessment experts, teacher trainers, and faculty members from teacher preparation colleges. Additionally, ERO and EGRP experts participated in both workshops and reviewed the performance level descriptors (PLDs) and the benchmarks.

7. Benchmarking Process Methodology

The ten-step process in setting the early grade benchmarks was as follows:

1) Task Team Formation and Expert Selection: The EGR SC selected an eight-member task team led by the DG of the ERO. The EGRP in consultation with the task team appointed a national expert to oversee the process, review literature, conduct workshops, and prepare reports.

2) Development of Concept Note and Roadmap: The task team and experts developed a concept note and a roadmap for the benchmark setting process. They reviewed national and international benchmark setting practices and agreed on a process. The task team made plans to organize meetings and conduct workshops according to the road map.

3) Key Reading Skills Identification: The task team selected two reading skills for national benchmark setting: ORF and RC.

4) Proficiency Levels Determination: The task team determined four proficiency levels: pre-basic, basic, proficient, and advanced.

5) Preparation of PLDs: The national expert reviewed GPF for reading and the Nepali language curriculum for grades 1-3. They prepared an initial draft of the PLDs across three reading skills: decoding, ORF, and RC. The national expert presented this draft at a two-day workshop with task team members and EGR specialists, where the PLDs were reviewed and revised.

6) Selection of a Benchmark Setting Method: A modified version of the Angoff method for benchmark setting was selected to set three cut-off points between adjacent proficiency levels.

This method was selected given the item types and criterion-referenced interpretation of the NARN and EGRA scores.

7) Training for SMEs: A half-day workshop was organized to train the ten SMEs, called panelists, in setting cut-off scores using the modified Angoff method. The training included orientation on the benchmark setting process, EGRA tools, proficiency levels, descriptors, cut-off points, rating sheets, and rounds. They reviewed the method and materials, with questions and discussions.

8) Benchmark-Setting Meeting Organization: The ERO organized a one-and-a-half-day meeting to set the benchmarks at grades 1, 2 and 3. For round 1, each panelist estimated ORF in the two grade levels using the number of CWPM that students at the basic, proficient, and advanced levels could read. Then they estimated RC for grades 2 and 3 students using the number of questions that students at the basic, proficient, and advanced level could answer correctly. The data for round 1 were calculated and presented to the panelists, with discussion. For round 2, the panelists used their understanding and experience from round 1 to refine their ratings. The round 2 data were calculated, and the consistency was checked. Three cut-off points for four levels of ORF and RC were established through average ratings from each panel member.

9) Review, Validation, and Recommendations: The task team reviewed the benchmark setting to confirm its validity. They examined the consistency of panelists' ratings to confirm their reliability. The task team evaluated the three cut-off points between adjacent proficiency levels with score ranges for each of the four proficiency levels in ORF and RC. The ORF cut-off points were rounded to the nearest five points.

10) Field Testing and Endorsement: Following a presentation and discussions with the EGR SC, the ERO field tested the proposed benchmarks with a group of students. Upon analyzing the achievement data in relation to the benchmarks, adjustments were made to the grade 2 ORF benchmark for the basic level. The benchmarks show a progression in ORF and RC for students in grades 1-3. These benchmarks were recommended for use with grade-level appropriate reading passages. After the adjustments and explanations, the EGR SC endorsed the final benchmarks and the accompanying document.

8. Benchmark Results

The following three tables present the results of the benchmarking activities led by the ERO and the EGR SC in 2021 (ERO, 2022a).

This table shows three cut-off points for ORF and RC by early grade:

Reading sub-skill	Basic	Proficient	Advanced
Grade 1			
ORF (CWPM)	15	25	35
RC (% correct response)	20	40	60
Grade 2			
ORF (CWPM)	15	30	40
RC (% correct response)	30	50	70
Grade 3			
ORF (CWPM)	20	35	45
RC (% correct response)	30	60	80

This table presents four proficiency levels of ORF and RC with ranges by early grade:

Reading sub-skill	Pre-basic	Basic	Proficient	Advanced
Grade 1				
ORF (CWPM)	Less than 15	15 to 24	25 to 34	35 and more
RC (% correct response)	Less than 20	20 to 39	40 to 59	60 and above
Grade 2				
ORF (CWPM)	Less than 15	15 to 29	30 to 39	40 and more
RC (% correct response)	Less than 30	30 to 49	50 to 69	70 and above
Grade 3				
ORF (CWPM)	Less than 20	20 to 34	35 to 45	45 and more
RC (% correct response)	Less than 30	30 to 60	60 to 79	80 and above

This final table gives the combined benchmarks for ORF and RC by grade of the proficient early grade reader:

Reading sub-skills	Proficient
Grade 1	
ORF (CWPM) and RC (% correct responses)	25 and 40%
Grade 2	
ORF (CWPM) and RC (% correct responses)	30 and 50%

Reading sub-skills	Proficient
Grade 3	
ORF (CWPM) and RC (% correct response)	35 and 60%

9. Uses of the Benchmarks

As initially established, between 2017 and 2021, EGRAs used the combined benchmark of 45 CWPM for ORF and 80% correct responses for RC. Additional proficiency levels for ORF used on the 2020 NARN were 1) No reader: A student who cannot read a single word correctly in one minute; 2) Initial reader: A student who can read up to 1–15 words correctly in one minute; 3) Emergent reader: A student who can read between 16 to 44 words correctly in one minute; and 4) Fluent reader: A student who can read over 45 words correctly in one minute. Data was reported by the EGRAs and NARN using the

The ERO, in collaboration with the NEGRP, updated the benchmarks in 2021 (ERO, 2022a). However, those benchmarks have not been used. The sub-national, group-administered CB-EGRA used statistical technique to estimate the percentage of grade 3 students achieving the previous combined ORF and RC benchmark (RTI, 2022). NARN has not taken place over the past three years. It is anticipated that the new benchmarks will be used in forthcoming assessments since their development was officially within the MOEST, under the guidance of ERO and endorsement by the ERG SC.

Aside from reporting on student achievement of the benchmarks, there have been initial discussions on target setting. It is a complement to benchmarking as targets are established on an annual or longer-term basis using available data. Targets represent the aim of reducing the proportion of students at lower proficiency levels while increasing the proportion of those at higher levels. They should be grounded in the interventions or programs designed to enhance EGR. As suggested by the benchmark-setting report of 2021, it is crucial for policymakers, experts, and practitioners to collaborate in establishing targets for the EGR skills of Nepali students.

10. Benchmarking Lessons Learned

Because of the challenges associated with using a single benchmark of 45 CWPM for ORF and

80% correct responses for RC established in 2017, the ERO, together with NEGRP, revised and expanded the benchmarks in 2021. The updated benchmark was based on an internationally accepted benchmarking method (modified Angoff) along with PLDs derived from the Nepali national curriculum and the GPF. It outlines four levels of learning with three cut-off scores for each of ORF and RC, which allows placement of learners on a continuum and offers them opportunities to progress to higher levels of achievement. This revised benchmark has gained widespread acceptance in Nepal and is expected to be used in future early reading assessments.

One of the issues discussed during the benchmark setting process was the difference between proficiency and minimum proficiency. SDG 4.1.1 and the GPF focus on minimum proficiency while the Nepali benchmarks focus on proficiency. This may need revisiting. For instance, for international reporting purposes, it is possible that the basic level is more appropriate for minimum proficiency than the proficiency level.

Another issue is the number of comprehension questions in each set of assessment tools. Previous EGRAs conducted in Nepal used one reading paragraph with five comprehension questions in each set of tools. Recent guidance and evidence have shown that a minimum of 10 questions in RC is necessary to implement the 2021 reading benchmark. Thus, it is recommended to include two reading passages with five questions each for the next NARN and EGRA assessments. This approach will not only support in categorizing students at different achievement levels but will also enhance the overall reliability of the sub-content areas and the assessments. Regarding the time required for assessments, administering a test on reading and comprehension of two passages would take approximately ten minutes, which should not pose significant field implementation or assessment cost issues.

The next NARN and EGRAs are recommended to implement the following based on the workshop report: 1) use the new benchmarks and proficiency levels, as agreed upon in the 2021 benchmark setting workshop; 2) apply the modified Angoff process – as needed for the upper grades – and document the choice of the specific Angoff version; 3) produce additional evidence of consistency in the ratings – again as needed for the upper grades – for use during the benchmarking and in the reporting; 4) focus on building consensus in the benchmarking process, including meetings within the MOEST; 5) organize a meeting between the ERO and relevant stakeholders to request

compliance in benchmarking process, application of the new benchmarks to the data, and reporting of SDG 4.1.1 results to UIS.

References

CDC (2017). *Reading Benchmark (Text in Nepali).* CDC/GON.

CDC (2019). *National Curriculum Framework (NCF) for School Education in Nepal (Text in Nepali).* CDC/GON.

CDC (2019). *Basic Level Curriculum, Grades 1-3.* CDC, GON.

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

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

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

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

ERO (2017a). *Classroom-Based Early Grade Reading Assessment: A Framework*. ERO/GON.

ERO (2017b). *Report of Classroom-Based Early Grade Reading Assessment: Bhaktapur*. ERO/GON.

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

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

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

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

ERO (2022a). *Early Grade Reading Benchmark*. ERO/GON.

ERO (2022b). *Report on National Assessment of Student Achievement 2020 in Mathematics, Science, Nepali and English for Grade 8*. ERO, GON.

MOE (2014). *National Early Grade Reading Program (NEGRP), 2014/15-2019/20*. GON, MOE.

NORC (2017). *Impact Evaluation of the Early Grade Reading Program (EGRP) in Nepal: Design Report.* USAID/NORC.

NORC (2018). *USAID Early Grade Reading Program in Nepal Impact Evaluation Midline Results.* USAID/NORC.

NORC (2020). *Impact Evaluation of the Early Grade Reading Program in Nepal.* USAID/NORC.

Poudel, L. N. and 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.

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

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.*

Room to Read (2019). *Impact Evaluation of the Literacy Program in Nepal: 2019.* Room to Read.

RTI (2014). *Nepal Early Grade Reading Assessment (EGRA) Study*. USAID/RTI.

RTI (2020). *USAID's Early Grade Reading Program in Nepal: Final Progress Report, Years 1-5.* USAID/MOEST/RTI.

RTI (2021a). *Request for Proposal for Consultant to Support the Revision to the Reading Benchmark.* USAID/RTI.

RTI (2021b). *USAID's Early Grade Reading Program II (EGRP II) in Nepal: Baseline Report Vol. 1 and 2.* USAID/MOEST/RTI.

RTI (2022). *USAID's Early Grade Reading Program II (EGRP II) in Nepal: Endline Report.* USAID/MOEST/RTI.

WFP (2019). *USDA McGovern Dole Food for Education Program in Nepal: Evaluation Report.* WFP.

WFP (2022). *USDA McGovern Dole Food for Education Program in Nepal: Evaluation Report.* WFP.

World Education (2017). *Early Grade Reading Assessment Results 2015-2017*. World Education.

Annex: Descriptions of Interviews Participants

No.	Name	Agency	Position(s)/Experience
1	Purushottam Ghimire	Curriculum Development Centre (CDC)	Participated in EGR benchmark setting activities of 2017 and 2021; worked with EGR-related government activities for about 10 years; currently working at CDC.
2	Dr. Shyam Prasad Acharya	Curriculum Development Centre (CDC)	He worked for 13 years with ERO in designing assessment tools, analyzing data, and preparing reports; also worked in designing the 2020 NARN; currently working at CDC.
3	Yubraj Adhikari	Education Review Office (ERO)	Coordinates the research and development activities as well as the EGRA-related activities at ERO.
4	Narayan Jha	Education Review Office (ERO)	He worked for more than 20 years in the GON at public examination and student assessment offices; currently working as the unit chief of National Assessment of Student Achievement of ERO.
5	Devi Ram Acharya	Ministry of Education, Science and Technology (MOEST)	He worked for ERO in designing tools and analyzing data for national assessment; currently working at the Foreign Aid Coordination Unit of MOEST.
6	Dr. Ananda Paudel	Chemonics USAID Early Learning Program	He worked for MOEST for about 20 years in education and EGR; worked for different organizations in education and EGR; currently working for Chemonics as the Program Director.
7	Narayan Krishna Shrestha	Plan International USAID Equity and Inclusion Project	He worked for MOEST for about 20 years including as the planning and foreign aid coordination unit head; worked with RTI on EGR; currently working for Plan International as the Deputy Chief of Party for a USAID supported program.
8	Dhan Singh Dhami	Plan International USAID Equity and Inclusion Project	He worked for MOEST for over 15 years; worked with RTI on EGR; currently working for Plan International as a Technical Advisor for a USAID supported program.
9	Sagar Mani Neupane	Information, Training and Development (ITAD)	He worked in research and data analysis at the university grants commission; worked with RTI on benchmarking and with Room to Read; currently working for ITAD as a Senior Evaluation Lead.
10	Indra Bahadur Shrestha	Foundation for Educational Change (FEDUC)	He worked for RTI as the coordinator for the EGRA endline assessment in 2022 on behalf of FEDUC; currently working for FEDUC as a coordinator for activities related to research and consulting services in education.

uis.unesco.org

© UNESCO-UIS 2025