



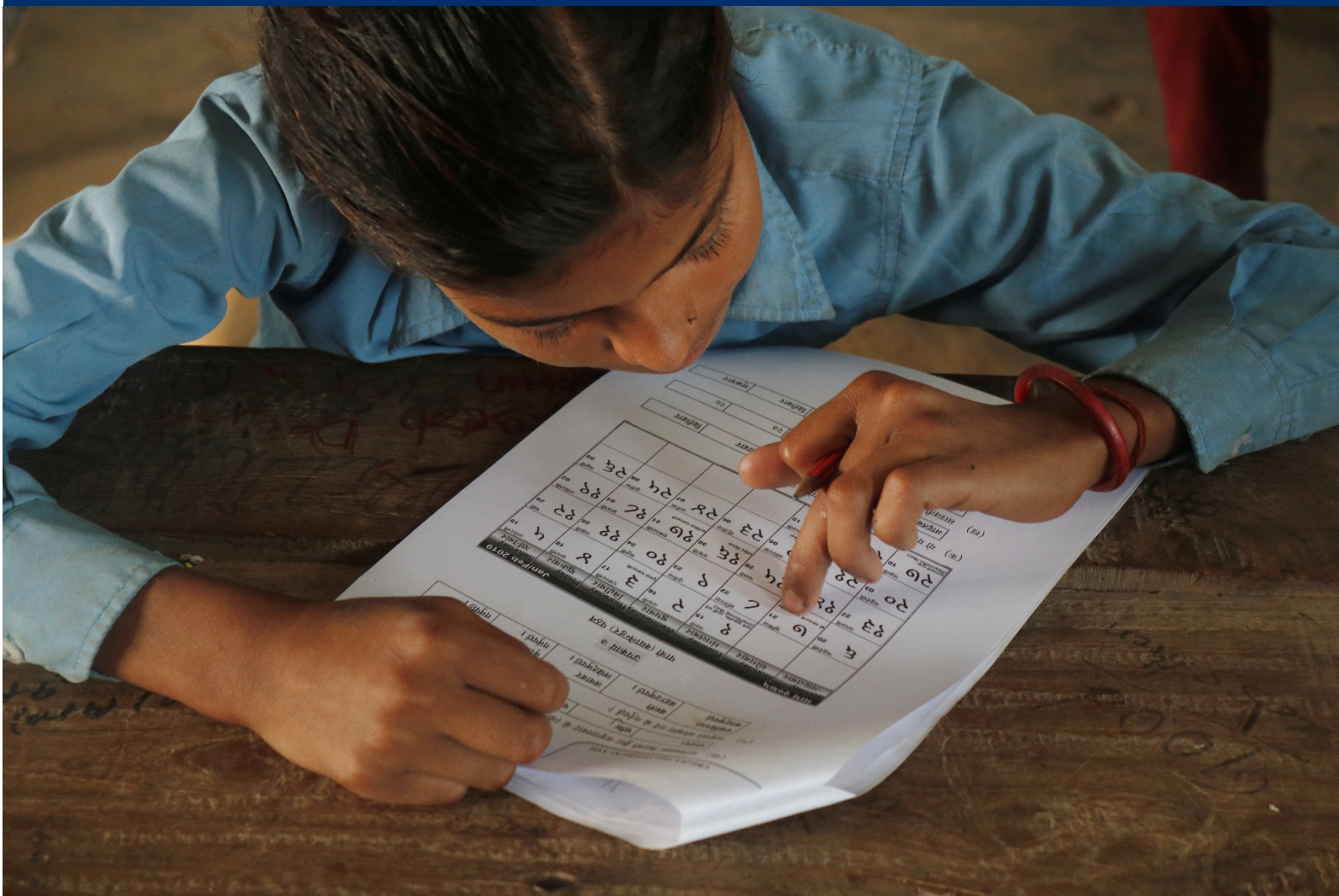
USAID
FROM THE AMERICAN PEOPLE



**Government of Nepal,
Ministry of Education,
Science and Technology**

USAID's Early Grade Reading Program II (EGRP II) in Nepal

Baseline Report Vol. 1, Student Reading Performance in the Early Grades



August 2021

This publication was prepared for review by the United States Agency for International Development. It was prepared by RTI International.

USAID's Early Grade Reading Program II (EGRP II) in Nepal

Baseline Report Vol. 1: Student Reading Performance in the Early Grades

Cooperative Agreement No. 72036720CA00001

Prepared for
Jayakrishna Upadhyay
Agreement Officer's Representative
Education Development Office
United States Agency for International Development/Nepal
GPO Box No. 295
U.S. Embassy, Maharajgunj
Kathmandu, Nepal
Email: jupadhyay@usaid.gov

Prepared by
Sagar Mani Neupane, Swadesh Maharjan, Simon King, Jodie Fonseca, and Wendi
Ralaingita
RTI International
P.O. Box 12194
Research Triangle Park, NC 27709-2194
USA
<http://www.rti.org/>

The authors' views expressed in this report do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

RTI International is a trade name of Research Triangle Institute. RTI and the RTI logo are U.S. registered trademarks of Research Triangle Institute.

Cover photo: A grade 3 student in Nawalparasi West District participating in a classroom-based early grade reading assessment during the baseline evaluation conducted by EGRP II. (Photo credit: Swadesh Maharjan)

Table of Contents

List of Tables	ii
List of Figures.....	iii
Acronyms and Abbreviations	v
Acknowledgments	vi
Executive Summary	1
1 Background	4
2 Study Design.....	6
2.1 Research Questions.....	6
2.2 Sample Design	6
2.3 Study Instruments	8
2.4 Study Quality Assurance.....	10
3 Study Findings	13
3.1 Grade 2 Findings.....	13
3.1.1 Overall Reading Achievement (Grade 2)	13
3.1.2 Reading Achievement by Subtask (Grade 2).....	14
3.1.3 Reading Achievement by Sex and L1 (Grade 2)	25
3.2 Grade 3 Findings.....	27
3.2.1 Overall Reading Achievement (Grade 3)	27
3.2.2 Reading Achievement by Subtask (Grade 3).....	28
3.2.3 Reading Achievement by Sex and Language (Grade 3).....	39
4 Linking EGRA and CB-EGRA Data.....	42
4.1 Rationale for Linking.....	42
4.1.1 The Model to Predict ORF (Grade 2)	43
4.1.2 The Model to Predict Reading Comprehension (Grade 2)	44
4.1.3 The Model to Predict ORF (Grade 3)	45
4.1.4 The Model to Predict Comprehension (Grade 3).....	46
4.2 Extrapolation of Reading Achievement.....	47
4.2.1 Equivalent Scoring.....	47
4.2.2 Students Who Met the Reading Benchmark.....	48
4.2.3 Comparing EGRP II Baseline Findings with the 2020 NARN and CB-EGRA Scores from Previous Years	48
4.2.4 Reading Ability Categories.....	50
5 Summary and Conclusions	51
6 Study Limitations	54
References	56
Annex A: EGRP II MEL Indicator Reporting.....	57

List of Tables

Table 1:	Study sample size	6
Table 2:	Description of grade 2 CB-EGRA assessment tool.....	8
Table 3:	Description of grade 3 CB-EGRA assessment tool.....	9
Table 4:	Average item scores for the grade 2 letter/matra identification subtask	15
Table 5:	Average item scores for the grade 2 word and sentence identification subtask	17
Table 6:	Average item scores for the grade 2 vocabulary subtask	18
Table 7:	Average item scores for the grade 2 dictation subtask	20
Table 8:	Average item scores for the grade 2 listening comprehension subtask (25 words).....	22
Table 9:	Average item scores for the grade 2 reading comprehension subtask (60 words).....	23
Table 10:	Average item scores for the grade 2 calendar reading subtask	25
Table 11:	Average percentage scores of grade 2 students, by subtask and sex.....	26
Table 12:	Average percentage scores of grade 2 students, by language	26
Table 13:	Average item scores for the grade 3 word and sentence identification subtask	30
Table 14:	Average item scores for the grade 3 vocabulary subtask	31
Table 15:	Average item scores for the grade 3 word separation subtask	32
Table 16:	Average item scores for the grade 3 dictation subtask	34
Table 17:	Average item scores for the grade 3 listening comprehension subtask (30 words).....	35
Table 18:	Average item scores for the grade 3 reading comprehension subtask (60 words).....	37
Table 19:	Average item scores for the grade 3 calendar reading subtask	39
Table 20:	Average percentage scores of grade 3 students, by subtask and sex.....	40
Table 21:	Average percentage scores of grade 3 students, by subtask and language.....	41
Table 22:	Equivalent CB-EGRA scores for emergent and fluent benchmarks (in cwpm)	48
Table 23:	Percentage of learners who met the reading benchmark (45 words and 80% comprehension), by grade	48
Table 24:	Categories of readers, by grade	50

List of Figures

Figure 1:	Sample size required for the desired level of precision.....	7
Figure 2:	Screen shots of data collection, real-time visualization, and monitoring systems	11
Figure 3:	Average percentage scores, by subtask (grade 2).....	13
Figure 4:	Distribution of overall average percentage scores (grade 2).....	14
Figure 5:	Student stimulus for the grade 2 letter/matra identification subtask	14
Figure 6:	Distribution of scores for the grade 2 letter/matra identification subtask	15
Figure 7:	Student stimulus for the grade 2 word and sentence identification subtask	16
Figure 8:	Distribution of scores for the grade 2 word and sentence identification subtask	16
Figure 9:	Student stimulus for the grade 2 vocabulary subtask	17
Figure 10:	Distribution of scores for the grade 2 vocabulary subtask	18
Figure 11:	Student stimulus for the grade 2 dictation subtask.....	19
Figure 12:	Distribution of scores for the grade 2 dictation subtask	19
Figure 13:	Student stimulus for the grade 2 listening comprehension subtask.....	21
Figure 14:	Distribution of scores for the grade 2 listening comprehension subtask (25 words).....	21
Figure 15:	Student stimulus for the grade 2 reading comprehension subtask	22
Figure 16:	Distribution of scores for the grade 2 reading comprehension subtask (60 words).....	23
Figure 17:	Student stimulus for the grade 2 calendar reading subtask	24
Figure 18:	Distribution of scores for the grade 2 calendar reading subtask	25
Figure 19:	Average percentage scores of grade 2 students, by sex.....	25
Figure 20:	Average percentage scores of grade 2 students, by language	26
Figure 21:	Average percentage scores, by subtask (grade 3).....	27
Figure 22:	Distribution of overall average percentage scores (grade 3).....	28
Figure 23:	Student stimulus for the grade 3 word and sentence identification subtask	29
Figure 24:	Distribution of scores for the grade 3 word and sentence identification subtask	29
Figure 25:	Student stimulus for the grade 3 vocabulary subtask	30
Figure 26:	Distribution of scores for the grade 3 vocabulary subtask	31
Figure 27:	Student stimulus for the grade 3 word separation subtask	31
Figure 28:	Distribution of scores for the grade 3 word separation subtask	32
Figure 29:	Student stimulus for the grade 3 dictation subtask.....	33

Figure 30:	Distribution of scores for the grade 3 dictation subtask	33
Figure 31:	Student stimulus for the grade 3 listening comprehension subtask.....	34
Figure 32:	Distribution of scores for the grade 3 listening comprehension subtask (30 words).....	35
Figure 33:	Student stimulus for the grade 3 reading comprehension subtask (60 words)	36
Figure 34:	Distribution of scores for the grade 3 reading comprehension subtask (60 words).....	36
Figure 35:	Student stimulus for the grade 3 calendar reading subtask	38
Figure 36:	Distribution of scores for the grade 3 calendar reading subtask	39
Figure 37:	Average percentage scores of grade 3 students, by sex.....	40
Figure 38:	Average percentage scores of grade 3 students, by language	40
Figure 39:	Regression model to describe the CB-EGRA average percentage score from ORF (grade 2).....	44
Figure 40:	Regression model to describe the overall CB-EGRA percentage score from reading comprehension (grade 2)	45
Figure 41:	Regression model to describe the CB-EGRA average percentage score from ORF (grade 3).....	46
Figure 42:	Regression model to describe the overall CB-EGRA percentage score from reading comprehension (grade 3)	47
Figure 43:	Percentage of grade 3 students who met the reading benchmark in the 2021 EGRP II baseline and the 2020 NARN	49
Figure 44:	Comparison of average CB-EGRA scores between previous CB-EGRA assessments and the EGRP II baseline	49

Acronyms and Abbreviations

ACT	A standardized test to measure secondary school students' academic college readiness (mainly US)
CB-EGRA	Classroom-Based Early Grade Reading Assessment
COVID-19	Coronavirus Disease 2019 (SARS-CoV-2)
cwpm	Correct Words Per Minute
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
EGRA	Early Grade Reading Assessment
EGRP II	Early Grade Reading Program II
ERO	Education Review Office
GON	Government of Nepal
IC	Integrated Curriculum
IND	Indicator Code
L1, L2	First Language, Second Language
LEU	Local Education Unit
MEL	Monitoring, Evaluation, and Learning
MOE	Ministry of Education (former name of MOEST)
MOEST	Ministry of Education, Science and Technology
N/A	Not Applicable
NARN	National Assessment for Reading and Numeracy
NEGRP	National Early Grade Reading Program
ORF	Oral Reading Fluency
<i>p</i> -value	Probability of occurrence by chance (example: *** $p < .01$)
RTI	RTI International (registered trademark and trade name of Research Triangle Institute)
SAT	A standardized test broadly used for college admission screening (mainly US)
St	Subtask
USAID	United States Agency for International Development

Acknowledgments

The authors would like to acknowledge the contributions of all stakeholders and agencies who made it possible to complete this baseline assessment. We appreciate the kind support of officials from the Education Review Office, who not only provided the data collection tools from their item bank but also rigorously supported the full baseline evaluation process, from the master training of trainers to data collection in schools. The Early Grade Reading Program II (EGRP II) monitoring, evaluation, and learning coordinators, who led the entire process in their regions, also played crucial roles in administering the assessment. EGRP II's central and regional office teams as well as district coordinators and local-level program officers—who monitored the classroom-based early grade reading assessment (CB-EGRA) data collection process and also collected mini-EGRA data—were equally invaluable for this study. Last but not least, we appreciate the efforts of all the teachers who worked as CB-EGRA assessors, and the head teachers who facilitated data collection in their schools despite the hardships caused by the coronavirus disease 2019 pandemic.

Executive Summary

The Early Grade Reading Program II (EGRP II) is a 2-year, United States Agency for International Development (USAID)-funded program of technical assistance to the Government of Nepal (GON) that is being implemented from June 1, 2020, to May 31, 2022. EGRP II's support to the GON is being provided in the context of the shift toward the recently developed integrated curriculum (IC), ongoing decentralization in Nepal's education governance system, and prolonged disruptions to teaching and learning due to the coronavirus disease 2019 (COVID-19) pandemic. EGRP II is implemented in 38 National Early Grade Reading Program (NEGRP) districts, covering 396 Local Education Units (LEUs). The program provides intensive support for the implementation of the NEGRP minimum package¹ in 22 districts where the Ministry of Education, Science, and Technology (MOEST) is expanding early grade reading activities (referred to as Levels 1 and 2) and continued technical assistance for the 16 districts that were targeted under the first Early Grade Reading Program (EGRP) that ran from 2015–2020 (called Level 3).

To understand the overall impact over the program period, EGRP II conducted a baseline study in February 2021, which will be followed by an endline study in 2022. This report (referred to as Volume 1) is complementary to a separate but related report (Volume 2) that provides the baseline findings for EGRP II home- and community-based schooling intervention in Province 2 (Neupane et al., 2021).

The baseline study aimed to answer the following research questions: (1) How do grade 2 and 3 students from the program districts perform in reading skills? (2) In what ways do those levels of reading performance differ for boys and girls? (3) Are there differences in the reading performance of students who speak Nepali as a first language (L1) versus those who speak Nepali as a second language (L2)? (4) What model describes the relationship between the classroom-based early grade reading assessment (CB-EGRA) and fluency, comprehension, and reading ability of the students? (5) What are the baseline percentages of emergent student readers and fluent student readers in grades 2 and 3 in program districts?²

The CB-EGRA instrument was the key tool used in the baseline study. It was developed by Nepal's Education Review Office (ERO), under the MOEST, and is a group-administered assessment instrument used to measure the reading abilities of early-grade students. The instrument measures four core reading components (phonological awareness, grapho-phonemic awareness, vocabulary, and comprehension), plus writing. For the study sample, 45 schools, with 758 students from grade 2 and 804 students from grade 3, were selected randomly using a sampling design that ensured estimates were representative of EGRP II's population (i.e., the universe of students enrolled in EGRP II-supported schools). Moreover,

¹ NEGRP minimum package: A costed set of interventions designed to improve early grade reading. It encompasses curriculum development, teaching and learning materials, teacher training and support, community and parent engagement, and monitoring and learning assessment. USAID's first Early Grade Reading Program, implemented from 2015 to 2020, assisted the GON in developing the minimum package.

² The emergent reader category (15 correct words per minute [cwpm]) was identified in the GON's 2020 National Assessment of Reading and Numeracy (NARN) study (ERO 2020). As of August 2021, the GON was in the process of adopting different categories of readers in addition to the current national benchmark for fluent readers (45 or more cwpm), similar to the 2020 NARN categories. Anticipating this revision to the benchmarks, this baseline report discusses the different categories of readers and not just the fluent-reader category.

to be able to statistically equate the CB-EGRA scores with relevant early grade reading assessment (EGRA) scores, we conducted a mini-EGRA³ of five children from each grade from each sampled school. The CB-EGRA was conducted by trained schoolteachers and the mini-EGRA was conducted by EGRP II staff. The EGRP II team put in place a number of measures to ensure data quality throughout the process, including rigorous training for the teachers; a real-time data collection and reporting system; monitoring by at least one EGRP II staff member in all schools when the teacher administered the CB-EGRA; real-time data plotting; and an instant feedback system.

The CB-EGRA tools for both grade 2 and grade 3 consisted of seven different subtasks (St). Students were given either full credit or no score for each of the 21 questions, with no partial points awarded. Overall achievement scores were calculated using a composite average of the scores across the subtasks, presented as a percentage.

On average, grade 2 students were able to correctly answer 6 out of the 21 total questions in the assessment. The average grade 2 scores for each subtask were 50.3% for letter/matra⁴ identification; 45.5% for word and sentence identification; 36.5% for vocabulary; 9.4% for dictation; 51.5% for listening comprehension; 25.3% for reading comprehension; and 19.8% for calendar reading.

These results indicate that grade 2 students performed somewhat better on the listening comprehension subtask and that the average student was able to correctly respond to about half of the questions in this subtask. However, in general, students found the dictation subtask most difficult and generally left more than 90% of the items in this subtask incomplete or incorrect.

Similarly, on average, grade 3 students were able to correctly answer 7 out of the 21 total questions in the assessment. The average grade 3 scores for each subtask included 51.1% for word and sentence identification; 27.3% for vocabulary; 23.8% for word separation; 15.3% for dictation; 54.8% for listening comprehension; 44.8% for reading comprehension; and 23.1% for calendar reading.

These results indicate that, as in grade 2, students in grade 3 performed better on listening comprehension, with the average student able to solve about half of the questions from that subtask. Also Similar to the grade 2 findings, dictation was the most difficult subtask in the grade 3 assessment.

Student achievement was not significantly different by sex for either grade. In grade 2, on average, girls and boys were both able to correctly answer 6 questions out of 21. In grade 3, girls and boys both could answer 7 out of 21 questions correctly on average. However, student performance varied significantly by their home language. For grade 2 children, the average score for students who indicated that they speak Nepali as their L1 was 7 out of 21 questions answered correctly, whereas for students speaking Nepali as their L2, the average score was 4 questions correct out of 21. A similar result was observed in grade 3, where the average score for L1 students was 8 questions out of 21 correct while for L2 students it was 4

³ The “mini-EGRA” included only reading fluency and comprehension subtasks, whereas full EGRAs typically include more than just those two subtasks.

⁴ *Matras* are Nepali consonant letters accompanied by vowel signs.

questions out of 21. These results indicated a difference of about one grade level in reading outcomes between L1 and L2 children.

In addition to answering the research questions noted above, as part of the baseline activity, EGRP II aimed to develop a model to link the CB-EGRA with EGRA reading benchmarks. This effort would ensure that there is a simple method for assessing progress on early grade reading skills in Nepal while also reporting on standard and custom learning outcome indicators (e.g., ES. 1-1) for EGRP II. The EGRP II team anticipates that this model will be helpful for extrapolating reading fluency, reading comprehension, and overall reading ability using a tool that is simpler and cheaper than the full EGRA, and that has wide stakeholder buy-in in Nepal.

The process of assessment linking is a common and accepted practice to create equivalent scores between two assessments. For this evaluation, EGRP II conducted extensive statistical analysis to develop a rigorous model for using student performance on the CB-EGRA to predict their oral reading fluency (ORF) and comprehension skills.

The following are the statistical models (explained in depth, with graphics, in the body of the report) that the team developed and used to extrapolate children's early grade reading ability from CB-EGRA scores through this analytical process.

- Average grade 2 CB-EGRA percentage score = $19.901 + 0.911 \times \text{ORF}$
- Average grade 2 CB-EGRA percentage score = $24.003 + 9.201 \times \text{average comprehension}$
- Average grade 3 CB-EGRA percentage score = $22.399 + 0.817 \times \text{ORF}$
- Average grade 3 CB-EGRA percentage score = $28.149 + 7.674 \times \text{average comprehension}$

Using this methodology, EGRP II can report on the percentage of emergent and fluent readers, but not on student performance at the nonreader level (ORF = 0 correct words per minute [cwpm]) or the initial reader level (ORF between 1 and 15 cwpm). This limitation occurs because most of the CB-EGRA subtasks contain multiple-choice questions, thereby enabling children to avoid nonzero scores by chance.

Using the statistical equating approach, EGRP II determined that 7.4% of grade 2 children and 12.6% of grade 3 children met the GON's current national benchmark for reading fluency (45 cwpm with 80% comprehension). Furthermore, 27.8% of grade 2 and 29.7% of grade 3 students fell into the emergent reader category.

These EGRP II baseline findings generally align with results from a similar grade 3 assessment conducted in 2020 by the GON called the National Assessment for Reading and Numeracy (NARN; see ERO 2020). However, the EGRP II baseline findings were substantially lower than the CB-EGRA results from recent years. This baseline report discusses the potential reasons for these similarities and differences in recent reading assessment results in Nepal as well as caveats that must be considered when analysts are attempting to compare diverse assessment findings.

I Background

EGRP II is a 2-year, USAID-funded program of technical assistance to the GON that is being implemented from June 1, 2020, to May 31, 2022. EGRP II's support to the GON is being provided in the context of the shift toward the recently developed IC, ongoing decentralization in Nepal's governance system, and prolonged disruptions to teaching and learning due to the COVID-19 pandemic.

Building on the foundation of the first EGRP from 2015 to 2020, EGRP II aims to improve early grade literacy for students in grades 1–3 in Nepali public schools by supporting IC development and rollout (Objective 1), building local capacity for early grade reading service delivery (Objective 2), improving teacher professional support (Objective 3), and assisting with the COVID-19 response in the education sector (Objective 4).

EGRP II is implemented in 38 NEGRP districts, covering 396 LEUs. EGRP II has grouped the 38 target districts into three levels, as follows.

- **Level 1** includes the 10 districts that were meant to begin in-school implementation in 2020–2021, as well as the 8 districts that are meant to begin NEGRP implementation in the 2021–2022 school year: Achham, Baglung, Bara, Bhojpur, Dailekh, Doti, Kapilvastu, Khotang, Mahottari, Myagdi, Nawalparasi West, Rautahat, Rolpa, Salyan, Sarlahi, Sindhuli, Sindhupalchok, and Siraha.
- **Level 2** consists of the next four NEGRP rollout districts: Dhanusha, Rasuwa, Tanahun, and Taplejung.
- **Level 3** includes the 16 EGRP-supported districts where NEGRP initially rolled out: Banke, Bardiya, Bhaktapur, Dadeldhura, Dang, Dhankuta, Dolpa, Kailali, Kanchanpur, Kaski, Manang, Mustang, Parsa, Rupandehi, Saptari, and Surkhet.

The program provides intensive support for implementation of the NEGRP minimum package in the 22 Level 1 and 2 districts, and continued technical assistance for the 16 districts that were targeted under EGRP (Level 3). EGRP II operates from a Kathmandu central office as well as four regional offices. Supported by one district coordinator per district, one local-level program officer in eight Province 2 districts, and other regionally based technical staff, EGRP II works closely with LEUs and other local government staff to plan for and roll out NEGRP activities. The district coordinators are embedded in Education Development Coordination Units at the district level, and the local level program officers are embedded within LEU offices, to support LEUs in implementing activities such as training rollout, monitoring and use of data for decision making, and building of LEU skills in teacher professional support.

To understand the project impact over the program period, EGRP II conducted a baseline study in February 2021 and will undertake an endline in 2022. Although the project started in June 2020, we timed the baseline study to align with the end of the academic year in Nepal. The public school year usually ends in February–March, although the COVID-19 pandemic resulted in some disruptions to the usual timing and the 2020–2021 school year ultimately was extended for a few months beyond March 2021.

To assess student reading ability, the CB-EGRA was conducted by trained teachers in the sampled schools. The CB-EGRA was developed by Nepal's ERO, under the MOEST, as a

group-administered assessment of reading abilities for students in the early primary grades. The CB-EGRA assesses four reading components (phonological awareness, grapho-phonemic awareness, vocabulary, and comprehension) and writing. ERO has developed a CB-EGRA item bank, and this instrument has become an important assessment tool under the NEGRP and the national School Sector Development Plan.

However, because it is a group-based test, the CB-EGRA does not directly assess students' reading fluency. To overcome this limitation of the CB-EGRA, EGRP II simultaneously conducted a subsample-based mini-EGRA consisting of an oral reading passage and related comprehension subtasks. The aim was to use a statistical model to produce equivalence scores between skills measured by the CB-EGRA and the EGRA-measured skills of reading fluency and comprehension. By describing this statistical model, EGRP II has produced a tool that can be used in future assessments, tapping into the CB-EGRA assessment approach and avoiding the need to conduct a more expensive and complex EGRA.

2 Study Design

2.1 Research Questions

The EGRP II baseline study was designed to answer a specific set of questions.

1. How do grade 2 and 3 students from the program districts perform in reading skills?
2. In what ways do those levels of reading performance differ for boys and girls?
3. Are there differences in the reading performance of students who speak Nepali as a first language (L1) versus those who speak Nepali as a second language (L2)?
4. What model describes the relationship between CB-EGRA scores and the fluency, comprehension, and reading ability of the students?
5. What are the baseline percentages of emergent and fluent student readers in grades 2 and 3 in program districts?

2.2 Sample Design

EGRP II is working in 38 program districts covering 396 palikas⁵ and supporting approximately 13,500 schools. As such, 328,929 students from grade 2 and 333,968 from grade 3 made up the population for the study. Using a 95% confidence level, 45 schools were sampled at random for the study. The sample is presented in *Table 1*.

Table 1: Study sample size

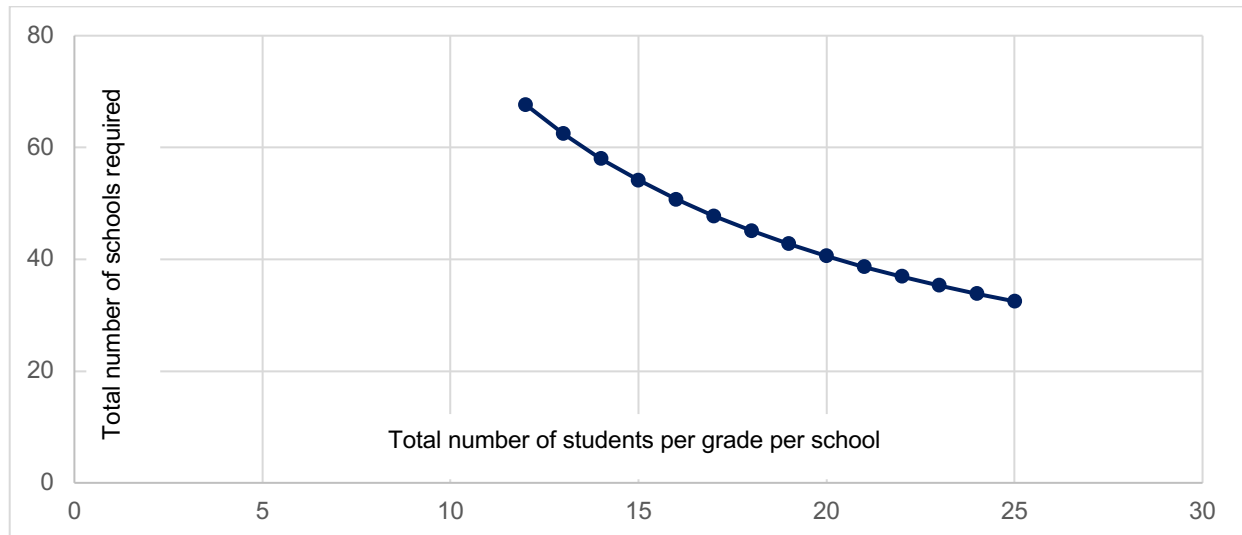
District	No. of sampled schools	No. of students assessed with CB-EGRA (baseline)						No. of students assessed with mini-EGRA (pilot)					
		Grade 2			Grade 3			Grade 2			Grade 3		
		Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
Achham	5	33	51	84	38	47	854	9	16	25	13	12	25
Bara	14	164	140	304	117	204	321	34	36	70	24	46	70
Bhojpur	6	23	40	63	31	34	65	9	20	29	9	19	28
Nawalparasi West	6	44	44	88	56	70	126	13	26	29	9	20	29
Rasuwa	3	23	27	49	24	27	51	7	12	19	9	10	19
Surkhet	7	53	41	101	60	46	106	16	19	35	19	16	35
Tanahun	4	33	36	67	32	40	70	8	11	19	8	13	21
Total	45	373	379	752	361	465	826	96	130	226	91	136	227
Grand total		1,578						453					

From the 45 schools, 752 students (boys: 49.6%; girls: 50.4%) from grade 2 and 826 students (boys: 43.7%, girls: 56.3%) from grade 3 were sampled for the EGRP II baseline. Of the grade 2 sample, 44.1% were learners with Nepali as L1 and 55.9% were learners with Nepali as L2. In grade 3, the sample consisted of 42.7% learners with Nepali as L1 and 57.3% with Nepali as L2. Overall, most of the sampled students (56.6%) had Nepali as their L2.

⁵ In Nepal's federal system of governance, palikas are the equivalent of municipalities. There are 753 palikas (both rural and urban) across 77 districts within 7 provinces in the country.

The sample size was determined based on the desire to maximize the precision of resulting estimates while limiting overall data collection costs. The optimal sample size was calculated using historical reading data from the 2020 NARN. *Figure 1* shows the relationship between the number of schools and number of students per school required to meet the maximum precision level.

Figure 1: Sample size required for the desired level of precision



In this assessment, EGRP II used the approach followed by ERO to conduct the CB-EGRA (ERO 2017), which targets 18 students on average as the number of students sampled from each school. Thus, by considering a confidence interval width of $\pm 3.5\%$, at a 95% confidence level, a standard deviation of 17.98 (taken from NARN 2020 data), and an intra-cluster correlation of 0.36, a design effect of 2.83 of was calculated. This led to determination of a total sample size of 812 students from each grade. Taking an average of 18 students per grade per school, we sampled 45 schools for the study. We considered different socio-cultural and geographical attributes when selecting the sample districts and municipalities. One district was selected from each province so that we could obtain a balance with regard to the language majority, level of EGRP II’s interventions, and topographical distribution, as presented in Figure 1. From each province, one palika was selected randomly and, to balance the sampling weight, we adjusted the number of schools to be sampled randomly from each palika. Initial student selection within each school, for administration of the CB-EGRA, was also random. While adjusting the number of sampled schools from each palika, we selected the sample number so that the ratio of sample weights among the cluster would not exceed 10. Because of student absenteeism on the day of assessment, we were able to administer the CB-EGRA to 752 students from grade 2 and 826 students from grade 3. These variations from the ideal sample size, however, did not limit the precision level of the overall study.

Moreover, we aimed to equate the CB-EGRA scores with the mini-EGRA scores to extrapolate oral reading fluency, which cannot be assessed directly with the CB-EGRA instrument. For this purpose, we subsampled five students from each grade in all of the sample schools, from among the larger sample of students who were assessed using the CB-EGRA. The purpose of the subsample was to create a 1:1 correspondence between the CB-

EGRA and reading fluency benchmarks. A subsample is not intended to be representative; rather, it is designed to sample scores relevant to the extrapolated ORF levels. As such, EGRP II used a purposive sample approach to collect EGRA scores near to our desired benchmarks. The “good” scores were in the fluent benchmark range, and the “average” scores were in the emergent benchmark range. Therefore, the subsample of students for EGRA administration was selected purposively, according to the classroom teacher’s estimation, to ensure that two students were high performing; the next two were average or in the middle of the class in terms of reading performance; and the remaining one was to be a low performer in the class.

2.3 Study Instruments

The CB-EGRA was used to collect students’ reading proficiency data for the baseline. The CB-EGRA is a curriculum-based tool that assesses children’s reading skills. Grade 2 and grade 3 children were assessed by their Nepali subject teachers who were trained to administer the CB-EGRA.

The CB-EGRA has a total of seven subtasks and each subtask includes three items, for a total of 21 items. For both grades 2 and 3, most subtasks entailed multiple-choice questions with five answer options (one correct answer and four distractors). However, the dictation subtasks for grades 2 and 3 and the word separation subtask for grade 3 were not multiple choice. For both grades, the classroom teacher followed a teacher’s guide while administering the assessment to students. While conducting the assessment, the teacher instructed the whole class at once on each of the subtasks. Two separate CB-EGRA assessment tools were used for grade 2 and grade 3. *Table 2* and *Table 3* provide the details of the tools that were used for each grade in the study.

Table 2: Description of grade 2 CB-EGRA assessment tool

No.	Subtask name	Items	Type	No. of distractors for each item	Example?	Time (minutes)		Subtask weight
						Example/ teacher instruction	Assessment	
1	Letter/matra identification	3	Multiple choice	5	Yes	2 min	3 min	1
2	Word and sentence identification	3	Multiple choice	5	Yes	2 min	3 min	2
3	Vocabulary	3	Multiple choice	5	Yes	2 min	3 min	3
4	Dictation	3	Writing	N/A	No	1 min	6 min	7
5	Listening comprehension	3	Multiple choice	5	No	4 min	4 min	4
6	Reading comprehension	3	Multiple choice	5	No	2 min	5 min	6
7	Calendar reading	3	Multiple choice	5	No	2 min	3 min	2

Note. N/A = not applicable.

Subtask 1: Letter/matra identification assesses students’ ability to identify the first letter or matra from the word that the teacher says.

Subtask 2: Word and sentence identification assesses students’ ability to identify the word or a sentence that the teacher reads aloud.

Subtask 3: Vocabulary assesses students’ vocabulary knowledge. Students are asked to state the definition, a synonym, and an antonym of each vocabulary word.

Subtask 4: Dictation assesses students’ writing skills. For this subtask, students have to write the entire sentence correctly as the teacher dictates. The teacher reads the sentence three times.

Subtask 5: Listening comprehension measures the number of comprehension questions that students answer correctly, based on a story of 25 words that the teacher reads aloud two times.

Subtask 6: Reading comprehension measures the number of comprehension questions that students answer correctly after they read a 60-word paragraph.

Subtask 7: Calendar reading measures students’ ability to comprehend a calendar, which can be considered a visual literacy skill (ability to view and comprehend multimodal texts).

Table 3: Description of grade 3 CB-EGRA assessment tool

No.	Subtask name	No. of Items	Type	No. of distractors for each item	Example?	Time (minutes)		Subtask weight
						Example/ teacher instruction	Assessment	
1	Word and sentence identification	3	Multiple choice	5	Yes	2 min	3 min	1
2	Vocabulary	3	Multiple choice	5	Yes	2 min	3 min	2
3	Word separation	3	Multiple choice	N/A	Yes	2 min	5 min	5
4	Dictation	3	Multiple choice	N/A	No	1 min	6 min	6
5	Listening comprehension	3	Multiple choice	5	No	4 min	4 min	4
6	Reading comprehension	3	Multiple choice	5	No	2 min	5 min	5
7	Calendar reading	3	Multiple choice	5	No	2 min	3 min	2

Note. N/A = not applicable.

Subtask 1: Word and sentence identification assesses students’ ability to identify the word or a sentence that the teacher reads aloud.

Subtask 2: Vocabulary assesses students’ vocabulary knowledge. Students are asked to state the definition, a synonym, and an antonym for each vocabulary word.

Subtask 3: Word separation assesses the children’s ability to decode words. It measures how well children can separate the words in a sentence when all the words are joined together.

Subtask 4: Dictation assesses students’ writing skills. For this subtask, students have to write the entire sentence correctly as the teacher dictates. The teacher reads the sentence three times.

Subtask 5: Listening comprehension measures the number of comprehension questions the students answer correctly, based on a story of 30 words that the teacher reads aloud two times.

Subtask 6: Reading comprehension measures the number comprehension questions that students answer correctly after reading a 60-word passage.

Subtask 7: Calendar reading measures students’ ability to comprehend the calendar, which can be considered a visual literacy skill (ability to view and comprehend multimodal texts).

In addition to CB-EGRA, a mini-EGRA, which was administered to a subsample of students as discussed above, consisted of a test of ORF, where students read a 60-word passage out loud and then answered five comprehension questions based on the passage. The number of words the students were able to read correctly per minute (the ORF rate) and the number of questions answered correctly comprised the data collected using the mini-EGRA.

2.4 Study Quality Assurance

Quality assurance was prioritized throughout the entire process of the study. In the first phase, the Kathmandu-based EGRP II monitoring, evaluation, and learning (MEL) team, along with ERO technical personnel, provided a training of trainers to EGRP II technical leads and regional MEL coordinators. This 2-day training focused on the theoretical and practical aspects of the CB-EGRA and EGRA and the logistics that would be required while the trainees were collecting the data. Using the KoBo Toolbox platform, the mini-EGRA tools were digitized and rendered on tablets. The MEL team also developed a monitoring platform and digitized it using KoBo Toolbox. Using Microsoft Power Query, the team extracted KoBo Toolbox data to Excel for real-time visualization and monitoring.

The EGRP II MEL coordinators, along with the Kathmandu-based team members, subsequently rolled out the CB-EGRA training to teachers from the sampled schools who would administer the CB-EGRA, while the EGRP II district coordinators were trained on mini-EGRA administration and quality monitoring. After the training, the teachers administered a CB-EGRA in the presence of EGRP II staff to ensure the quality and reliability of the administration. Through the tools mentioned above, the team ensured that there was real-time reporting on progress and advised on any challenges that arose during the assessment.

Figure 2 shows screen shots of the assessment monitoring system along with the real-time data visualization system.

Figure 2: Screen shots of data collection, real-time visualization, and monitoring systems

Baseline 2021 Status

▼ **Baseline Status Dashboard**

▼ **» Assessor Information**

NAME OF ASSESSOR	* DATE OF ASSESSMENT *
Sagar Neupane	2021-04-02

▼ **» विद्यालय र विद्यार्थी सम्बन्धी विवरण**

S1: EMIS CODE	* S2: SCHOOL NAME :	S3: PROVINCE :	
S4: DISTRICT :	S5: PALIKA :	S6: WARD NUMBER :	S7: TOLE/GAUN :
TOTAL NUMBER OF GRADE 2 CHILDREN ADMITTED : 300	TOTAL NUMBER OF GRADE 3 CHILDREN ADMITTED : 250	TOTAL NUMBER OF GRADE 2 CHILDREN PRESENT TODAY : 80	TOTAL NUMBER OF GRADE 3 CHILDREN PRESENT TODAY : 70

▼ **» Status of Baseline**

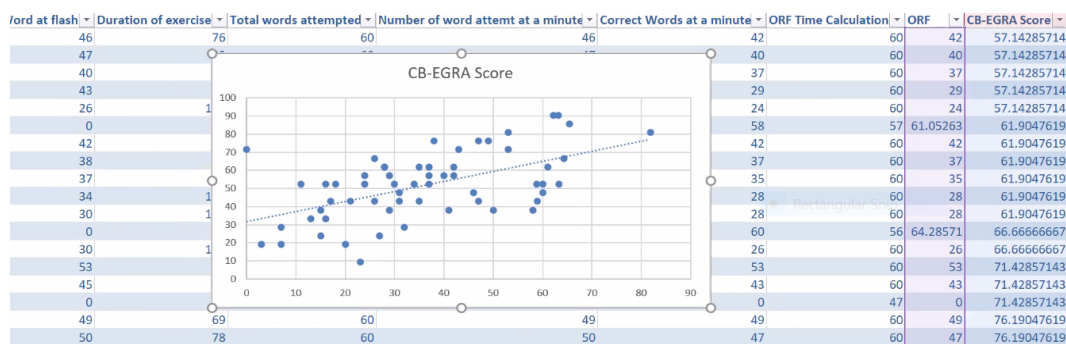
TOTAL NUMBER OF CHILDREN PARTICIPANTS ON CB-EGRA *	
<input checked="" type="radio"/> Yes <input type="radio"/> No	
TOTAL NUMBER OF GRADE 2 STUDENT *	TOTAL NUMBER OF GRADE 3 STUDENT *
20	20
TOTAL NUMBER OF CHILDREN PARTICIPANTS ON MINI-EGRA *	
<input type="radio"/> Yes <input checked="" type="radio"/> No	

Early Grade Reading Program II

Status of Baseline 2021

Enter Data

District	# Sampled School	# Completed School	# Remaining School	Total Students Admitted		Total Students present		CB-EGRA		Mini-EGRA	
				Grade 2	Grade 3	Grade 2	Grade 3	Grade 2	Grade 3	Grade 2	Grade 3
Achham	5	5	0	120	113	80	84	80	84	25	25
Bara	14	14	0	734	686	360	354	319	305	70	70
Bhojpur	6	6	0	67	74	54	69	54	69	30	28
Nawalparasi West	6	6	0	148	179	88	125	88	125	28	30
Rasuwa	3	3	0	51	54	49	51	49	51	21	20
Surkhet	7	7	0	137	146	101	113	101	100	35	35
Tanahun	4	4	0	79	81	67	70	67	70	20	21
Total	45	45	0	1336	1333	799	866	758	804	229	229



This monitoring system allowed the EGRP II MEL team to monitor the progress of the assessment. In addition, it enabled the team to provide case-by-case support when required. The data visualization, shown in the last image of Figure 2, helped the team to understand the distribution of CB-EGRA scores along with EGRA ORF scores in real-time. Midway through data collection, the team realized that the count of zero scores in the EGRA was higher than expected, which could undermine the ability to analyze the relationship between CB-EGRA and EGRA performance. To overcome this challenge, we requested our data collectors to stop purposefully selecting low performers for the mini-EGRA assessment. This enabled EGRP II to obtain a more balanced set of data on learner performance.

3 Study Findings

This section presents the findings from the study. For the CB-EGRA data for both grades 2 and 3, we calculated sample weights based on the number of provinces, districts, and palikas; number of schools in each palika; number of students sampled from each school against the total enrollment; and total number of students present on the day of the assessment. The average percentage scores were calculated based on the sample weights and subtask weights.⁶ We used IBM SPSS Statistics version 21 to analyze the data using the Complex Sample module. Using this approach to sample weighting affords confidence that the baseline results represent the estimated population.

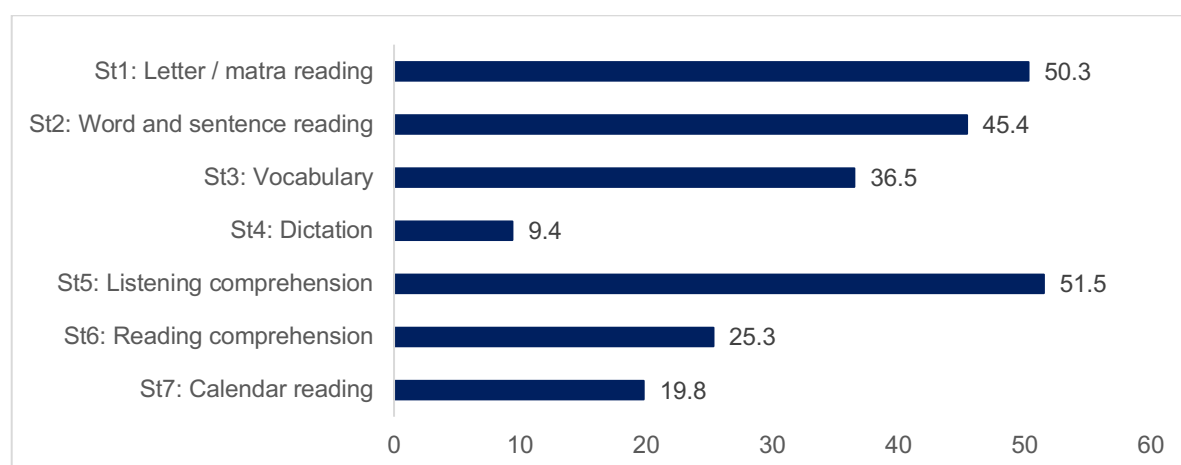
3.1 Grade 2 Findings

3.1.1 Overall Reading Achievement (Grade 2)

Of the subtasks assessed in grade 2, students performed best on listening comprehension, with an average score of 51.5%. This score indicates that students were, on average, able to respond correctly to about half of the questions from the listening comprehension subtask. In contrast, students had the most difficulty with the dictation subtask. The average percentage score on dictation, 9.4%, signifies that those students made errors in over 90% of the items in this subtask. Similarly, in general, students struggled with the calendar-reading subtask. The average percentage score for this subtask was 19.8%, which means that, on average, children were able to answer only one-fifth of the items correctly.

A breakdown of the average scores for grade 2 students for each subtask is presented in *Figure 3*.

Figure 3: Average percentage scores, by subtask (grade 2)



On average, grade 2 students were able to answer 6 out of 21 questions correctly in the overall assessment. The distribution of average percentage scores is presented in *Figure 4*. In

⁶ The ERO subject committee, in consultation with subject experts from Nepali universities, allocated different weights to the subtasks as presented in Table 2 and Table 3, based upon the difficulty level. The main purpose of the weighting was to calculate overall reading achievement by using weights for all of the subtasks.

this figure, the overall average percentage score is categorized into five different groups: 0, 1%–20%, 21%–40%, 41%–60%, 61%–80% and 81%–100%.

Figure 4: Distribution of overall average percentage scores (grade 2)

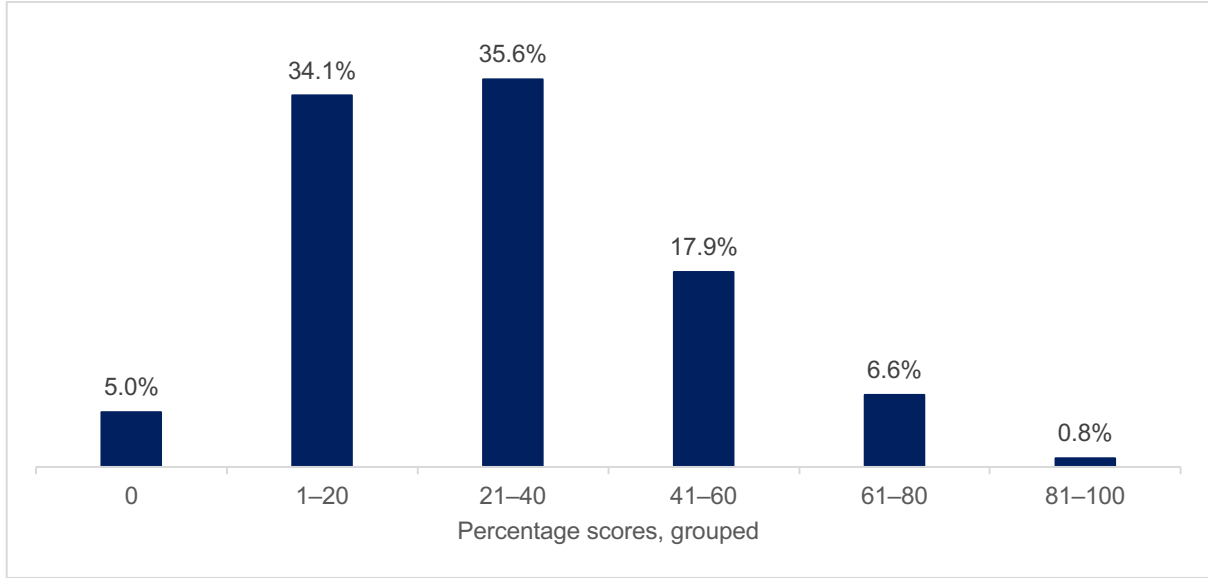


Figure 4 shows that about two-thirds of students achieved a score of 40% or less. Very few students (0.8%) could answer 80% or more questions correctly.

3.1.2 Reading Achievement by Subtask (Grade 2)

The following analysis provides details about the average grade 2 percentage scores for the different subtasks.

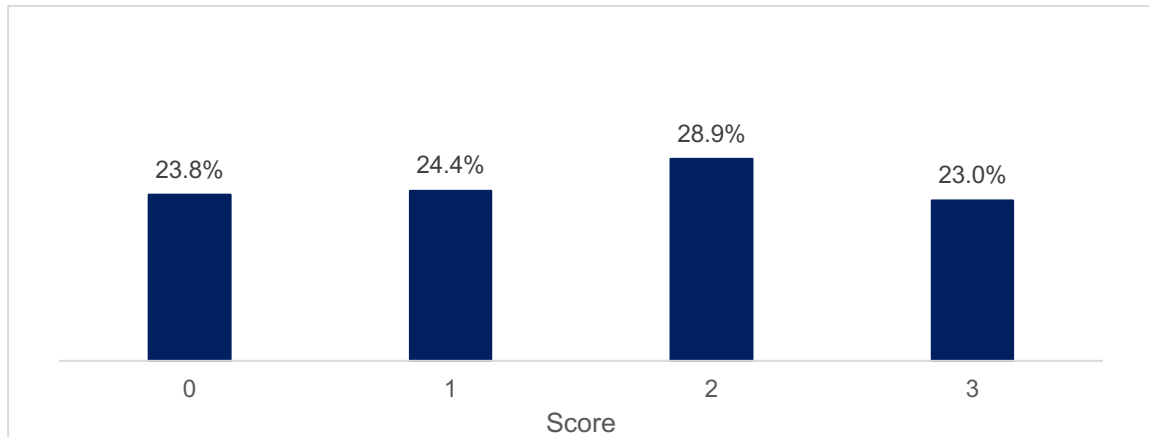
Subtask 1 was to identify the first letter/matra from the word that was said by the teacher, repeated two times. The subtask was intended to assess the students' ability to recognize the first letter/matra in a word. The items in the subtask were multiple choice. There were five possible answers in each item, including one correct option and four distractors. **Figure 5** is a screen shot of the student stimulus for the grade 2 letter/matra identification subtask.

Figure 5: Student stimulus for the grade 2 letter/matra identification subtask

उपकार्य १					
वर्ण र मात्रा पहिचान					
(क)	र	ख	ह	स	म
(ख)	ली	प	पु	फु	त
(ग)	इया	स्या	फ्या	ज्या	प्या

About half of the students correctly responded to one question or fewer in this subtask. Less than one-fourth (23.0%) of students were able to correctly respond to all the questions. The score distribution for the subtask is presented in *Figure 6*.

Figure 6: Distribution of scores for the grade 2 letter/matra identification subtask



The overall average score on letter/matra identification was 50.3%, which means that out of three items, students were able to answer less than two questions on average. The first item focused on identifying a simple vowel or consonant letter. As indicated in *Table 4*, about 60% of the students were able to answer this item correctly and the same proportion of students were able to answer the second question, which was to identify a simple matra. Less than one-third of the students were able to answer the third question, which was identification of mixed letters (a half letter and a matra combined). This result suggests that students in the study may not yet have mastered mixed letters in grade 2.

Table 4: Average item scores for the grade 2 letter/matra identification subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
1a	Identify vowel or consonant letter	58.6%	4.1%
1b	Identify simple matra	60.3%	5.1%
1c	Identify mixed letter/matra	32.1%	5.2%

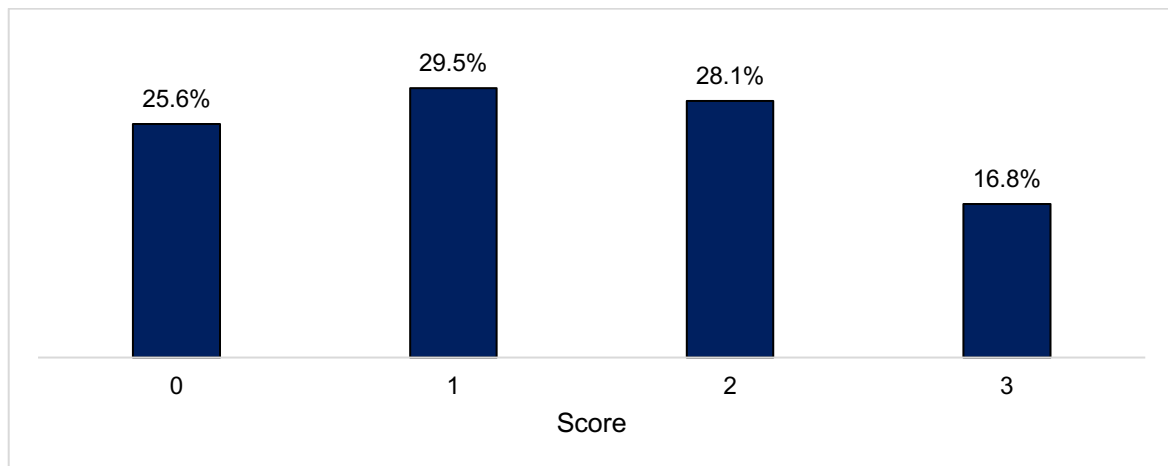
In **Subtask 2**, students had to identify the word or short sentence that the teacher said, repeating two times. Among the three items in the subtask, the first item was to identify a word and the second and third items were to identify sentences of three and four words, respectively. The items in the subtask were multiple choice. There were five possible responses for each item, with one correct option and four distractors. *Figure 7* provides a screen shot of the student stimulus for the grade 2 word and sentence identification subtask.

Figure 7: Student stimulus for the grade 2 word and sentence identification subtask

उपकार्य २					
शब्द र वाक्य पहिचान					
(क)	किल्ला	डल्ला	बिल्ला	चल्ला	जिल्ला
(ख)	जात्रामा भिड थियो ।	जात्रामा रमिता थियो ।	जात्रामा पिड थियो ।	जात्रामा नाच थियो	जात्रामा हल्ला थियो ।
(ग)	फ्याउराको कुराले चिल रिसायो ।	फ्याउराको कुराले काग डरायो ।	फ्याउराको कुराले काग रिसायो ।	फ्याउराको कुराले परेवा रिसायो ।	फ्याउराको कुराले सिंह करायो ।

The distribution of scores for this subtask (*Figure 8*) indicates that only 16.8% of the students were able to solve all the questions asked, while just over one-quarter of students could not answer a single question.

Figure 8: Distribution of scores for the grade 2 word and sentence identification subtask



The overall average score on the subtask was 45.4%. This result indicates that out of three questions, students were able to solve fewer than two items correctly on average. Item-level score disaggregation (*Table 5*) indicates that the first item was comparatively easy for the students as it was correctly solved by 62.4%. At the same time, the students found the third question—on identification of a four-word sentence—comparatively harder. Only one-fourth (25.7%) of students were able to solve the question correctly.

Table 5: Average item scores for the grade 2 word and sentence identification subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
2a	Identify one word	62.4%	4.1%
2b	Identify three-word sentence	48.0%	3.4%
2c	Identify four-word sentence	25.7%	2.9%

Subtask 3 assessed student vocabulary. The first item focused on defining a word, whereas the second and third items focused on knowledge of antonyms and synonyms. The items in the subtask were multiple choice. There were five possible responses for each item, with one correct option and four distractors. A screen shot of the student stimulus is presented in *Figure 9*.

Figure 9: Student stimulus for the grade 2 vocabulary subtask

उपकार्य ३
शब्दभण्डार

(क) बाघ बस्ने ठाउँलाई भनिन्छ ।

गुँड	कुर	गुफा	दुलो	गोठ
------	-----	------	------	-----

(ख) थोरैको उल्टो अर्थ दिने शब्द हो ।

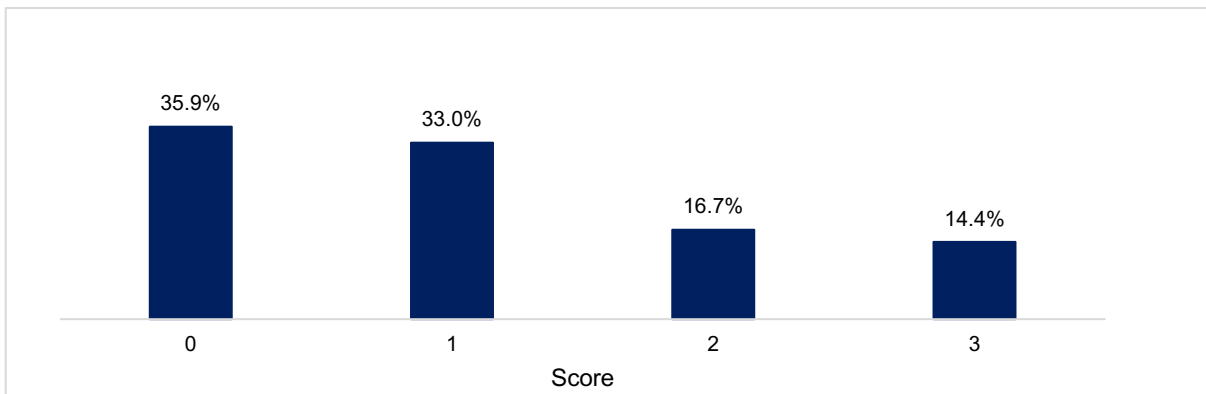
धेरै	अलिकति	कस्ति	अपुग	केही
------	--------	-------	------	------

(ग) तागत शब्दको अर्थ हो ।

कमजोर	बल	निर्धो	नाजुक	दुर्बल
-------	----	--------	-------	--------

The score distribution for the grade 2 vocabulary subtask (*Figure 10*) shows that less than one-third (31.1%) of the students were able to solve two or more questions. More than one-third (35.9%) of the students could not solve a single question in this subtask.

Figure 10: Distribution of scores for the grade 2 vocabulary subtask



The overall average score for the grade 2 vocabulary subtask was 36.5%. The result indicates that out of three questions, students were able to solve only one question correctly on average. Disaggregation of average scores by item (*Table 6*) indicates that half of the students (50.6%) were able to respond correctly to the item related to word definition whereas students struggled far more to correctly answer the item related to antonyms. Only one-fourth of the students were able to answer the question on antonyms, whereas one-third of the students responded correctly to the question on synonyms.

Table 6: Average item scores for the grade 2 vocabulary subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
3a	Define a word	50.6%	4.7%
3b	Antonyms	25.7%	4.3%
3c	Synonyms	33.2%	5.1%

Subtask 4 assessed writing skills and was a dictation task. In this subtask, students were asked to write sentences correctly as the teacher said them, repeating each item three times. The first item in the subtask was to write a three-word sentence, whereas the second was a four-word sentence. The third was also a four-word sentence with words that were more difficult. A screen shot of the student stimulus is presented in *Figure 11*.

Figure 11: Student stimulus for the grade 2 dictation subtask

उपकार्य ४
श्रुतिलेखन

(क) खोलामा माछा छन् ।
(ख) हामी विद्यालय जान्छौं ।
(ग) वनमा बाघले बाख्रालाई तर्सियो ।

(क) |
(ख) |
(ग) |

The average percentage score for this subtask was 9.4%. This result indicates that, on average, the number of items students were able complete without errors was less than one. As noted in **Figure 12**, only a small number of students (1.5%) were able to complete all three items without any errors.

Figure 12: Distribution of scores for the grade 2 dictation subtask

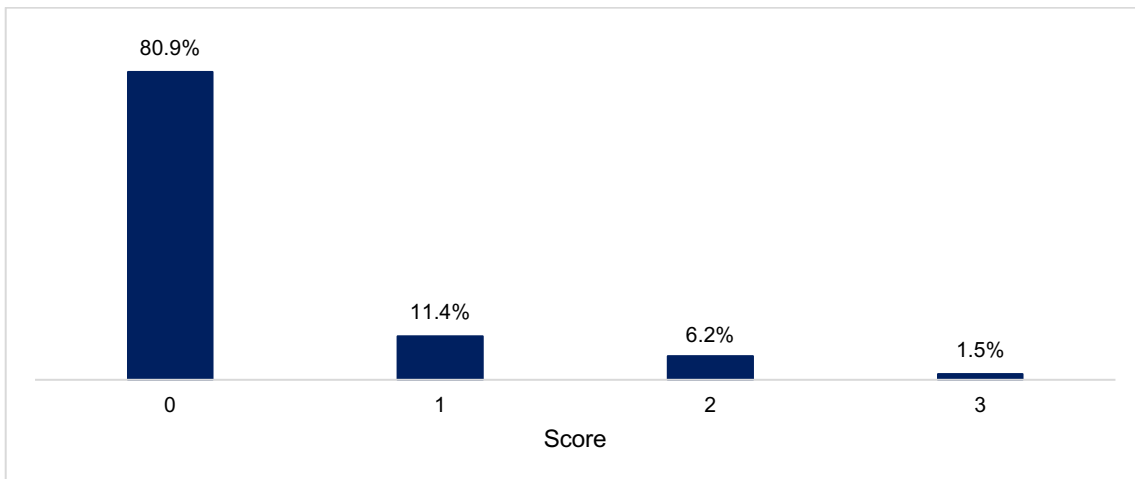


Table 7 displays the average percentage scores of the students on different items in the subtask. For the first question, related to dictating a three-word sentence, the average score was 18.1%; on the second question, the average was 8.3%. For the third question—dictation of four-word sentences with a higher difficulty of words—the average score was only 1.9%.

Table 7: Average item scores for the grade 2 dictation subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
4a	Three-word sentence	18.1%	4.6%
4b	Four-word sentence	8.3%	2.7%
4c	Four-word sentence, difficult words	1.9%	0.7%

Subtask 5 assessed the listening comprehension ability of students. The teacher read a 25-word passage and asked three questions about it. The first question was in short-answer format and could be answered based on information provided explicitly in the first or second sentence of the paragraph. The second question’s answer was also found directly in the text. The third was an inferential question where students had to build answers from information in at least two sentences in the text. The items in the subtask were multiple choice, with five answer options, including one correct option and four distractors. The student stimulus is presented in *Figure 13*.

Figure 13: Student stimulus for the grade 2 listening comprehension subtask

उपकार्य ५
श्रुतिबोध

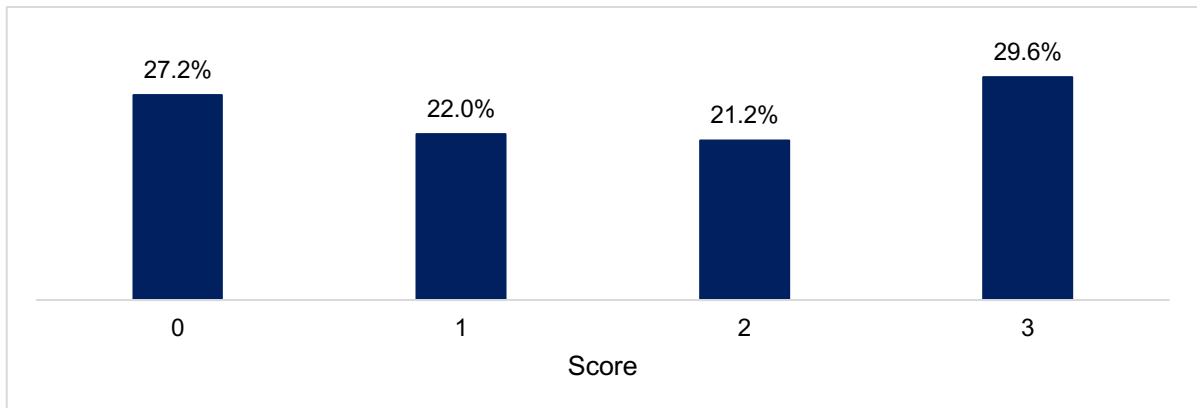
कमल विद्यालयबाट फर्कदै थिए । अचानक पानी पर्न थाल्यो । कमलसँग छाता थिएन । उनले बारीमा कर्कलाको बोट देखे । उनले कर्कलाको पात टिपेर ओढे । उनी भिजेबाट सजिलै जोगिए ।

- क. कमल कहाँबाट फर्कदै थिए ?
ख. कमलले केको बोट देखे ?
ग. कमल रुइनबाट कसरी जोगिए ?

(क)	घरबाट	विद्यालयबाट	पसलबाट	सहरबाट	जङ्गलबाट
(ख)	सालको	पिपलको	चिउरीको	कर्कलाको	हलेदोको
(ग)	छाता ओढेर	कर्कलाको पात ओढेर	घुम ओढेर	ओत लागेर	प्लास्टिक ओढेर

Analysis of each item in the subtask (*Figure 14*) identified that 29.6% of the students were able to solve all three questions from this subtask, while more than one quarter could not answer any of the questions.

Figure 14: Distribution of scores for the grade 2 listening comprehension subtask (25 words)



The overall average score for Subtask 5 was 51.1%. Looking at the disaggregated results for each item in the subtask (*Table 8*), students were found nearly equally competent to answer the questions that came directly from the text (54.2% and 49.8% of students were able to answer the first and second items, respectively) or inferential type of questions (the third

item, with 49.2% of students able to answer correctly). Thus, half of the students were able to answer each item from this subtask correctly.

Table 8: Average item scores for the grade 2 listening comprehension subtask (25 words)

Subtask	Description	Percentage of students who answered correctly	Standard error
5a	Short answer, explicit	54.2%	4.8%
5b	Short answer, explicit	49.8%	4.5%
5c	Inferential from at least two sentences	49.2%	3.8%

Subtask 6 assessed reading comprehension ability. Students had to read a passage of 60 words and answer three questions based on the text. The first and second questions could be answered directly by referring to the text, and the third question was inferential and demanded that the student consider information from two or more sentences from the text. The items in the subtask were multiple choice. There were five answer options, with one correct option and four distractors. The student stimulus is presented in *Figure 15*.

Figure 15: Student stimulus for the grade 2 reading comprehension subtask

उपकार्य ६

पठनबोध

किसानको बारी थियो । बारीमा नासपातीका बिरुवा थिए । केही सुन्तलाका बिरुवा पनि थिए । उनी बिरुवाको हेरविचार गर्थे । एक दिन गोरुले फलफूलका बिरुवा खाइदियो । किसान निकै दुःखी भए । उनले बिरुवा जोगाउने उपाय सोचे । उनले बारीमा बार लगाए । बार नाघेर गाईगोरु भित्र जान सकेनन् । उनले बारीमा नयाँ बिरुवा पनि थपे । किसानका फलफूलका बिरुवा जोगिए । बिरुवा हुकिँदै गए । यो देखेर किसान धेरै खुसी भए ।

(क) किसानको बारीमा के केका बिरुवा थिए ?

नासपाती र मेवा	सुन्तला र कागती	नासपाती र सुन्तला	सुन्तला र आरु	नासपाती र आरु
----------------	-----------------	-------------------	---------------	---------------

(ख) किसान किन दुःखी भए ?

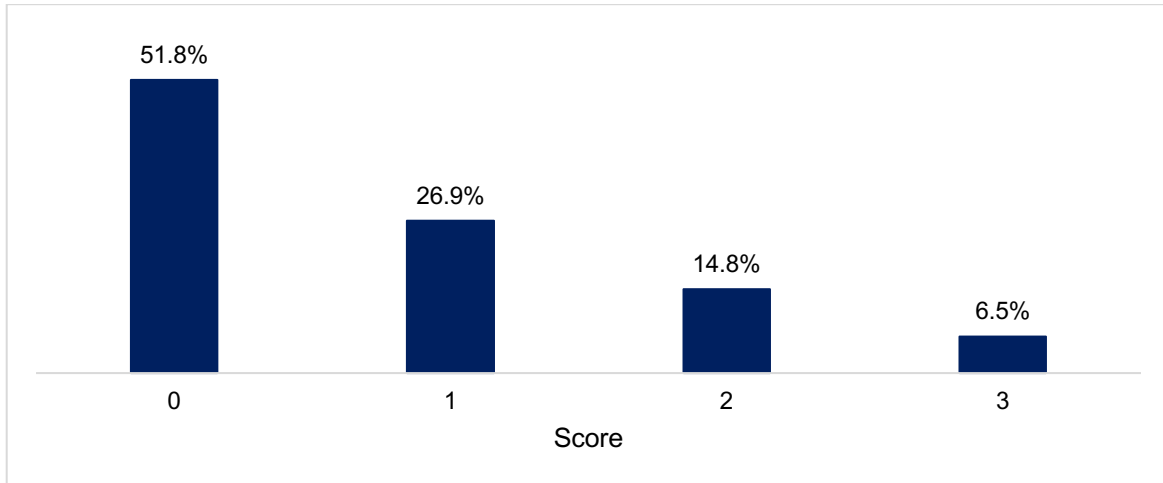
गोरुले फलफूलका बिरुवा खाएकाले	गोरुले घाँसका बिरुवा खाएकाले	गोरुले तोरीका बिरुवा खाएकाले	गोरुले धानका बिरुवा खाएकाले	गोरुले सागका बिरुवा खाएकाले
-------------------------------	------------------------------	------------------------------	-----------------------------	-----------------------------

(ग) किसानका बिरुवाहरू कसरी जोगिए ?

पाले बसेर	बार लगाएर	गोरु बाँधेर	पर्खाल लगाएर	कुकुर पालेर
-----------	-----------	-------------	--------------	-------------

The distribution of scores on each item in the subtask (*Figure 16*) shows that more than half of the students could not solve a single question in this subtask; at the other end of the distribution, the percentage of students who solved all three questions was low at 6.5%.

Figure 16: Distribution of scores for the grade 2 reading comprehension subtask (60 words)



The average percentage score for this subtask was 25.3%. This result indicates that the average correct answer per student was less than one question out of three. In general, students found the first item in the subtask slightly more difficult than the second and third items (*Table 9*).

Table 9: Average item scores for the grade 2 reading comprehension subtask (60 words)

Subtask	Description	Percentage of students who answered correctly	Standard error
6a	Short answer, explicit	20.4%	2.4%
6b	Short answer, explicit	29.1%	3.4%
6c	Inferential from two or more sentences	26.5%	2.9%

Subtask 7 was related to calendar reading. Being able to view and make sense of a calendar is considered part of visual literacy, which is the ability to view and understand multimodal texts. In this subtask, a month from the Nepali calendar was provided and three questions based on the calendar shown were asked. The first question required identifying the day and date, while the second question involved understanding the relationship between festival and date. The third question was to count the total number of a certain type of day (e.g., Saturday) in the month. The overall average percentage score on the subtask was 25.3%. This result shows that students were able to answer fewer than one question out of three correctly in this subtask. The items in the subtask were multiple choice. There were four distractors in each item in addition to one correct option. *Figure 17* shows the student stimulus.

Figure 17: Student stimulus for the grade 2 calendar reading subtask

उपकार्य ७

पात्रो (क्यालेन्डर) पठन

माघ २०७५			Jan/Feb 2019			
Sunday आइतबार	Monday सोमबार	Tuesday मंगलबार	Wednesday बुधबार	Thursday बिहीबार	Friday शुक्रबार	Saturday शनिबार
		नवमी १ 15 माघे संक्रान्ति, माघी पर्व (सम्बन्धितलाइ बिदा)	दशमी २ 16	एकादशी ३ 17 पुत्रदा एकादशी	द्वादशी ४ 18	त्रयोदशी ५ 19
चतुर्दशी ६ 20	पूर्णिमा ७ 21 श्री स्वस्थानी व्रत आरम्भ	प्रतिपदा ८ 22	तृतीया ९ 23	चतुर्थी १० 24	पञ्चमी ११ 25	षष्ठी १२ 26
सप्तमी १३ 27	अष्टमी १४ 28	नवमी १५ 29	दशमी १६ 30 सहिद दिवस	एकादशी १७ 31 षटतिला एकादशी	द्वादशी १८ 01	त्रयोदशी १९ 02
चतुर्दशी २० 03	औशी २१ 04	प्रतिपदा २२ 05 सोनास लोसार सम्बन्धितलाइ बिदा	द्वितीया २३ 06	तृतीया २४ 07	तृतीया २५ 08	चतुर्थी २६ 09
पञ्चमी २७ 10 सरस्वती पूजा (शिक्षणसंस्था बिदा)	षष्ठी २८ 11	सप्तमी २९ 12				

(क) यो महिनाको १० गते कुन बार परेको छ ?

आइतबार	सोमबार	मङ्गलबार	बुधबार	बिहीबार
--------	--------	----------	--------	---------

(ख) सरस्वती पूजा कति गते परेको छ ?

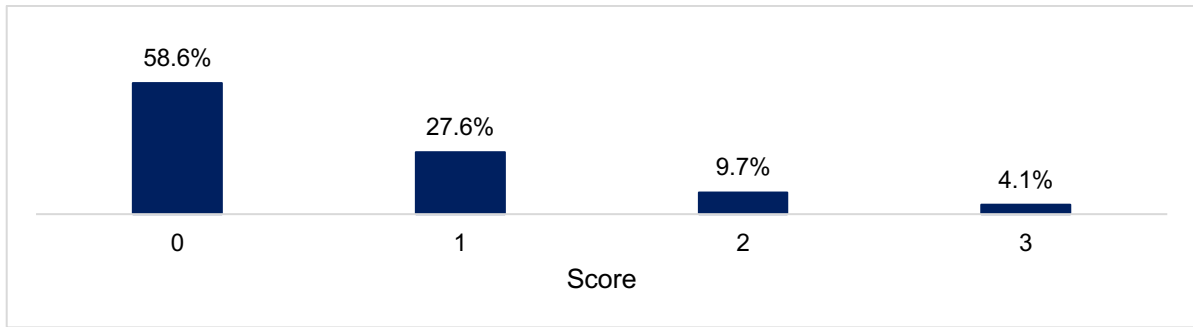
१५	१७	२७	२५	२८
----	----	----	----	----

(ग) यो महिनामा कति ओटा शनिबार छन् ?

३	४	५	६	७
---	---	---	---	---

Figure 18 shows that 58.6% of the students were not able to answer a single question from this subtask. Very few students (13.8%) were able to answer two or more questions from the subtask. These findings indicate that students struggled with viewing and understanding the calendar, which could signify difficulty with visual literacy, although we cannot be sure that this result would translate to all types of multimodal texts.

Figure 18: Distribution of scores for the grade 2 calendar reading subtask



As shown in *Table 10*, the difficulty level of all three questions in this subtask was nearly the same. The first question was answered correctly by 17.6% of students, while 19.6% and 22.2% of students were able to answer the second and third questions correctly, respectively.

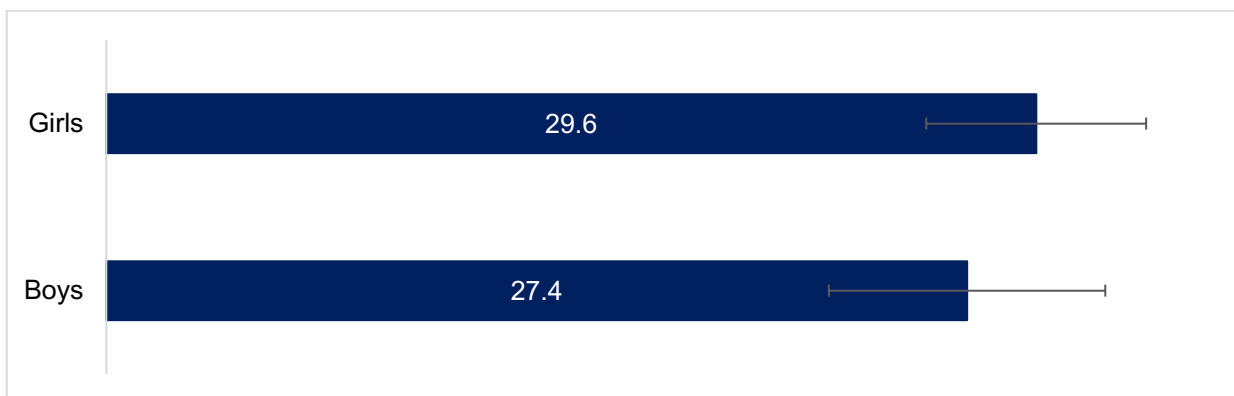
Table 10: Average item scores for the grade 2 calendar reading subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
7a	Day and date	17.6%	2.7%
7b	Festival and date	19.6%	2.7%
7c	Number of specific days (e.g., Saturdays) in a month	22.2%	3.5%

3.1.3 Reading Achievement by Sex and LI (Grade 2)

Student reading achievement was disaggregated by the sex of students to understand whether scores varied between boys and girls, as shown in *Figure 19*. The findings show that the average percentage score of girls was slightly more than that of boys. The difference, however, was not statistically significant. (The 95% confidence intervals are represented by thin black lines at the end of each blue bar.)

Figure 19: Average percentage scores of grade 2 students, by sex



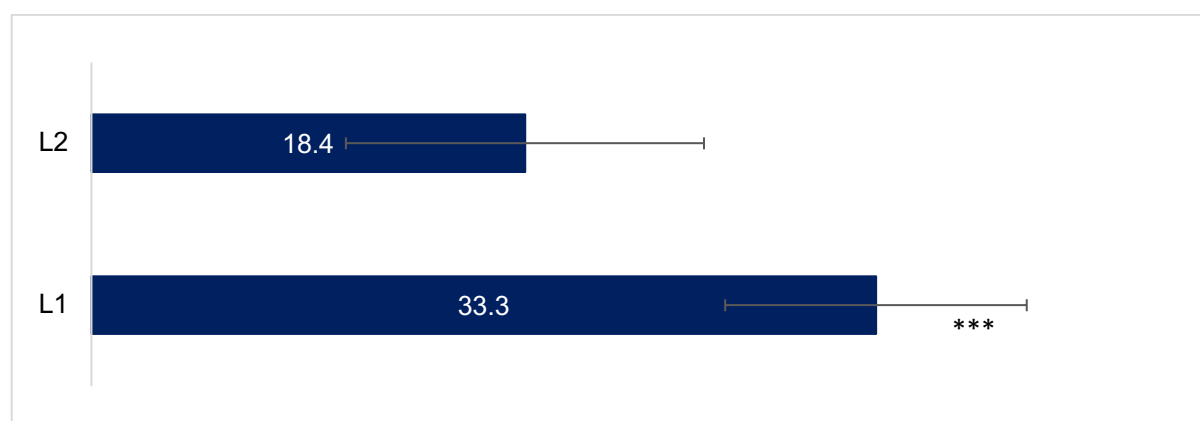
Similar results held for all subtasks: the achievement difference was not significantly different in any case (*Table 11*).

Table 11: Average percentage scores of grade 2 students, by subtask and sex

Subtask	Average percent score (boys)	Average percent score (girls)	Difference (boys – girls)
St1: Letter/matra reading	48.5	52.1	-3.7
St2: Word and sentence identification	43.7	46.9	-3.3
St3: Vocabulary	35.2	37.8	-2.7
St4: Dictation	8.8	10.0	-1.2
St5: Listening comprehension	48.0	53.9	-5.9
St6: Reading comprehension	25.0	25.7	-0.7
St7: Calendar reading	19.4	20.2	-0.7

L1, on the other hand, was found to be a significant factor in students’ reading achievement. As shown in **Figure 20**, the average score for Nepali L1 students was roughly 7 out of 21 correct, whereas it was only approximately 4 out of 21 correct among Nepali L2 students. The overall difference by language, of about 15 percentage points, was statistically significant, as were the differences for all the subtasks except dictation and calendar reading.

Figure 20: Average percentage scores of grade 2 students, by language



*** $p < .01$.

Table 12 shows the students’ reading achievement disaggregated by language for each subtask.

Table 12: Average percentage scores of grade 2 students, by language

Subtask	Average percent score (L1)	Average percent score (L2)	Difference (L1 – L2)
St1: Letter/matra identification	59.0	31.9	27.2***
St2: Word/sentence identification	52.8	29.5	23.3***
St3: Vocabulary	43.5	21.8	21.7**
St4: Dictation	11.2	5.8	5.4
St5: Listening comprehension	59.9	32.2	27.7***
St6: Reading comprehension	29.3	17.0	12.2***
St7: Calendar reading	21.4	16.4	5.1

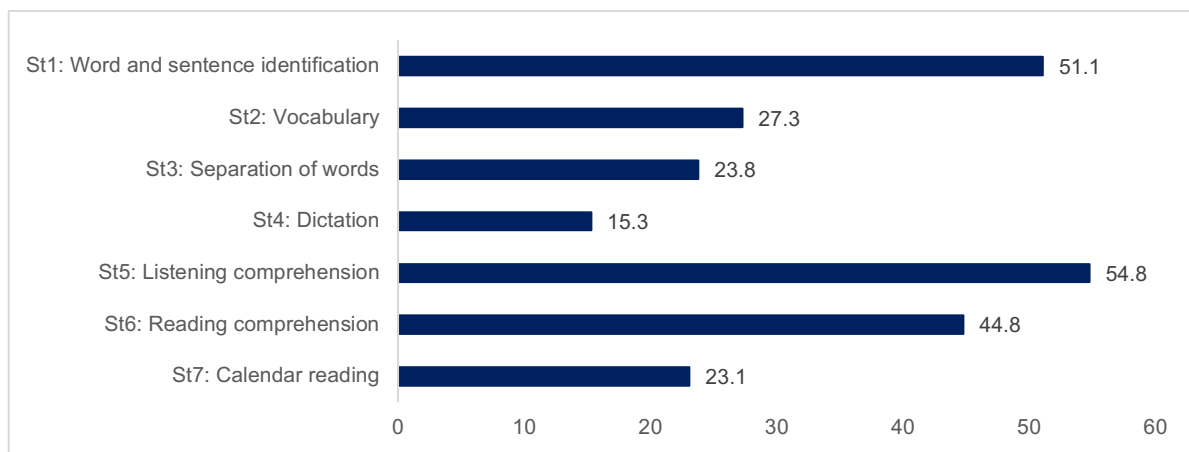
*** $p < .01$, ** $p < .05$.

3.2 Grade 3 Findings

3.2.1 Overall Reading Achievement (Grade 3)

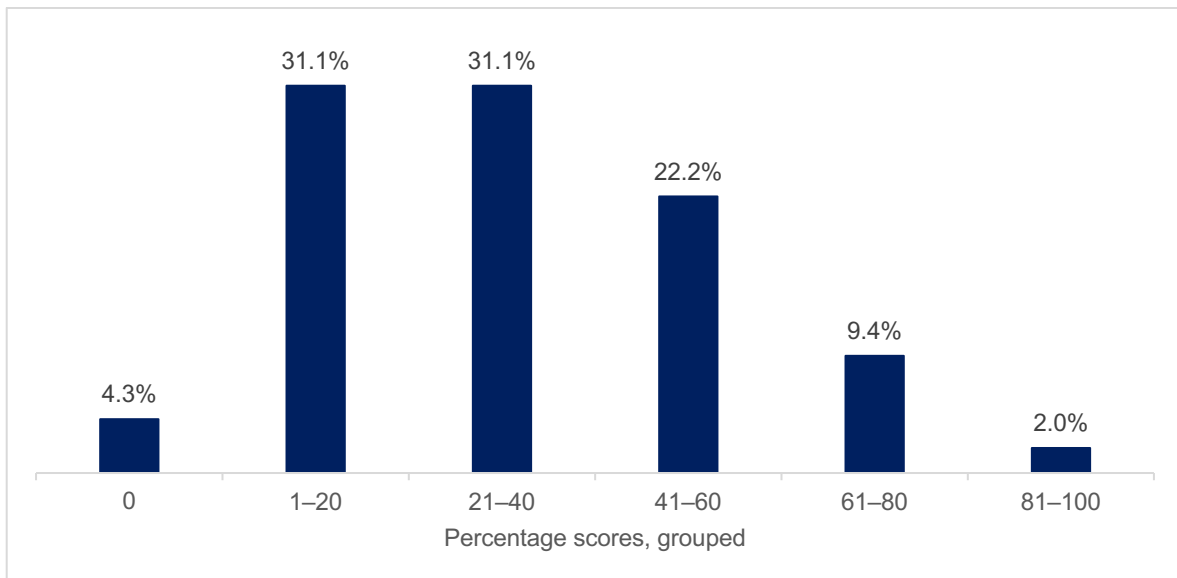
Similar to grade 2, students in grade 3 performed highest on listening comprehension, with an average score of 54.8%, indicating that students were able to solve about half of the questions from the listening comprehension subtask, on average. However, the writing skills that were assessed from the dictation and word separation subtasks were most difficult for the students. The average score for the dictation subtask was 15.3% and for word separation was 23.8%. Similarly, students struggled with calendar reading. The average score for this subtask was 23.1%, which indicates that the students answered only one-fifth of the questions correctly, on average. Details on the average scores of grade 3 students on each subtask are presented in *Figure 21*.

Figure 21: Average percentage scores, by subtask (grade 3)



On average, grade 3 students were able to answer 7 out of 21 questions correctly in the overall assessment. The distribution of the average scores is presented in *Figure 22*, with the percentage scores categorized into five groups. This figure shows that about two-thirds of the total students achieved 40% or less. Very few students (2.0%) were able to answer 81% or more questions correctly.

Figure 22: Distribution of overall average percentage scores (grade 3)



3.2.2 Reading Achievement by Subtask (Grade 3)

This section shares analysis of average percentage scores of grade 3 students for the different subtasks.

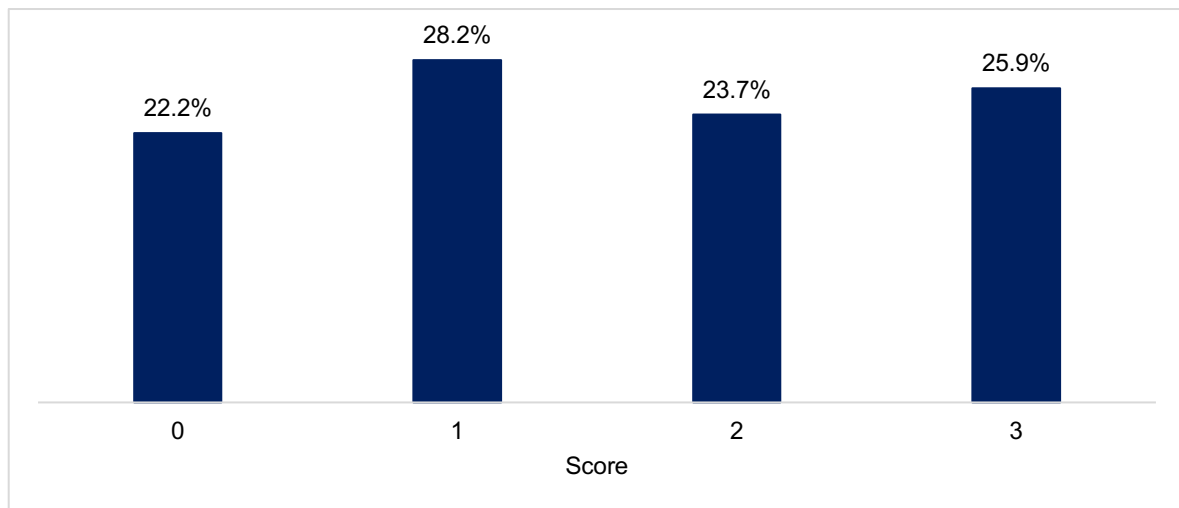
Subtask 1 was to identify the word or short sentence that the teacher said, repeating two times. Among the three items in the subtask, the first item was to identify a word and the second and third items were to identify sentences of four and five words, respectively. The items in the subtask were multiple choice. There were five possible responses for each item, with one correct option and four distractors. The student stimulus appears in *Figure 23*.

Figure 23: Student stimulus for the grade 3 word and sentence identification subtask

उपकार्य १					
शब्द र वाक्य पहिचान					
(क)	राम्रो	न्यास्रो	च्यात्यो	त्यान्द्रो	काँक्रो
(ख)	कौवाले गुँडबाट टाढासम्म देख्यो ।	कौवाले धुरीबाट टाढासम्म देख्यो ।	कौवाले रुखबाट टाढासम्म देख्यो ।	कौवाले आँगनबाट टाढासम्म देख्यो ।	कौवाले डिलबाट टाढासम्म देख्यो ।
(ग)	सङ्ग्रहालयमा पुराना सामग्री सुरक्षित राखिन्छ ।	सङ्ग्रहालयमा नयाँ सामग्री प्रदर्शन गरिन्छ ।	सङ्ग्रहालयमा विदेशी सामग्री सुरक्षित राखिन्छ ।	सङ्ग्रहालयमा खेलकुदका सामग्री सजाएर राखिन्छ ।	सङ्ग्रहालयमा वाद्यवादनका सामग्री सुरक्षित राखिन्छ ।

The distribution of scores for this subtask (*Figure 24*) also indicates that about 50% of students were able to respond correctly to zero items or to only one item in the subtask. Only one-fourth (25.9%) of students were able to solve all of the questions in the subtask.

Figure 24: Distribution of scores for the grade 3 word and sentence identification subtask



The overall average score on the word and sentence identification subtask was 51.1%. The result indicates that out of three questions, students were able to solve less than two items correctly on average. Item-level disaggregation (*Table 13*) indicates that the first item was comparatively easy, as it was correctly solved by 62.7% of students. It is surprising that more

students were able to identify a five-word sentence (50.9%) than a four-word sentence (39.7%). This result might have been due to the choices of distractors for the questions.

Table 13: Average item scores for the grade 3 word and sentence identification subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
1a	Identify word	62.7%	3.6%
1b	Identify four-word sentence	39.7%	3.5%
1c	Identify five-word sentence	50.9%	3.0%

Subtask 2 assessed student vocabulary. The first item focused on defining a word, whereas the second and third items focused on knowledge of antonyms and synonyms. The items in the subtask were multiple choice. There were five possible responses for each item, with one correct option and four distractors. The subtask details are presented in *Figure 25*.

Figure 25: Student stimulus for the grade 3 vocabulary subtask

उपकार्य २
शब्द भण्डार

(क) सहयोग गर्ने मन भएको मानिसलाई भनिन्छ ।

मनकारी	आज्ञाकारी	स्वाभिमानी	इमान्दार	मिहिनेती
--------	-----------	------------	----------	----------

(ख) स्वस्थ शब्दको उल्टो अर्थ दिने शब्द हो ।

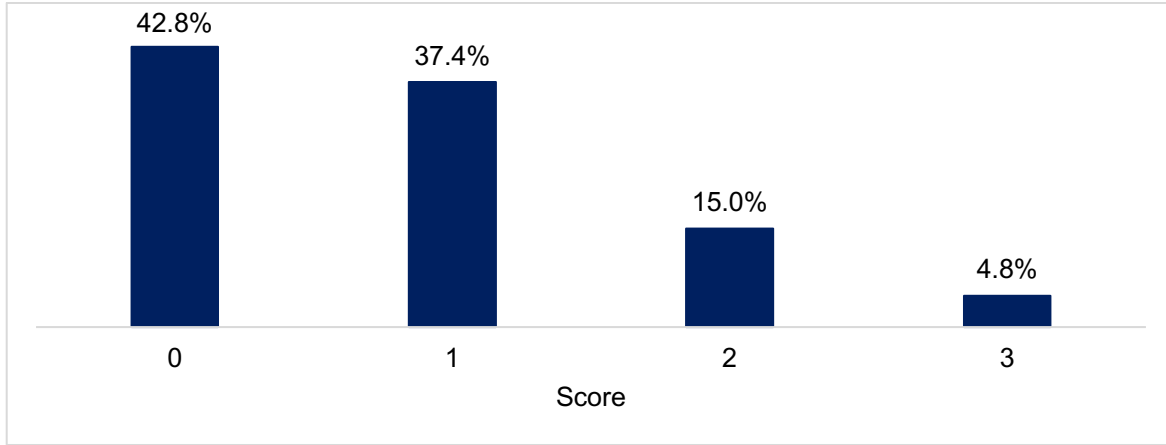
लोभी	फुर्तिलो	रोगी	जाँगरिलो	चलाख
------	----------	------	----------	------

(ग) सम्पदाको अर्थ हो ।

सम्मान	सन्तोष	सफल	सचेत	सम्पत्ति
--------	--------	-----	------	----------

The distribution of scores (*Figure 26*) shows that 42.8% of the students were unable to respond correctly to any items. Another 37.4% of students solved only one question in this subtask. This result shows that many students struggled with vocabulary.

Figure 26: Distribution of scores for the grade 3 vocabulary subtask



The overall average percentage on the vocabulary subtask was 27.3%. Similarly, the disaggregated average scores by subtask (*Table 14*) indicate that fewer than one-third of the students were able to respond correctly to any of the three items.

Table 14: Average item scores for the grade 3 vocabulary subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
2a	Define a word	28.9%	2.9%
2b	Antonyms	23.6%	2.2%
2c	Synonyms	29.2%	2.4%

Subtask 3 assessed students' ability to separate the words in a sentence in which all words were joined together—that is, they appeared without spaces between words. Three-word, four-word, and five-word sentences were asked in the first, second, and third questions, respectively. The student stimulus is presented in *Figure 27*.

Figure 27: Student stimulus for the grade 3 word separation subtask

उपकार्य ३
जोडिएका शब्द छुट्याउने

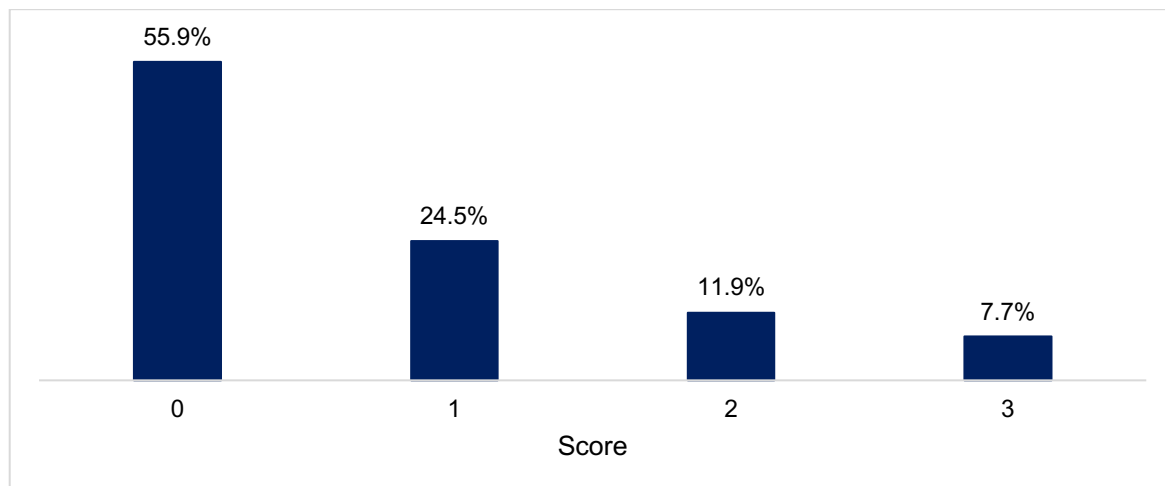
(क) पोखरीतुलोछ ।
..... |

(ख) बारीमागोलभैँडापाकेकोछ ।
..... |

(ग) दराजमाकिताबमिलाएराखिएकाछन् ।
..... |

Figure 28 indicates that more than half (55.9%) of students were not able to solve a single question correctly. In addition, about four-fifths (80.4%) of the students answered zero or one question from this subtask correctly.

Figure 28: Distribution of scores for the grade 3 word separation subtask



The overall average score on the word separation subtask was 23.8%. This result indicates that out of three questions, students were able to solve less than one question correctly on average. The disaggregated scores by subtask (**Table 15**) indicate that the first question was easier than the other two items. The first item—to separate a three-word sentence—was correctly solved by 41.8% of students, whereas the second and third items—separation of four-word and five-word sentences—were solved by only 14.7% and 14.8% of total students, respectively.

Table 15: Average item scores for the grade 3 word separation subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
3a	Three-word sentence	41.8%	3.7%
3b	Four-word sentence	14.7%	2.3%
3c	Five-word sentence	14.8%	2.6%

Subtask 4 assessed children’s dictation skills. In this subtask, students were asked to write sentences correctly as the teacher said them, repeating each item three times. The first item in the subtask was to write a three-word sentence, whereas the second was a four-word sentence. The third was a five-word sentence with more difficult words. **Figure 29** is the student stimulus for dictation.

Figure 29: Student stimulus for the grade 3 dictation subtask

उपकार्य ४
श्रुतिलेखन

(क) म भात खान्छु ।
(ख) आँगनमा पुतली उडेका छन् ।
(ग) हामीले ट्राफिक नियम पालना

(क) |
(ख) |
(ग) |

The average percentage score on this subtask was 15.3%, or less than one item out of three completed without errors. As indicated in *Figure 30*, a very minimal percentage of students (2.3%) was able to respond correctly to all three items on the dictation subtask.

Figure 30: Distribution of scores for the grade 3 dictation subtask

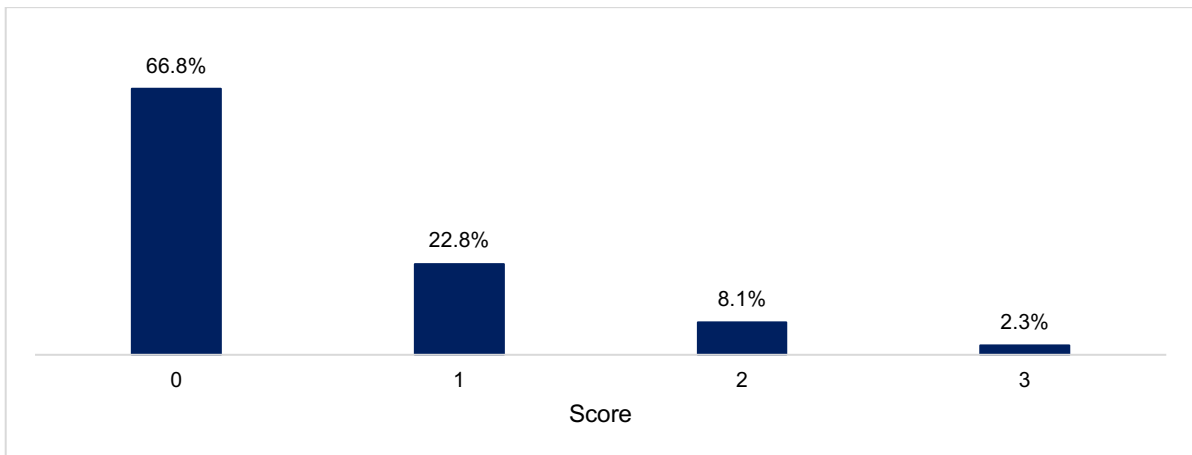


Table 16 also presents the average percentage of students who could correctly answer the different items in the subtask. The first question was dictation of a three-word sentence, and 31.0% of the students were able to respond correctly. The second and third questions were completed correctly by lower proportions of students. Only 10.0% were able to correctly complete the second item, and the third item was most difficult of all, with only 4.7% of students able to complete the item correctly.

Table 16: Average item scores for the grade 3 dictation subtask

Subtask	Description	% of students who answered correctly	Standard error
4a	Three-word sentence	31.0%	3.8%
4b	Four-word sentence	10.0%	1.8%
4c	Five-word sentence	4.7%	0.8%

Subtask 5 assessed the listening comprehension ability of students. The teacher read a 30-word passage and asked three questions about it. The first item was a short-answer question and could be answered based on information provided explicitly in the first or second sentence of the paragraph. The second question's answer also could be found directly in the text. The third was an inferential question for which students had to build answers from information in at least two sentences in the text. The items in the subtask were multiple choice, with five answer options, including one correct option and four distractors. The student stimulus is presented in *Figure 31*.

Figure 31: Student stimulus for the grade 3 listening comprehension subtask

उपकार्य ५
श्रुतिबोध

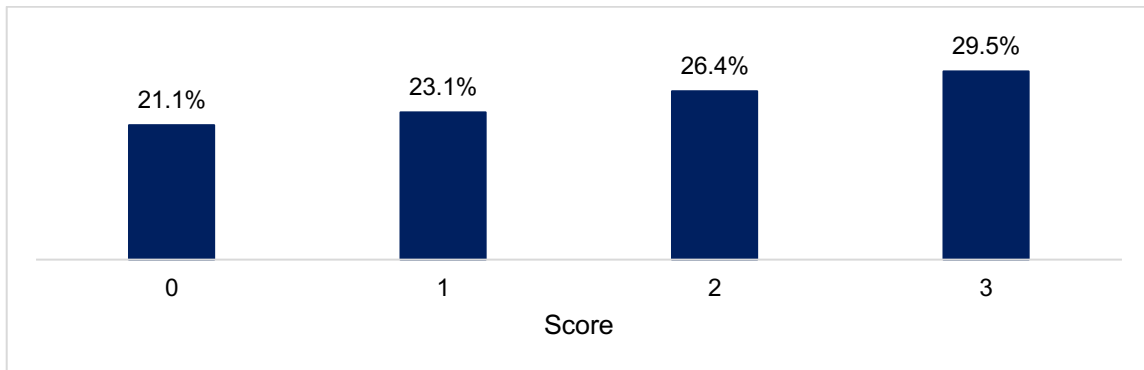
प्रज्ञा र बिशाखा विद्यालय जाँदै थिए। उनीहरूले बाटोमा भोला देखे। प्रज्ञाले हतपत भोला टिप्न खोजिन्। बिशाखाले प्रज्ञालाई भोला टिप्न दिइनन्। उनले अर्काको सामान टिप्न हुँदैन भनिन्। दुवैले शिक्षकलाई खबर गरे।

- (क) प्रज्ञा र बिशाखा कहाँ जाँदै थिए ?
(ख) प्रज्ञा र बिशाखाले कसलाई खबर गरे ?
(ग) बिशाखाले प्रज्ञालाई किन भोला टिप्न दिइनन् ?

(क)	विद्यालय	शौचालय	पुस्तकालय	भोजनालय	सङ्ग्रहालय
(ख)	साथीलाई	शिक्षकलाई	आमालाई	बुबालाई	दिदीलाई
(ग)	अर्काको सामान भएकोले	भोला नराम्रो भएकाले	भोला फोहर भएकाले	भोला फाटेको हुनाले	भोला खाली भएकाले

Analysis of each item in the subtask (*Figure 32*) identified that 29.5% of the students were able to solve all three questions from this subtask, while just over 20% could not answer any of the questions.

Figure 32: Distribution of scores for the grade 3 listening comprehension subtask (30 words)



The overall average score for the listening comprehension subtask was 54.8%. The result shows that students performed better on the listening comprehension task than on the other tasks. Looking at disaggregated results for the items in the subtask (*Table 17*), slightly fewer than two-thirds (64.3%) of the students were able to solve the first question, whereas the second question was solved by 56.0%. However, only 44.0% of students were able to solve the third, inferential question.

Table 17: Average item scores for the grade 3 listening comprehension subtask (30 words)

Subtask	Description	Percentage of students who answered correctly	Standard error
5a	Short answer, explicit	64.3%	3.9%
5b	Short answer, explicit	56.0%	4.2%
5c	Inferential from at least two sentences	44.0%	2.8%

Subtask 6 assessed reading comprehension ability. Students had to read a 60-word passage and then answer three questions about it. The first and second questions could be answered directly by referring to the text; the third question was inferential, demanding that the student consider information from two or more sentences from the text. The items in the subtask were multiple choice. There were five answer options, with one correct option and four distractors. *Figure 33* is the student stimulus for reading comprehension.

Figure 33: Student stimulus for the grade 3 reading comprehension subtask (60 words)

उपकार्य ६

पठनबोध

लाम्पाको गाउँ नजिकै बाक्लो जङ्गल थियो । गाउँलेहरू जङ्गलबाट घाँस दाउरा ल्याउँथे । मानिसहरू बिरामी हुँदा जङ्गलबाटै जडीबुटी ल्याउँथे । जङ्गलमा धेरै चराचुरुङ्गी र जनावरहरू बस्थे । बस्ती बढ्दै गयो । खेती र बसोबासका लागि वन फडानी बढ्दै गयो । वन विनाशले खडेरी र बाढी पहिरो बढ्दै गए । गाउँलेहरू चिन्तित भए । वन विनाशले गर्दा विपत्ति आएको थाहा पाए । सबैले वृक्षरोपण गरी वन संरक्षण गर्ने निधो गरे ।

(क) लाम्पाको गाउँनजिकै के थियो ?

पोखरी	पहाड	जङ्गल	खोला	मैदान
-------	------	-------	------	-------

(ख) मानिसहरू बिरामी हुँदा जङ्गलबाट के ल्याउँथे ?

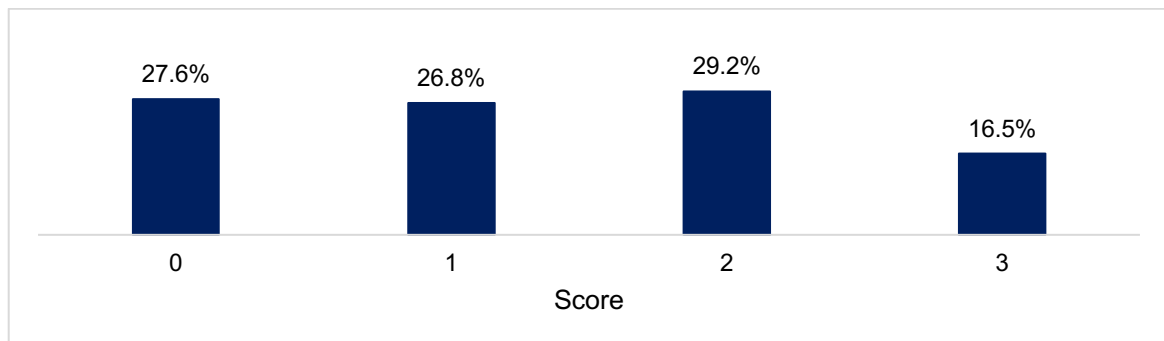
दाउरा	घाँस	पानी	सोतर	जडीबुटी
-------	------	------	------	---------

(ग) जङ्गल संरक्षण गर्न के गर्नुपर्छ ?

रुखबिरुवा काट्नुपर्छ	बस्ती बढाउनुपर्छ ।	जनावर लखेट्नुपर्छ ।	वृक्षरोपण गर्नुपर्छ ।	जङ्गलमा खेती गर्नुपर्छ ।
----------------------	--------------------	---------------------	-----------------------	--------------------------

The average percentage score for this subtask was 44.8%. This result indicates that the average correct answer per student was slightly more than one question out of three. The distribution in **Figure 34** shows that about one-fourth (27.6%) of the students could not answer a single question from this subtask. At the other end of the spectrum, the percentage of students who could answer all three questions was only 16.5%.

Figure 34: Distribution of scores for the grade 3 reading comprehension subtask (60 words)



Looking at the average scores on each of the items within the subtask (**Table 18**), more than half of the students were able to answer the first and second questions (53.4% and 54.2% respectively) whereas the third, inferential question was solved by only one-fourth (27.0%) of the students. This finding shows that students had more difficulty with the higher level (inferential) comprehension question.

Table 18: Average item scores for the grade 3 reading comprehension subtask (60 words)

Subtask	Description	Percentage of students who answered correctly	Standard error
6a	Short answer, explicit	53.4%	3.9%
6b	Short answer, explicit	54.2%	4.2%
6c	Inferential from two or more sentences	27.0%	2.6%

Subtask 7 involved calendar reading. In this subtask, a month from the Nepali calendar was provided and three questions based on the calendar shown were asked. The first question was to identify the day of the last date of the month, while the second was to understand the relationship between festival and date. The third question was to identify the last day of the previous month by looking at the calendar for the month. The items in the subtask were multiple choice, with one correct option and four distractors. *Figure 35* shows the details of the subtask.

Figure 35: Student stimulus for the grade 3 calendar reading subtask

उपकार्य ७

पात्रो (क्यालेन्डर) पठन

माघ २०७५			Jan/Feb 2019			
Sunday आइतबार	Monday सोमबार	Tuesday मंगलबार	Wednesday बुधबार	Thursday बिहीबार	Friday शुक्रबार	Saturday शनिबार
		नवमी १ 15 माघे संक्रान्ति, माघी पर्व (सम्बन्धितलाइ बिदा)	दशमी २ 16	एकादशी ३ 17 पुत्रवा एकादशी	द्वादशी ४ 18	त्रयोदशी ५ 19
चतुर्दशी ६ 20	पुर्णिमा ७ 21 श्री स्वस्थानी व्रत आरम्भ	प्रतिपदा ८ 22	तृतीया ९ 23	चतुर्थी १० 24	पञ्चमी ११ 25	षष्ठी १२ 26
सप्तमी १३ 27	अष्टमी १४ 28	नवमी १५ 29	दशमी १६ 30 सहिद दिवस	एकादशी १७ 31 घटतिला एकादशी	द्वादशी १८ 01	त्रयोदशी १९ 02
चतुर्दशी २० 03	औशी २१ 04	प्रतिपदा २२ 05 सोनाइ ल्होसुर सम्बन्धितलाइ बिदा	द्वितीया २३ 06	तृतीया २४ 07	तृतीया २५ 08	चतुर्थी २६ 09
पञ्चमी २७ 10 सरस्वती पूजा (विशेषणसंस्था बिदा)	षष्ठी २८ 11	सप्तमी २९ 12				

(क) यो महिनाको अन्तिम दिन कुन बार परेको छ ?

आइतबार	सोमबार	मङ्गलबार	बुधबार	बिहीबार
--------	--------	----------	--------	---------

(ख) सरस्वती पूजाको अघिल्लो दिन कति गते परेको छ ?

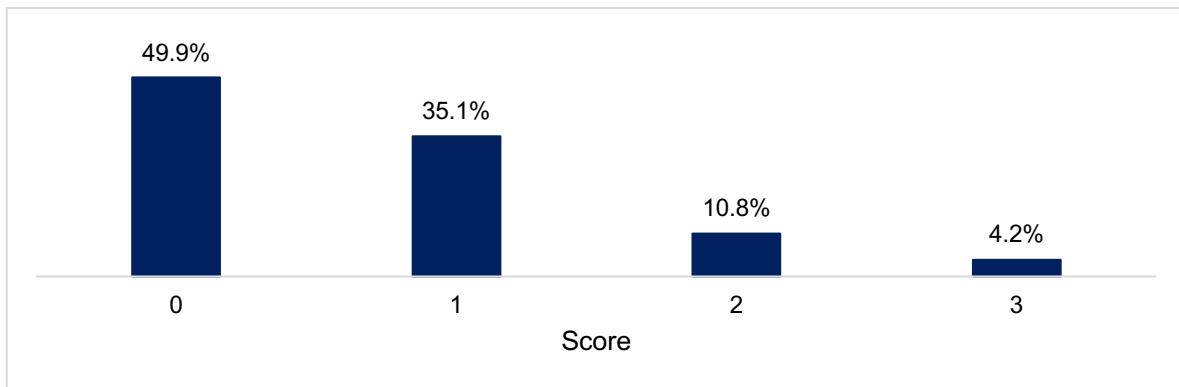
२४	२५	२६	२७	२८
----	----	----	----	----

(ग) अघिल्लो महिनाको अन्तिम दिन कुन बार पर्छ ?

सोमबार	मङ्गलबार	बुधबार	बिहीबार	शुक्रबार
--------	----------	--------	---------	----------

Figure 36 shows that about half (49.9%) of the students were not able to answer a single question from this subtask. Very few students (4.2%) were able to answer all three questions.

Figure 36: Distribution of scores for the grade 3 calendar reading subtask



The overall average percentage for the subtask was 23.1%. This result shows that students were, on average, able to answer less than one question out of three correctly in this subtask. The analysis of average scores on each of the items in the subtasks (*Table 19*) indicates that, as with the grade 2 results, students struggled with viewing and understanding the calendar, which could signify difficulty with visual literacy—although we cannot be sure that this result would translate to all types of multimodal texts.

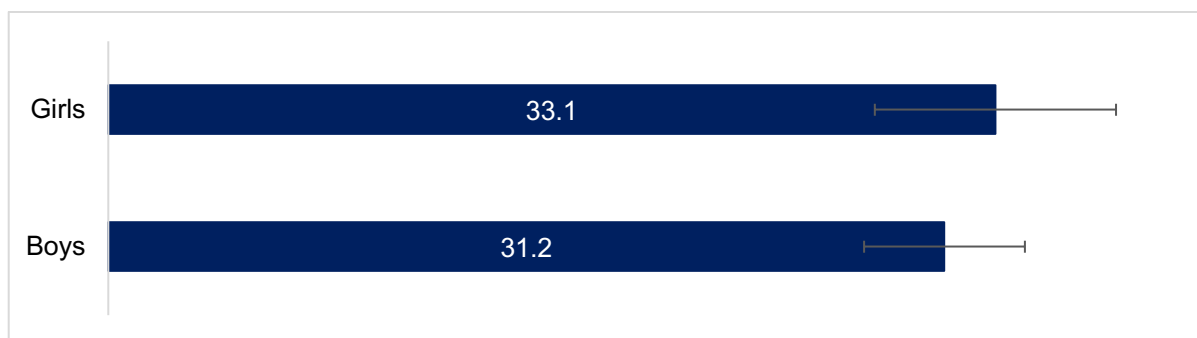
Table 19: Average item scores for the grade 3 calendar reading subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
7a	Day of last date of month	26.7%	2.9%
7b	Festival and date	23.8%	2.3%
7c	Last day of previous month	18.8%	2.2%

3.2.3 Reading Achievement by Sex and Language (Grade 3)

Student reading achievement was disaggregated by the sex of the students to learn whether scores varied for boys and girls. As shown in *Figure 37*, the average percentage score for girls was slightly higher than for boys. The difference, however, was not statistically significant.

Figure 37: Average percentage scores of grade 3 students, by sex



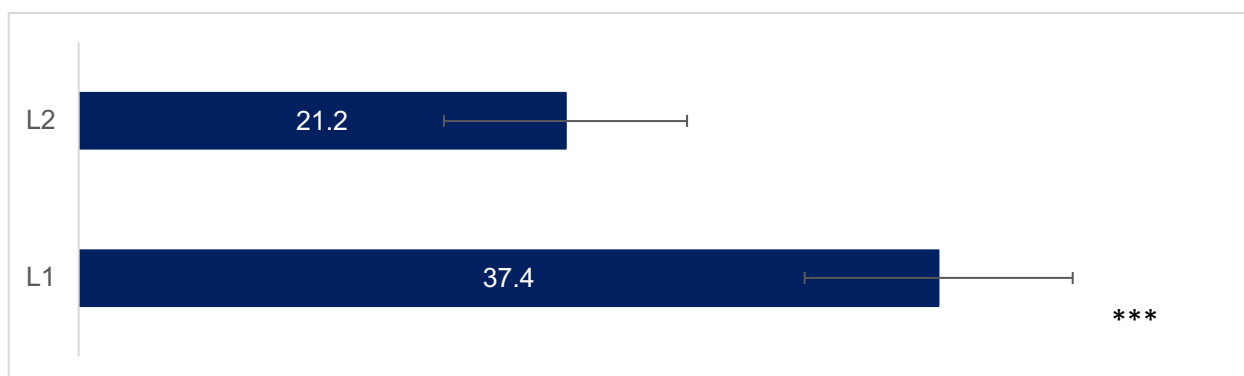
A similar result was found throughout the subtasks in the assessment, with no statistically significant achievement differences for any subtasks (*Table 20*).

Table 20: Average percentage scores of grade 3 students, by subtask and sex

Subtask	Average percent score (boys)	Average percent score (girls)	Difference (boys – girls)
St1: Word and sentence identification	48.9	53.1	-4.2
St2: Vocabulary	25.1	29.2	-4.1
St3: Separation of words	20.9	26.3	-5.4
St4: Dictation	13.6	16.7	-3.1
St5: Listening comprehension	57.9	52.0	6.0
St6: Reading comprehension	43.7	45.8	-2.1
St7: Calendar reading	22.0	24.1	-2.1

As with grade 2, in grade 3, language was also found to be a significant factor in students’ reading performance. As shown in *Figure 38*, the average score for Nepali L1 students nearly 8 out of 21 correct, whereas it was roughly 4 correct out of 21 for Nepali L2 students. The difference between the two groups of more than 16 percentage points on the overall score was statistically significant.

Figure 38: Average percentage scores of grade 3 students, by language



*** $p < .01$.

In addition, the difference by language was significant on all subtasks except vocabulary and calendar reading. *Table 21* shows the students' reading achievement disaggregated by language for each subtask.

Table 21: Average percentage scores of grade 3 students, by subtask and language

Subtask	Average percent Score (L1)	Average percent score (L2)	Difference (L1 – L2)
St1: Word and sentence identification	58.5	35.4	23.1***
St2: Vocabulary	29.5	22.6	6.9
St3: Separation of words	28.3	14.2	14.1**
St4: Dictation	18.0	9.4	8.6*
St5: Listening comprehension	64.7	33.7	31.0***
St6: Reading comprehension	52.0	29.5	22.5***
St7: Calendar reading	24.7	19.7	4.9

*** $p < .01$, ** $p < .05$.

4 Linking EGRA and CB-EGRA Data

4.1 Rationale for Linking

One of the major objectives of this study was to develop a model by which oral reading fluency, comprehension, and overall reading ability can be measured using an established assessment tool in the Nepal context. For this, we collected mini-EGRA scores on a subsample basis from each school that was sampled for the CB-EGRA. We then conducted analyses to arrive at the statistical models that best described the ORF, comprehension, and reading ability of the students in the study. The statistical equating model developed through this process is simple enough to be understood by local-level education stakeholders for extrapolating students' ORF and comprehension skills based on CB-EGRA scores in the future. This section provides background information on assessment linking and equating, and describes the statistical modeling that was used to equate EGRA and CB-EGRA scores, as well as the reading outcome values that were extrapolated from the model.

Assessment linking is a process in which scores on a first educational assessment are linked to a second educational outcome scale without having a second assessment having to be administered every time. This linking is achieved by using a statistical equating process that allows for predicting what a student's score would be on the second test based on their score on the first, using data from students who have been assessed on both tests.

The assessment linking approach is widely used and accepted internationally. For example, when high school students are applying to universities in the United States, they often take either the ACT or SAT. However, these assessments are different in content and approach, so it has been challenging to determine what is an equivalent score on the ACT for a student's score on the SAT, and vice-versa. Compounding the situation is that some students take only one of the tests while a university might accept only the other. This challenge has been mitigated through assessment linking using data from students who have taken both assessments to establish a relationship between results on the two tests (College Board and ACT 2018). The College Board, the U.S. nonprofit that manages the SAT, has stated that this process can be used to compare SAT and ACT scores across students, establish admissions and scholarship policies, and convert scores for use in a model or index that predicts applicants' likely future enrollment and performance in college. However, the College Board went on to state that this process should not be used for averages or ranges.

The most similar assessment linking approach aligned with EGRP II's evaluation design is Istation,⁷ which is a group-administered, online reading assessment. The Istation assessment is similar in its subtask composition to a CB-EGRA. Istation has replaced one-on-one assessments of reading fluency and comprehension—for example, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; see University of Oregon n.d.) assessment, similar to the EGRA—in several states in the United States. Instead, students take an Istation online assessment with multiple-choice and open-response questions, which is then converted to a text fluency measure scale. This scale has been shown to be highly correlated with outcomes

⁷ Istation home page: <https://www.istation.com>.

on the traditional oral reading task (see, for example, Campbell et al. 2018; Mathes et al. 2016).

Because the CB-EGRA does not include an ORF measure, but some EGRP II reporting indicators demand ORF, we used assessment linking to identify a statistical model that would relate CB-EGRA scores to the ORF scores that are typically derived from the EGRA. In order to do this, we first created a composite score to calculate the overall CB-EGRA average percentage score. Then, to establish the relationship between the overall CB-EGRA average percentage score and ORF, we trialed common approaches to assessment linking (Muraki et al. 2000) to maximize statistical precision and reliability and determine the best model fit for our data. This iterative process looked specifically at the variance explained between the two assessments. The process included fitting curve-fitting models and fitting only certain subtasks in the CB-EGRA. Ultimately, we concluded that the best model for equating CB-EGRA and reading fluency was a simple linear regression model utilizing the most possible items from CB-EGRA (i.e., all subtasks). Retaining all subtasks was important for creating a wider range of scores for the CB-EGRA and a more precise model fit. At the same time, we also used the subtask weights set by ERO’s technical experts, in which the most challenging literacy subtasks were afforded the highest weights.

Rather than equating for the entire range of scores between the two tasks, which is the typical reason for equating in the ACT, SAT, and DIBELS examples provided above, our purpose was to create equivalent scores for emergent and fluent levels of reading skill. These equivalent scores would then be used to estimate the number of students reading at each level. One key limitation of this approach—discussed further in the Limitations section below—is that the CB-EGRA can be linked to existing emergent (15 cwpm) and fluent (45 cwpm) benchmarks in Nepal, but cannot be used to create an average reading fluency or zero score percentage from the CB-EGRA.

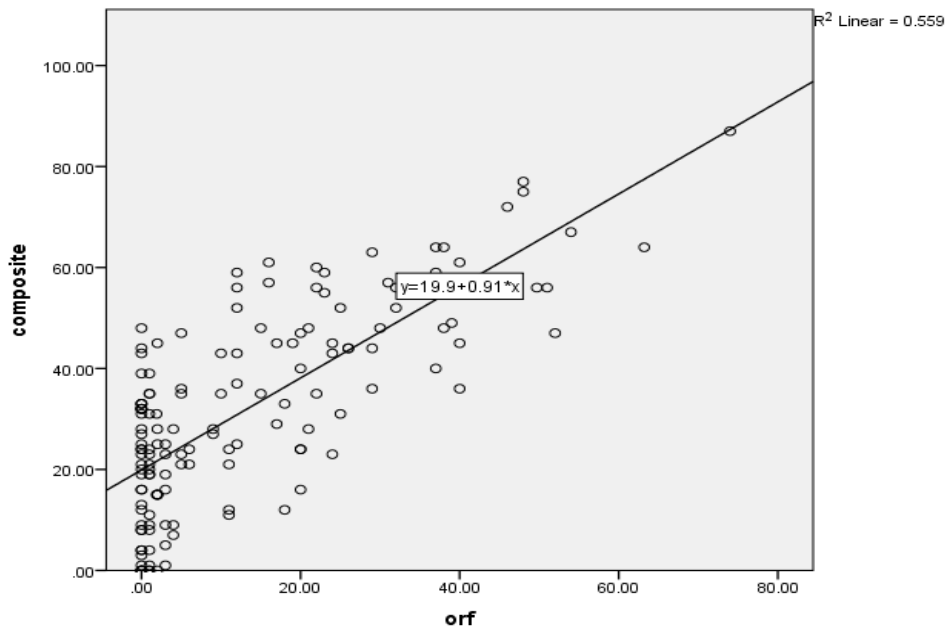
Next, this section describes the statistical models that we arrived at through the equating process.

4.1.1 The Model to Predict ORF (Grade 2)

Figure 39 is the model that related the ORF score and the CB-EGRA average scores for second grade.

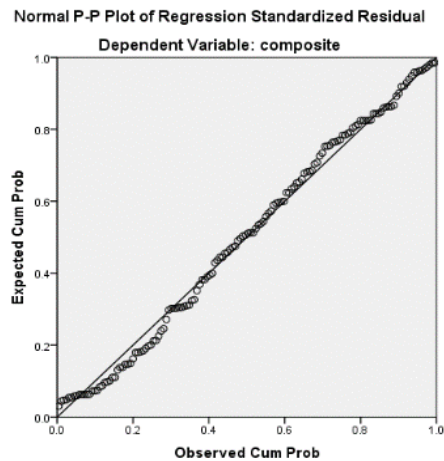
Figure 39: Regression model to describe the CB-EGRA average percentage score from ORF (grade 2)

Note. $y = a + bx + e$, where: a = constant; x = ORF, with b coefficient; e = error term; and y = predicted CB-EGRA composite score.



Constant	$a = 19.901^{***}$
$x = \text{ORF}$	$b = 0.911^{***}$
R	0.748
R^2	55.9%
Adjusted R^2	55.6%
No. of observations	147

*** $p < .01$.



The average CB-EGRA percentage score was the linear function of the ORF score with the correlation coefficient $r = 0.75$. The residuals were also normal for the model. Thus, we can claim that the ORF score describes the composite CB-EGRA average percentage score, and the model can be used for future predictions. The model can be expressed as:

$$\text{Average grade 2 CB-EGRA percentage score} = 19.901 + 0.911 \times \text{ORF}$$

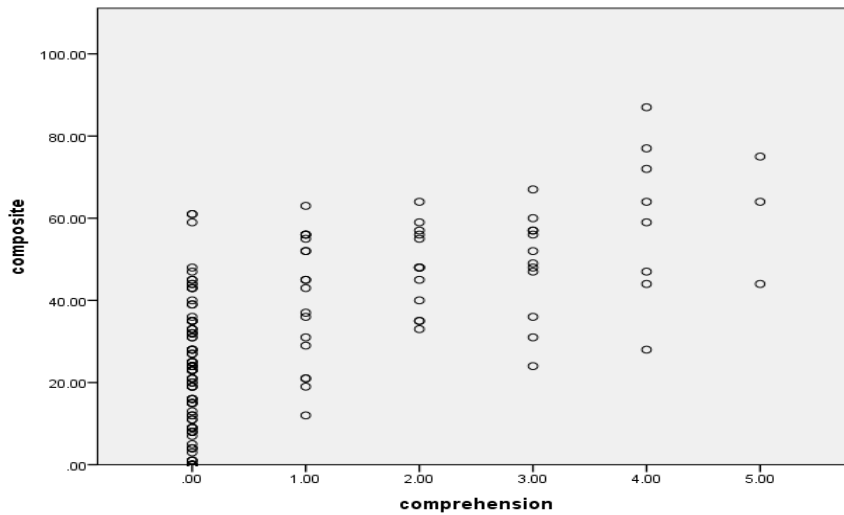
$$\text{Therefore, Grade 2 ORF} = (\text{average CB-EGRA percentage score} - 19.901) / 0.911$$

4.1.2 The Model to Predict Reading Comprehension (Grade 2)

Figure 40 is the model that calculates the average CB-EGRA percentage score by using the EGRA reading comprehension score.

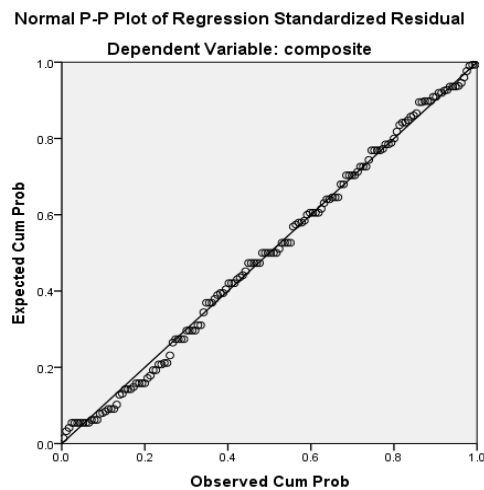
Figure 40: Regression model to describe the overall CB-EGRA percentage score from reading comprehension (grade 2)

Note. $y = a + bx + e$, where: a = constant; x = reading comprehension (number of correct answers out of 5 questions), with b coefficient; e = error term; and y = CB-EGRA average percentage score.



Constant	$a = 24.003^{***}$
x = Average CB-EGRA percentage score (measured in 100%)	$b = 9.201^{***}$
R	0.643
R^2	41.3%
Adjusted R^2	40.9%
No. of observations	147

*** $p < .01$.



The average CB-EGRA percentage score was the linear function of reading comprehension with the correlation coefficient $r = 0.64$. The residuals were also normal for the model. Thus, we can claim that the reading comprehension score describes the composite CB-EGRA average percentage score, and the model can be used for future predictions. The model can be expressed as:

$$\text{Average grade 2 CB-EGRA percentage score} = 24.003 + 9.201 \times \text{average comprehension}$$

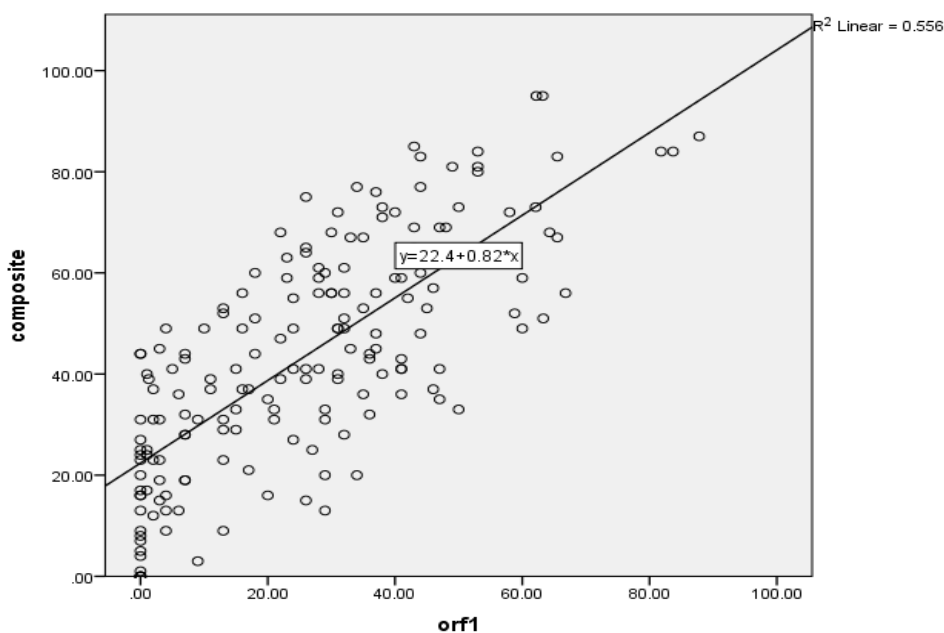
$$\text{Therefore, average grade 2 comprehension} = (\text{average CB-EGRA percentage score} - 24.003) / 9.201$$

4.1.3 The Model to Predict ORF (Grade 3)

Figure 41 is the model that calculates overall CB-EGRA average scores by using the EGRA ORF score.

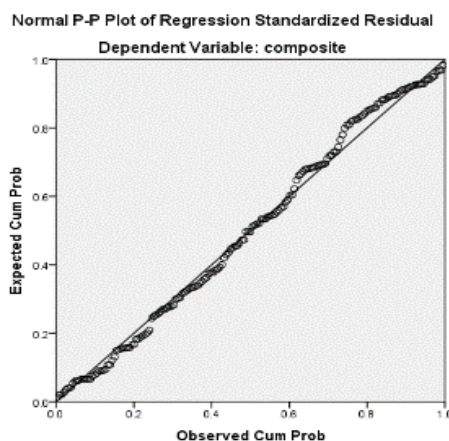
Figure 41: Regression model to describe the CB-EGRA average percentage score from ORF (grade 3)

Note. $y = a + bx + e$, where: a = constant, x = ORF, with b coefficient; e = error term; and y = predicted average CB-EGRA percentage score.



Constant	$a = 22.399^{***}$
x = Average CB-EGRA percentage score (measured in 100%)	$b = 0.817^{***}$
	$R = 0.746$
	$R^2 = 55.6\%$
	Adjusted $R^2 = 55.4\%$
No. of observations	176

*** $p < .01$.



The overall CB-EGRA percentage score was the linear function of ORF with the correlation coefficient $r = 0.75$. The residuals were also normal for the model. Thus, we can claim that the composite CB-EGRA average percentage score is described as ORF for grade 3, and the model can be used for future predictions. The model can be expressed as:

$$\text{Average grade 3 CB-EGRA percentage score} = 22.399 + 0.817 \times \text{ORF}$$

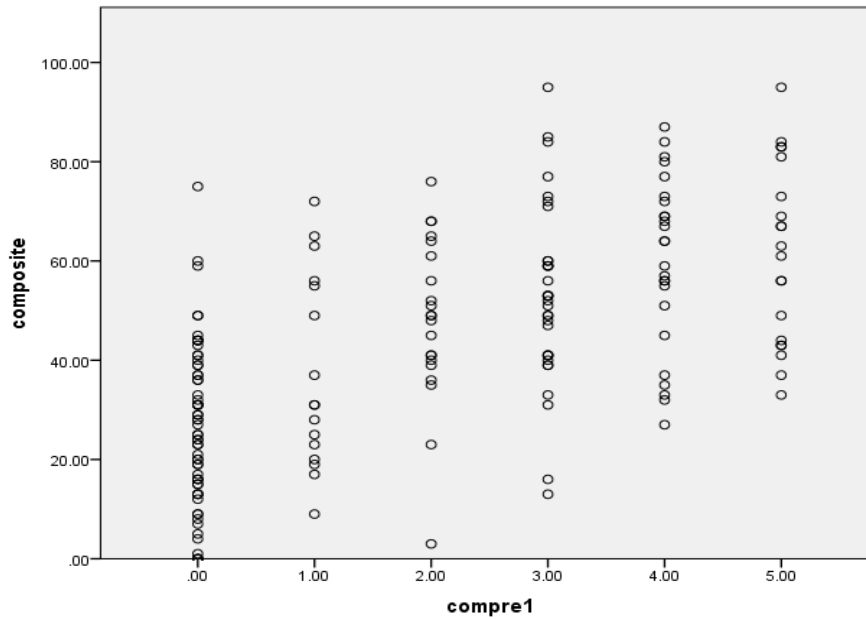
$$\text{Therefore, grade 3 ORF} = (\text{average CB-EGRA percentage score} - 22.399) / 0.817$$

4.1.4 The Model to Predict Comprehension (Grade 3)

Figure 42 is the model that calculates ORF by using CB-EGRA average scores.

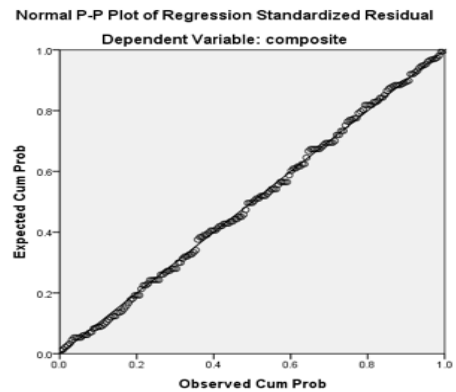
Figure 42: Regression model to describe the overall CB-EGRA percentage score from reading comprehension (grade 3)

Note. $y = a + bx + e$, where: a = constant; x = average CB-EGRA percentage score, with b coefficient; e = error term; and y = predicted reading comprehension (number of correct answers out of 5 questions).



Constant	$a = 28.149^{***}$
x = Average CB-EGRA percentage score (measured in 100%)	$b = 7.674^{***}$
R	0.626
R^2	39.2%
Adjusted R^2	38.8%
No. of observations	176

*** $p < .01$.



The overall CB-EGRA percentage score was the linear function of reading comprehension with $r = 0.63$. The residuals were also normal for the model. Thus, we can claim that reading comprehension describes the composite CB-EGRA average percentage score for grade 3, and the model can be used for future predictions. The model can be expressed as:

$$\text{Average grade 3 CB-EGRA percentage score} = 28.149 + 7.674 \times \text{average comprehension}$$

$$\text{Therefore, grade 3 average comprehension} = (\text{average CB-EGRA percentage score} - 28.149) / 7.674$$

4.2 Extrapolation of Reading Achievement

4.2.1 Equivalent Scoring

Using the models above, we created equivalent CB-EGRA scores for emergent and fluent reader benchmarks (*Table 22*). These scores can be used to calculate the percentage of

students at baseline meeting Nepal’s emergent and fluent benchmarks in line with EGRP II’s performance indicators, as described in *Annex A*. These scores will also become benchmark equivalencies for all future CB-EGRAs.

Table 22: Equivalent CB-EGRA scores for emergent and fluent benchmarks (in cwpm)

Grade	Benchmark CB-EGRA scores	
	Emergent	Fluent
2	33.5	60.9
3	34.7	58.9

It is important that the scores appear to be similar across grades, such as 60.9 for the fluency benchmark for grade 2 and 58.9 for grade 3. However, the CB-EGRA assessment tools are different for grades 2 and 3 and the results are, therefore, not directly comparable between the grades.

4.2.2 Students Who Met the Reading Benchmark

The Government of Nepal has set 45 correct words per minute with 80% comprehension as Nepal’s national reading benchmark (MOE 2017). The values extrapolated from the CB-EGRA results in this baseline evaluation were further analyzed to identify the percentage of students who met this reading benchmark. As shown in *Table 23*, 7.4% of grade 2 children and 12.6% of grade 3 children in the EGRP II baseline sample met the benchmark.

Table 23: Percentage of learners who met the reading benchmark (45 words and 80% comprehension), by grade

Grade	Percentage of learners who met the benchmark	Standard error
2	7.4%	2.0%
3	12.6%	2.4%

4.2.3 Comparing EGRP II Baseline Findings with the 2020 NARN and CB-EGRA Scores from Previous Years

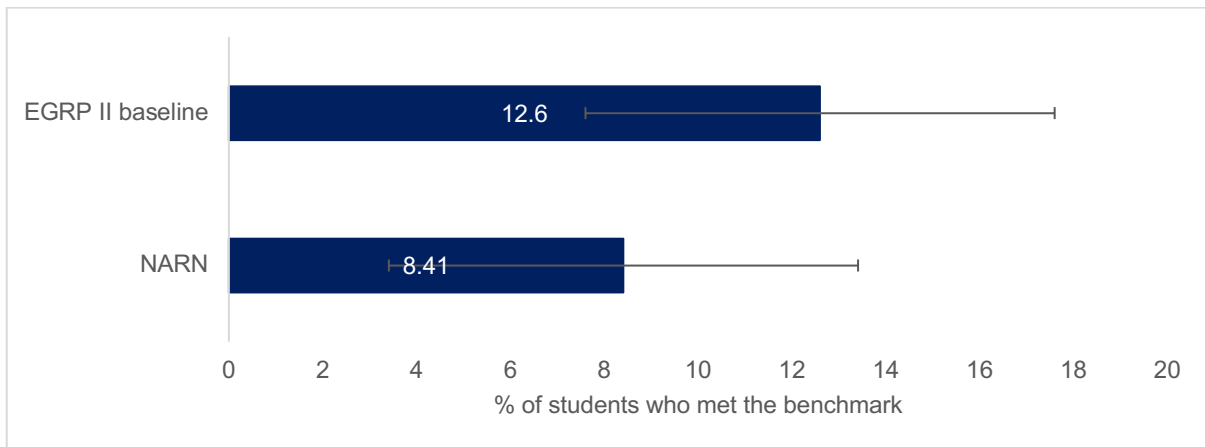
It may be useful for education decision-makers in Nepal to situate the EGRP II baseline findings within broader learning outcome trends in Nepal, particularly assessments that are similar in nature, such as the 2020 NARN and CB-EGRAs from previous years. However, it is also important to understand the potential limitations when direct comparisons of the findings are made between these different assessments.

For example, both EGRP II’s 2021 baseline and the 2020 NARN used a sample-based approach to estimate the percentage of students reaching different reading benchmarks. Consequently, the true population percentage lies within a range, called a confidence interval. For the EGRP II baseline, the estimate of grade 3 students who met the reading benchmark of 45 cwpm with 80% comprehension was 12.6%, with 95% confidence that the true population percentage was between 7.9% and 17.3%. If we compare these values with the 2020 NARN,

we find that the estimate for the NARN was 8.41% with a 95% confidence interval of between 6.8% and 10.0%.

Figure 43 demonstrates that the confidence intervals (the black lines at the end of each blue bar) for the EGRP II baseline and the 2020 NARN estimates overlap. Therefore, while the two estimates have a difference of over 4 percentage points, we cannot be certain that the EGRP II baseline percentage is higher than the NARN with any degree of statistical significance, due to the overlapping confidence intervals. On the other hand, we can conclude that average student performance as measured by both assessments was roughly similar.

Figure 43: Percentage of grade 3 students who met the reading benchmark in the 2021 EGRP II baseline and the 2020 NARN



Another important caveat to keep in mind is that the NARN and the EGRP II baseline were administered to different samples of schools and children. The assessments were also conducted in different years: the NARN in early 2020, before the COVID-19 pandemic; and the EGRP II baseline in the midst of the pandemic in early 2021. Consequently, direct comparisons in the average scores should be interpreted with caution.

Use of the CB-EGRA to help teachers gauge children’s early grade reading skills is one of the core elements of the National Early Grade Reading Program. Annual rollout of CB-EGRAs began in 2017. Typical scores from previous years were substantially higher, on average, than the average scores from this EGRP II baseline assessment in 2021. For instance, previous average grade 2 CB-EGRA scores ranged from 64% to 66%, while the average was 28.5% in the EGRP II baseline. Similarly, average grade 3 CB-EGRA scores ranged from 66% to 68% in the past, with an average of 32.2% in the EGRP II baseline. The differences are captured in **Figure 44** below.

Figure 44: Comparison of average CB-EGRA scores between previous CB-EGRA assessments and the EGRP II baseline

CB-EGRAs from previous years:	2021 EGRP II baseline:
Grade 2: 64%–66%	Grade 2: 28.5%
Grade 3: 66%–68%	Grade 3: 32.2%

These differences in average outcomes on the CB-EGRA could be due to factors such as learning disruptions caused by the COVID-19 pandemic during the 2020–2021 academic year, as well as differences in the samples for the various assessments. At the same time, the drop could also be due in part to how the CB-EGRA was administered during the EGRP II baseline. Specifically, targeted training for teachers conducting the assessment, combined with monitoring by EGRP II staff during test administration, constituted an extra layer of quality oversight for this baseline that is not typically present in CB-EGRAs carried out during the regular course of the academic year. The additional quality oversight could have resulted in lower-than-normal results this year if any possible “grade inflation” by teachers was minimized. Readers should keep these factors in mind when making direct comparisons between average CB-EGRA scores in previous years and average scores in this baseline.

4.2.4 Reading Ability Categories

As discussed in the Executive Summary, apart from the current national reading benchmark of 45 cwpm with 80% comprehension, the Government of Nepal has not yet officially defined reading ability levels or categories that would allow for more nuanced analysis of baseline results. However, in the 2020 NARN report (ERO 2020), ERO assigned readers to one of four categories. Those categories are nonreaders (ORF = 0), initial readers (ORF between 1 and 15), emergent readers (ORF between 16 and 44), and fluent readers (ORF 45 or more). As of finalizing this baseline report, EGRP II was in the process of supporting the GON to revise the national reading benchmark to include categories of readers beyond the “fluent reader” designation.

Because the CB-EGRA used multiple-choice questions with five answer options for most items in most of the subtasks, the likelihood of guessing correctly was 20%, and therefore there was less possibility of scoring very low or zero. As such, it is not meaningful to extrapolate the percentage of nonreaders and initial readers using the equating approach adopted in this evaluation. With this point in mind, **Table 24** provides only the percentages of students categorized as emergent or fluent readers.

Table 24: Categories of readers, by grade

Grade	Emergent reader	Fluent reader
2	27.8% (3.3%)	7.4% (2.0%)
3	29.7% (2.7%)	12.6% (2.4%)

Note. The numbers in parentheses represent standard errors.

5 Summary and Conclusions

The study described in this analysis report was intended to establish a baseline for EGRP II by assessing students' reading performance. For the study, a scientific sampling technique was used to select 45 schools from seven districts. The government's CB-EGRA tools for grades 2 and 3 were used for data collection.

As a group-administered assessment, the CB-EGRA cannot measure ORF. However, standard indicators, such as under the Millennium Development Goals, demand ORF data. Nepal's national reading benchmark (MOE 2017) also includes both ORF and reading comprehension measures. In order to address this gap, we simultaneously collected students' ORF and comprehension data, using a "mini-EGRA," from all sampled schools, on a subsample basis. The team then developed a statistical model to equate the CB-EGRA scores with the mini-EGRA scores. This model was helpful for extrapolating the ORF and comprehension scores for the EGRP II baseline and endline studies. In addition, it will be useful to the GON at the national and subnational levels—including district and palika officials—for identifying, reviewing, and reporting on key reading indicators, such as the number of children reaching the MOEST's early grade reading benchmark.

A total of five research questions were asked in this baseline study. The summary and conclusions from the study are presented as responses to each research question below.

Research Question 1: How do grade 2 and 3 students from the program districts perform in reading skills?

The overall reading achievement measure using the average percentage CB-EGRA score for grade 2 was 28.5%, and for grade 3 it was 32.2%. For both grades, a total of seven subtasks and 21 items were used to assess student reading ability. Thus, this finding means that on average, a child from grade 2 was able to correctly respond to about six items and a child from grade 3 was able to correctly respond to about seven items. As discussed above, these scores are much lower than those of CB-EGRAs conducted in previous years, which reported somewhere between 64% and 66% for grade 2 and between 66% and 68% for grade 3. This study did not produce direct evidence to explain the size of the score discrepancy, but we may conjecture that school closures for more than 10 months in the academic year due to COVID-19 could be an important contributing factor.

Overall, students scored lowest on subtasks measuring writing skills and calendar reading (visual literacy), while they performed highest on listening comprehension.

Research Question 2: In what ways do those levels of reading performance differ for boys and girls?

Reading performance was not significantly different between boys and girls in this baseline study. The average CB-EGRA performance of grade 2 boys was 27.4% and that of girls was 29.6%. Although there was a difference of 2.2 percentage points between girls and boys, this difference was not statistically significant. Similarly, boys from grade 3 scored 31.2% and girls scored 33.1%, a difference of 1.9 percentage points that was also not statistically significantly different. Thus, it can be concluded that there was no significant association between student sex and learning outcomes.

Research Question 3: Are there differences in the reading performance of students who speak Nepali as a first language (L1) versus those who speak Nepali as a second language (L2)?

Students were categorized into two groups according to their home language. Students with Nepali as their home language were categorized as L1 learners and students with languages other than Nepali as their home language were categorized as L2 learners. The study identified that students' home language played a significant role in their reading outcomes, with L1 students performing far better than L2 students on average.

For example, L1 students from grade 2 achieved an average CB-EGRA score of 33.3% whereas L2 students scored only 18.4% on average. The difference of about 15 percentage points was statistically significant. A similar result was found in grade 3, where L1 students achieved an average score of 37.4% while the score for L2 learners was 21.2% on average. The difference between L1 and L2 students of more than 16 percentage points was also statistically significant.

Based on these findings, we can conclude that L2 students were lagging behind their L1 peers in reading outcomes. Grade 3 L2 students achieved nearly the same average scores as L1 children in grade 2, indicating an achievement gap of nearly an entire grade based on home language.

Research Question 4: What model describes the extent relationship between CB-EGRA and fluency, comprehension, and reading ability of the students?

To establish the relationship between the average percentage score on CB-EGRA subtasks and the mini-EGRA (ORF and comprehension), we tested various models and relationships. As described in the discussion earlier in this report on linking EGRA and CB-EGRA data, the composite CB-EGRA score can now be linked to ORF and reading comprehension. We developed four models, which are presented as:

Average CB-EGRA percentage score for grade 2 = $19.901 + 0.911 \times \text{ORF}$

Average CB-EGRA percentage score for grade 2 = $24.003 + 9.201 \times \text{average comprehension}$

Average CB-EGRA percentage score for grade 3 = $22.399 + 0.817 \times \text{ORF}$

Average CB-EGRA percentage score for grade 3 = $28.149 + 7.674 \times \text{average comprehension}$

As a result of this process, EGRP II established robust statistical models for both grades 2 and 3 to extrapolate ORF and reading comprehension scores from CB-EGRA scores. National, district, and local-level governments and schools can now use this model to better understand the status of reading skills for their populations of students.

Research Question 5: What are the baseline percentages of emergent and fluent student readers in grades 2 and 3 in program districts?

We attempted to apply ERO's categorizations of reading ability using the ORF and reading comprehension scores that we had statistically extrapolated from the CB-EGRA and mini-EGRA results. Because of constraints associated with the multiple-choice items in the CB-EGRA, we calculated the figures only for emergent readers and fluent readers (and not for nonreaders or initial readers). Based on this analysis, we concluded that just over one-fourth

of students in both grades in the study sample were emergent readers (27.8% from grade 2 and 29.7% from grade 3). These results offer some cause for optimism that a larger percentage of children may be able to become fluent readers over time if they receive proper support in terms of instruction and materials.

6 Study Limitations

This section describes the limitations that should be considered by those who review and interpret the results of the EGRP II baseline.

Sample Size and Representativeness

The sample for this baseline study was intended to incorporate diversity in relation to geography, students' language, and level of EGRP II's interventions. However, the sample was not nationally representative. As such, findings and results are not generalizable at the national level.

At the same time, the sample was statistically sufficient to generalize the results within the program districts. However, due to resource limitations that affected the sample size, we were not able to generalize the results using lower levels of disaggregation by strata, such as school, district, and province.

Assessment Method

EGRP II adopted the Government of Nepal's tools and group-administered assessment approach to measure student achievement in reading. EGRP II utilized a two-layer cascade training approach, including a training of trainers and a training of classroom teachers, to promote quality and uniformity in administering the CB-EGRA across different locations. However, because it is a group-administered test, children's participation and achievement could theoretically have been affected by factors out of EGRP II's control. Such factors could have included, for example, the accuracy and clearness of each individual teacher's instructions, as well as the volume and tone of each teacher's voice in a group setting.

Lack of Estimates for Nonreaders and Initial Readers

EGRP II developed a statistical model to extrapolate ORF from the CB-EGRA results. As noted previously, because the CB-EGRA is primarily a multiple-choice assessment, it is possible that students obtained some correct answers by guessing. Students who responded to at least one question correctly obtained a nonzero ORF score using the predictive model. This result, however, differs from those observed during previous EGRAs in Nepal, in which many students scored zero on ORF even if they answered items correctly in other subtasks. With this factor in mind, EGRP II has not presented data on students falling in the nonreader or initial reader categories in this baseline, as might typically be done with an EGRA.

Equated Scores Are Estimates

The statistical models for equating EGRA and CB-EGRA scores presented in this baseline are based on the best fit between outcomes on the two tests. However, a key limitation in assessment linking is that the two linked assessments are not identical and therefore measure slightly different knowledge and skills. As such, an ORF score based on a student's CB-EGRA score is a statistically robust estimate rather than a perfect prediction of oral reading fluency and comprehension skill when directly measured. At the same time, conducting full-scale EGRAs requires greater cost and time commitments than CB-EGRAs, and CB-EGRAs have become more widely institutionalized within Nepal's education system. When designing this evaluation approach, EGRP II considered this trade-off between precision and

sustainability to be acceptable, and to offer a useful model for future early grade reading assessments both in Nepal and globally.

References

- Campbell, Laurie O., Glen W. Lambie, Claudia C. Sutter, Ashley R. Bickham, and Lindsay Pulse. 2018. *Measuring the Predictability of Istation Indicators of Progress Early Reading (ISIP-ER) Scores on Florida Standards Assessment (FSA) Scores*. Orlando, Florida: Morgridge International Reading Center, University of Central Florida. <https://www.istation.com/Content/downloads/studies/FSARReadingPredictability.pdf>.
- College Board and ACT. 2018. *Guide to the 2018 ACT®/SAT® Concordance*. <https://collegereadiness.collegeboard.org/pdf/guide-2018-act-sat-concordance.pdf>.
- Education Review Office (ERO). 2017. *Classroom-Based Early Grade Reading Assessment: A Framework*. Bhaktapur: ERO, MOEST, Government of Nepal. https://www.ero.gov.np/upload_file/files/post/1595301167_1390229304_1587620785_1577711485_CBEGRA_Framework-2017.pdf.
- Education Review Office (ERO). 2020. *National Assessment for Reading and Numeracy (Grade 3), 2020*. Sanathimi, Bhaktapur: ERO, MOEST, Government of Nepal. https://www.ero.gov.np/upload_file/files/post/1620735553_899016200_NARN_Final%20Report%20Final%20uploaded%20in%20the%20web.pdf.
- Mathes, Patricia, Joseph Torgesen, and Jeannine Herron. 2016. *Istation's Indicators of Progress (ISIP) Early Reading Technical Report: Computer Adaptive Testing System for Continuous Progress Monitoring of Reading Growth for Students Pre-K through Grade 3*. Dallas, Texas: Istation. https://www.istation.com/Content/downloads/studies/er_technical_report.pdf.
- Muraki, Eiji, Catherine M. Hombo, and Yong-Won Lee. 2000. "Equating and Linking of Performance Assessments." *Applied Psychological Measurement* 24, no. 4, 325–37. <https://doi.org/10.1177/01466210022031787>.
- Ministry of Education. 2017. *Reading Benchmark*. Kathmandu: Ministry of Education.
- Neupane, Sagar Mani, Swadesh Maharjan, Buddhi Chaudhary, Simon King, Jodie Fonseca, and Wendi Ralaingita. 2021. *USAID's Early Grade Reading Program II (EGRP II) in Nepal. Baseline Report Vol. 2, COVID-19 Response: The Home- and Community-Based Schooling Intervention*. Prepared for USAID under Cooperative Agreement No. 72036720CA00001. Research Triangle Park, North Carolina, USA: RTI International.
- University of Oregon. n.d. DIBELS home page. Eugene, Oregon: DIBELS. <https://dibels.uoregon.edu/>.

Annex A: EGRP II MEL Indicator Reporting

This annex summarizes the baseline values for learning outcome indicators in EGRP II's Monitoring, Evaluation, and Learning Plan, as measured through this baseline evaluation.

IND 01_ES. 1-1: Percent of learners targeted for United States Government assistance who attain a minimum grade-level proficiency in reading at the end of grade 2.

Overall: 7.4% (Numerator: 24,394, Denominator: 328,929)

Male: 7.7% (Numerator: 12,295, Denominator: 160,269)

Female: 7.2% (Numerator: 12,099, Denominator: 168,660)

IND 04_Custom: Percent of grade 2 and 3 students classified as fluent readers using national benchmarks

Grade 2

Overall: 7.4% (Numerator: 24,394, Denominator: 328,929)

Male: 7.7% (Numerator: 12,295, Denominator: 160,269)

Female: 7.2% (Numerator: 12,099, Denominator: 168,660)

Grade 3:

Overall: 12.6% (Numerator: 42,045, Denominator: 333,969)

Male: 10.5% (Numerator: 16,454, Denominator: 156,849)

Female: 14.4% (Numerator: 25,591, Denominator: 177,120)

IND 05_Custom: Percent of grade 2 and 3 students classified as emergent readers using national benchmarks.

Grade 2

Overall: 27.8% (Numerator: 91,562, Denominator: 328,929)

Male: 25.5% (Numerator: 40,870, Denominator: 160,269)

Female: 30.1% (Numerator: 50,692, Denominator: 168,660)

Grade 3:

Overall: 29.7% (Numerator: 99,208, Denominator: 333,969)

Male: 28.2% (Numerator: 44,270, Denominator: 156,849)

Female: 31.0% (Numerator: 54,938, Denominator: 177,120)