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

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

Cooperative Agreement No. 72036720CA00001

Cover photo: Grade 2 students in Tanahun 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

List of Figures

Acronyms and Abbreviations

CB-EGRA	Classroom-Based Early Grade Reading Assessment
COVID-19	Coronavirus Disease 2019 (SARS-CoV-2)
cwpm	Correct Words per Minute
EGRA	Early Grade Reading Assessment
EGRP II	Early Grade Reading Program II
ERO	Education Review Office
GON	Government of Nepal
IC	Integrated Curriculum
ICC	Intra-Cluster Correlation
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
NARN	National Assessment for Reading and Numeracy
NEGRP	National Early Grade Reading Program
ORF	Oral Reading Fluency
QED	Quasi-Experimental Design
RTI	RTI International (registered trademark and trade name of Research Triangle Institute)
St	Subtask
USAID	United States Agency for International Development

Acknowledgments

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 package[1] 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 EGRP (called Level 3).

In addition, EGRP II is implementing a home- and community-based schooling program in Province 2, covering 16 disadvantaged *palikas* in eight program districts, with plans for further expansion as the COVID-19 pandemic continues. The key objective of the home- and community-based schooling program is to help students catch up on learning that has been disrupted by the pandemic, in line with Nepal's curriculum expectations for grades 1–3.

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

This 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? (b) In what ways do those levels of reading performance differ for boys and girls? (3) What is the value added of the COVID-19 response component beyond the EGRP II support for the NEGRP minimum package? (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?

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, comprehension), plus writing. For the COVID-19 intervention's study sample, 47 schools, with 920 students from grade 2 and 899 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-

---

[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.

supported schools). The CB-EGRA was conducted by trained schoolteachers. The EGRP II team put in place a number of measures to ensure data quality, including rigorous training; 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 in Province 2 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. The overall achievement scores were calculated using a composite average of scores across the subtasks, presented as a percentage.

On average, grade 2 students were able to correctly answer 5 out of the 21 total questions in the assessment. The average grade 2 scores for each subtask were 46.2% for letter/*matra*[2] identification; 37.5% for word and sentence identification; 32.7% for vocabulary; 7.1% for dictation; 34.8% for listening comprehension; 24.7% for reading comprehension; and 31.7% for calendar reading.

These results indicate that grade 2 students performed somewhat better on letter/matra identification and listening comprehension. 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 6 out of the 21 total questions in the assessment. The average grade 3 scores for each subtask included 42.2% for word and sentence identification; 32.4% for vocabulary; 21.6% for word separation; 13.5% for dictation; 41.9% for listening comprehension; 40.6% for reading comprehension; and 33.0% for calendar reading.

These results indicate that, as in grade 2, students in grade 3 performed better on word and sentence identification and listening comprehension, with more than 40% of students on average able to solve the questions from the 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 6 out of 21 questions correctly on average. Because all children in the sample for this study speak Nepali as a second language (L2) and there were no children who speak Nepali as a first language (L1), it was not possible to compare achievement between Nepali L1 and L2 speakers as was done in Volume 1 of this report (see Neupane et al. 2021).

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

---

[2] *Matras* are the Nepali consonant letters along with vowel signs.

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 in Volume 1 of these two baseline reports [Neupane et al. 2021]), 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 = $19.901 + 0.911 \times ORF$

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

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

- Average grade 3 CB-EGRA percentage score = $28.149 + 7.674 \times$ 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 8.3% of grade 2 children and 13.4% of grade 3 children participating in the home- and community-based schooling intervention in Province 2 met the GON's current national benchmark for reading fluency (45 cwpm with 80% comprehension). Furthermore, 18.2% of grade 2 and 17.2% of grade 3 students fell into the emergent reader category.[3]

These EGRP II baseline findings for the home- and community-based schooling intervention 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, these Province 2 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.

---

[3] The emergent reader category (15 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 for the interventions in Province 2 discusses the different categories of readers and not just the fluent-reader category.

1 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 the 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 other regionally based technical staff, 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 *palika*[4] 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.

In addition to overall activities carried out across all target areas, EGRP II is implementing a home- and community-based schooling intervention under Objective 4, COVID-19 response, in selected disadvantaged palikas in Province 2. The purpose of these efforts is to enable students to catch up on learning that has been disrupted since March 2020 due to the COVID-19 pandemic. Through the home- and community-based schooling intervention, EGRP II provided initial orientation to district and palika-level government officials, followed by

---

[4] 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.

training of grades 1–3 teachers from the target schools on how to set up and run small learning clusters with their students. EGRP II also provided teaching and learning materials (tablets with preloaded content, stationery, and decodable readers) to support the quality of early grade reading instruction in line with curriculum expectations. Small grants to the schools also supported activity implementation and monitoring. Since it started in January 2021, the EGRP II COVID-19 response intervention has been implemented in 219 schools within 16 palikas across the eight districts of Province 2, with plans to scale up to an additional 16 palikas and approximately 264 new schools in the same province from August 2021 to February 2022.

To learn what impacts the home- and community-based schooling intervention may have, EGRP II conducted a baseline study in March 2021 focused on the initial set of 16 palikas, and will undertake an endline in 2022. Although EGRP II overall started in June 2020, we timed the baseline study to align with the end of the academic year, February–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.[5] 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.

---

[5] The mini-EGRA and the equating process were carried out on a sample that was different from but related to the sample described in this baseline report. Further details are provided in Volume 1 (see Neupane et al. 2021).

2   Study Design

2.1   Research Questions

The EGRP II baseline study of the COVID-19-related intervention in Province 2 was designed to answer a specific set of questions.

1. How do grade 2 and 3 students from the home- and community-based schooling intervention districts perform in reading skills?
2. In what ways do those levels of reading performance differ for boys and girls?
3. What is the value added of the COVID-19 response component above EGRP II's support for the NEGRP minimum package?
4. What are the baseline percentages of emergent and fluent student readers in grades 2 and 3 in the home- and community-based schooling intervention districts?

2.2   Sample Design

As noted above, to date, the home- and community-based schooling intervention has been implemented in eight program districts of Province 2, covering 16 palikas and supporting 219 schools. As such, 13,521 students from grade 2 and 12,519 from grade 3 made up the population for the study. Using a 95% confidence level, 47 schools were sampled at random for the study. The sample is presented in **Table 1.**

Table 1: Study sample size, Province 2 intervention

| District | No. of sampled schools | No. of students assessed with CB-EGRA (baseline) | | | | | |
| | | Grade 2 | | | Grade 3 | | |
		Boys	Girls	Total	Boys	Girls	Total
Dhanusha	11	123	90	213	100	122	222
Rautahat	12	81	147	228	91	148	239
Saptari	11	95	123	218	111	122	233
Sarlahi	13	111	150	261	99	106	205
Total	**47**	**410**	**510**	**920**	**401**	**498**	**899**
Grand total				**1,819**			

From the total of 47 schools, 920 students (boys: 45%; girls: 55%) from grade 2 and 899 students (boys: 45%; girls: 55%) from grade 3 were sampled for the home- and community-based schooling program's baseline study. None of the sampled grade 2 or grade 3 students spoke Nepali as their L1. Of the grade 2 sampled students, learners whose L2 was Nepali had as L1 either Bajjika (53.2%) or Maithili (46.8%). In grade 3, sampled students' L1 was Bajjika (49.2%), Bhojpuri (0.1%), or Maithili (50.7%).

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 calculation was made using historical reading data from the 2020 NARN. *Error! Reference source not found.* 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
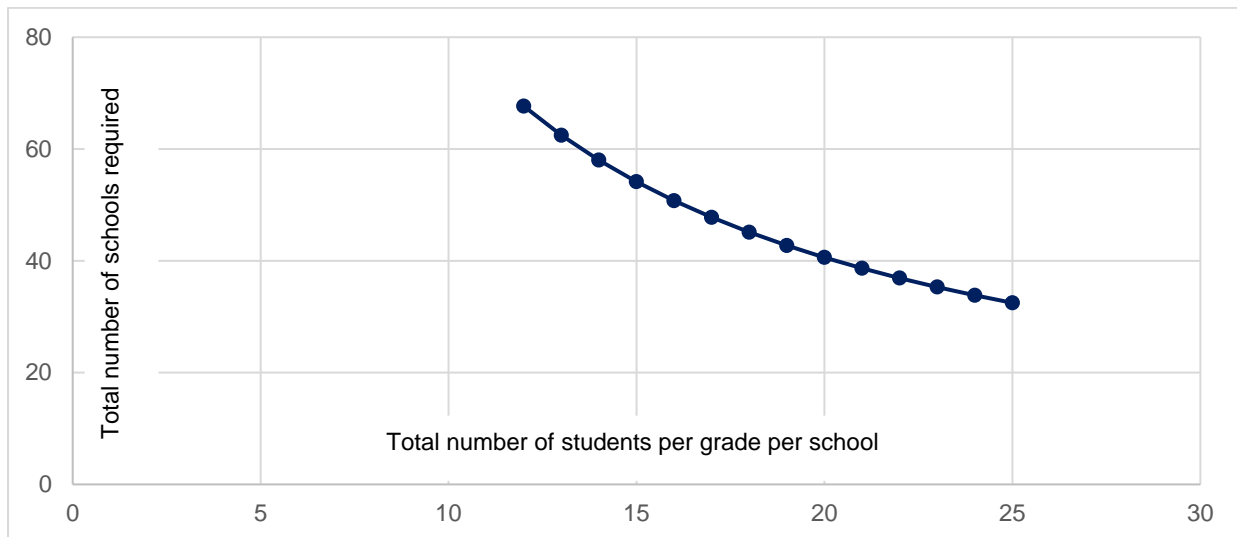
In this assessment, EGRP II used the approach that is being followed by ERO to conduct the CB-EGRA (ERO 2017), which targets 18 students on average as the number of students sampled from each school. Thus, by considering a confidence interval width of ±3.5%, at a 95% confidence level, a standard deviation of 17.98 (taken from NARN 2020 data), and an intra-cluster correlation of 0.36, a design effect of 2.83 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 47 schools for the study. We considered different socio-cultural and geographical attributes when selecting the sample districts and municipalities. Four Province 2 districts were selected such 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 district, 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 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 920 students from grade 2 and 899 students from grade 3. These variations from the ideal sample size, however, did not limit the precision level of the overall study.

Moreover, we statistically equated CB-EGRA scores with the mini-EGRA scores to extrapolate ORF, which cannot be assessed directly with the CB-EGRA instrument. This statistical equating process, including the mini-EGRA administered in a different but related sample, is described in more detail in Volume 1, EGRP II's overall baseline report (Neupane et al. 2021).

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. *Error! Reference source not found.* and *Error! Reference source not found.* provide the details of the tools that were used for grade 2 and 3 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 reading	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 the students' vocabulary knowledge. Students are asked to state the definition, a synonym, and an antonym of each vocabulary word.

Subtask 4: Dictation assesses the writing ability of the children. 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 subtask 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 the 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 that students answer correctly, based on a story of 30 words that the teacher reads aloud two times.

Subtask 6: Reading comprehension measures the number of 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).

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 the logistics that would be required while the trainees were collecting the data. 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 and local level program officers were trained on 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 :		TOTAL NUMBER OF GRADE 3 CHILDREN ADMITTED :	TOTAL NUMBER OF GRADE 2 CHILDREN PRESENT TODAY :	TOTAL NUMBER OF GRADE 3 CHILDREN PRESENT TODAY :
300		250	80	70

▼ **» Status of Baseline**

TOTAL NUMBER OF CHILDREN PARTICIPANTS ON CB-EGRA *

◉ Yes
◯ No

TOTAL NUMBER OF GRADE 2 STUDENT	*	TOTAL NUMBER OF GRADE 3 STUDENT	*
20		20	

TOTAL NUMBER OF CHILDREN PARTICIPANTS ON MINI-EGRA *

◯ Yes
◉ 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**

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
Dhanusha	11	11	0	723	770	261	289	147	162	0	0
Rautahat	12	12	0	761	690	381	346	247	233	0	0
Salrahi	13	13	0	833	674	309	227	264	200	0	0
Saptari	11	11	0	432	410	247	242	216	218	0	0
Total	**47**	**47**	**0**	**2749**	**2544**	**1198**	**1104**	**874**	**813**	**0**	**0**

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.

3   Study Findings

This section presents the overall 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 letter/matra identification, with an average score of 46.2%. This score indicates that students were, on average, able to respond correctly to nearly half of the questions from the letter/matra subtask. In contrast, students had the most difficulty with the dictation subtask. The average percentage score on dictation, 7.1%, signifies those students made errors in over 92% of the items in this subtask. Similarly, in general, students struggled with the reading comprehension subtask. The average percentage score for this subtask was 24.7%, which means that on average, children were able to answer only about one-quarter 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)
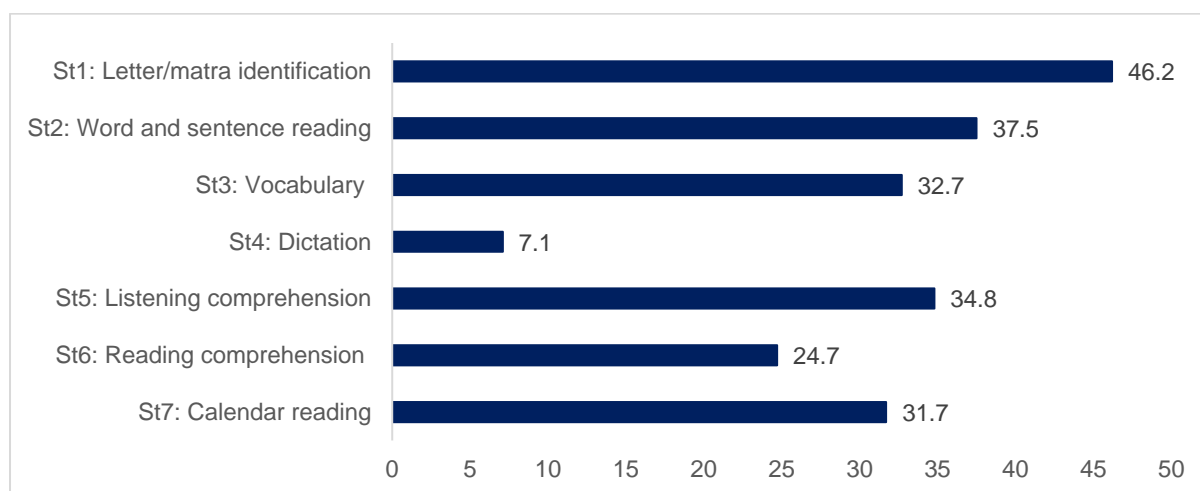
On average, grade 2 students were able to answer 5 out of 21 questions correctly in the assessment. The distribution of the average percentage score is presented in *Figure 4.* In this

---

[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 the overall reading achievement by using weights for all the subtasks.

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)
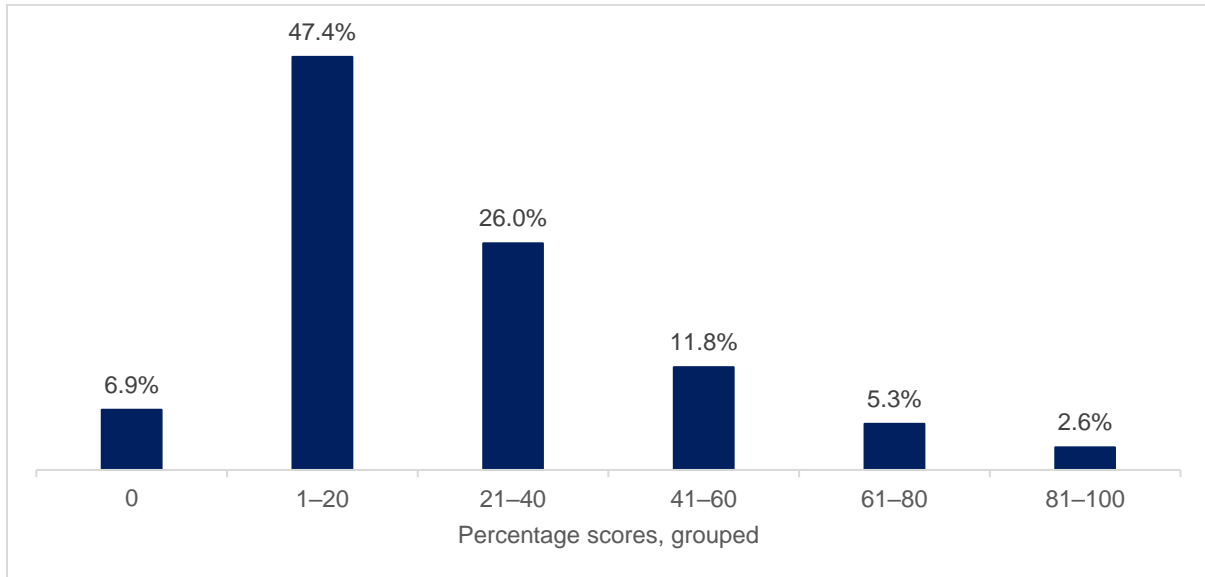
Figure 4 shows that about two-thirds of the total students achieved 40% or less. Very few students (2.6%) could answer 80% or more questions correctly.

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)	इ्या	म्या	फ्या	ज्या	प्या

About half of the students correctly responded to one question or fewer in this subtask. One-fifth (20.3%) 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
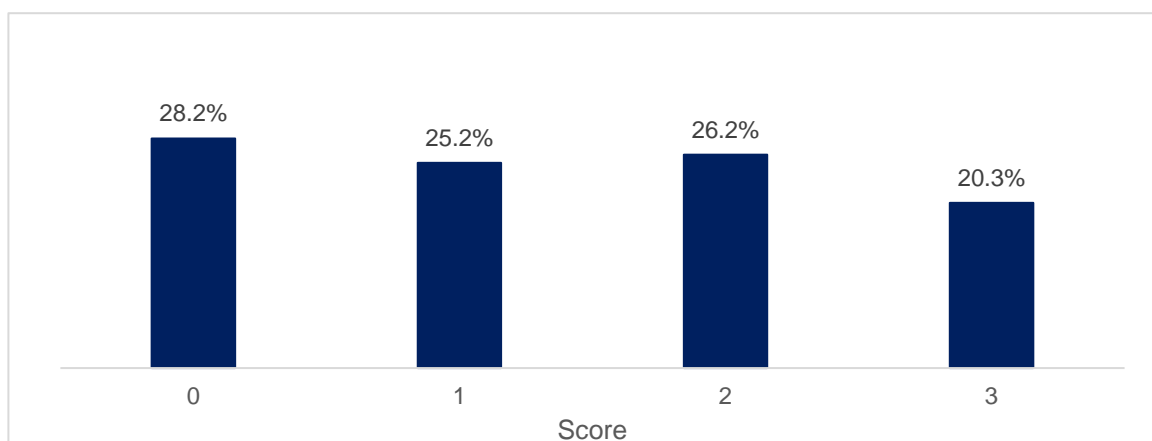
The overall average score on letter/matra identification was 46.2%, which means that out of three items, students were able to answer fewer than two questions on average. The first item focused on identifying a simple vowel or consonant letter. As indicated in *Table 4,* about 53.2% of the students were 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 the 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	53.2%	3.5%
1b	Identify simple matra	53.7%	3.4%
1c	Identify mixed letter/matra	31.7%	3.2%

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* Is 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 14.2% of the students were able to solve all the questions asked, while just over one-third of students could not answer a single question.

Figure 8:   Distribution of scores for the grade 2 word and sentence identification subtask
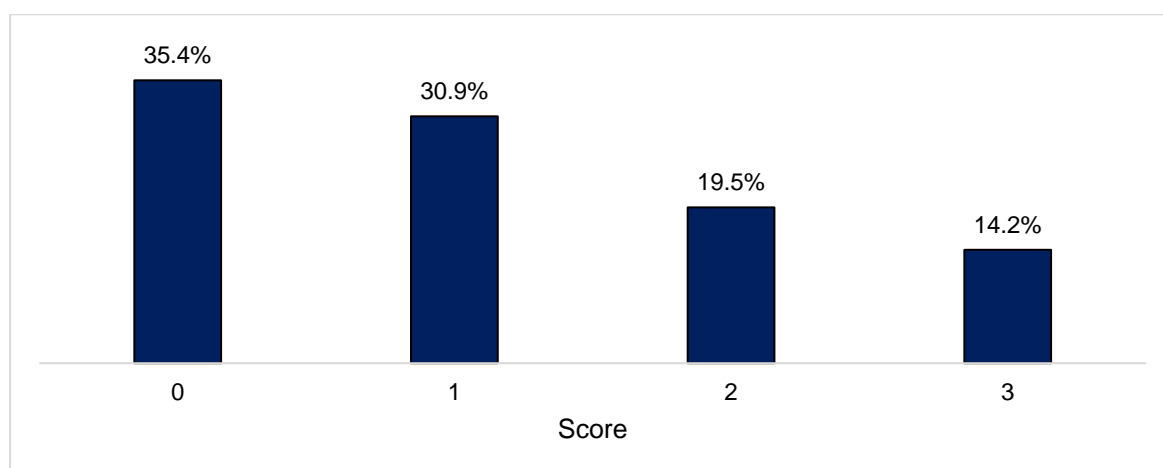
The overall average score on the subtask was 37.5%. This result indicates that, out of three questions, students were able to solve one item correctly on average. Item-level score disaggregation (*Table 5*) indicates that the first and second items were correctly solved by 45.5% and 50.2%. The students found the third question—identification of a four-word sentence—to be comparatively harder. Only one-fourth (26.9%) 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	45.5%	4.0%
2b	Identify three-word sentence	40.2%	2.7%
2c	Identify four-word sentence	26.9%	3.0%

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	कम्ति	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 (29.5%) of the students were able to solve two or more questions. About half (45.5%) of the students could not solve a single question in this subtask.

Figure 10:   Distribution of scores for the grade 2 vocabulary subtask
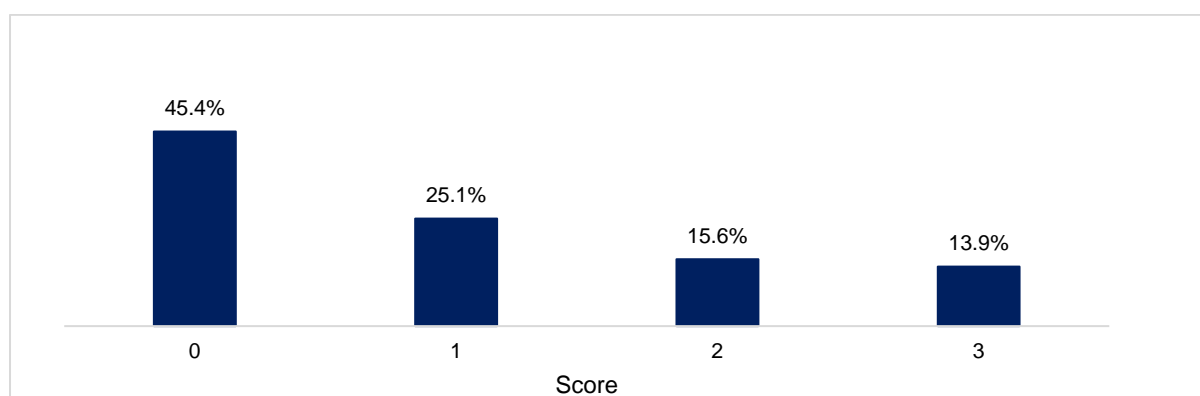
The overall average score for the grade 2 vocabulary subtask was 32.7%. The result indicates that out of three questions, students were able to solve only one question correctly on average. Disaggregation of average scores by item (***Table 6***) indicates that one-third of the students (39.4%) were able to respond correctly to the item related to definition, whereas students struggled to correctly answer the item related to antonyms. Only one-fourth of the students were able to answer the question related to 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	39.4%	3.3%
3b	Antonyms	27.8%	3.3%
3c	Synonyms	30.8%	3.3%

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 जान्छौँ ।
(ग) वनमा बाघले बाख्रालाई तर्सायो ।

(क) .... ।

(ख) .... ।

(ग) .... ।

The average percentage score for this subtask was 7.9%. 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 (2.6%) were able to complete all three items without any errors.

Figure 12: Distribution of scores for the grade 2 dictation subtask
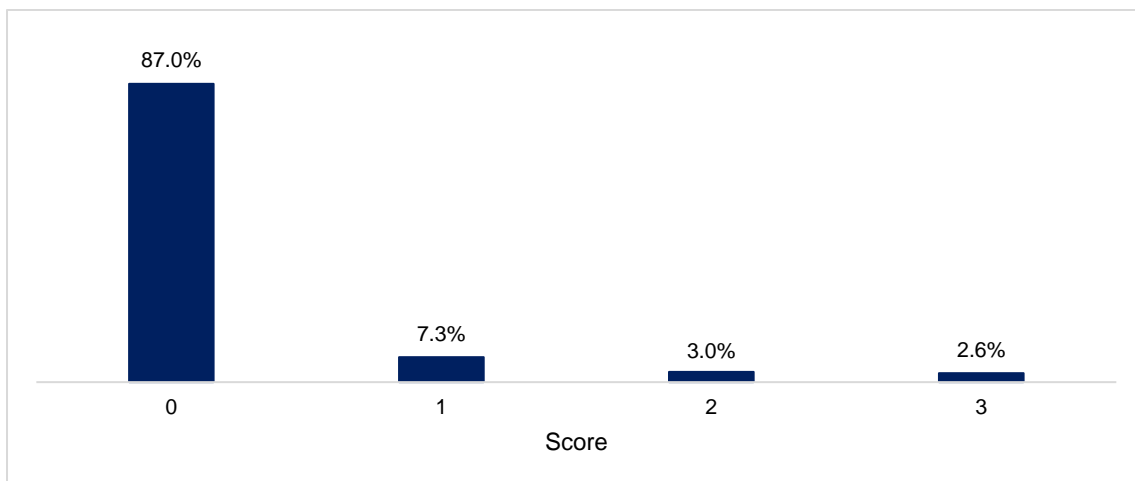
Table 7 displays the average percentage scores of the students on different items in the subtask. For the first question, related to dictating a three-word sentence, the average score was 10.7%; on the second question, the average was 5.7%. For the third question—dictation of four-word sentences with a higher difficulty of words—the average score was only 4.9%.

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	10.7%	2.3%
4b	Four-word sentence	5.7%	1.2%
4c	Four-word sentence, difficult words	4.9%	1.0%

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 directly found 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 of 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 फर्कँदै थिए । अचानक पानी पर्न थाल्यो । कमलसँग छाता थिएन । उनले बारीमा कर्कलाको बोट देखे । उनले कर्कलाको पात टिपेर ओढे । उनी भिज्नेबाट सजिलै जोगिए ।

क.   कमल कहाँबाट फर्कँदै 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 15.1% of the students were able to solve all three questions from this subtask, while more than 40% could not answer any of the questions.

Figure 14: **Distribution of scores for the grade 2 listening comprehension subtask (25 words)**
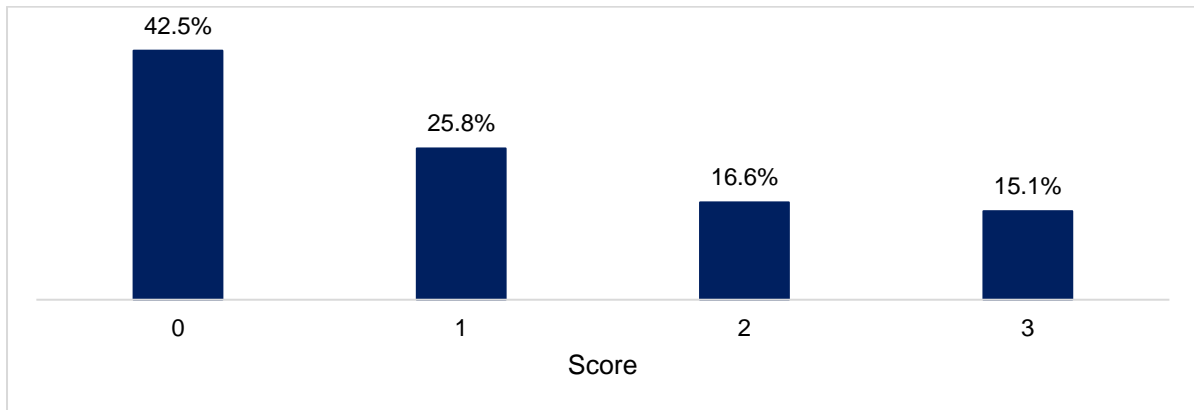
The overall average score for Subtask 5 was 34.7%. Looking at the disaggregated results for each item in the subtask (***Table 8***), students were found nearly equally competent to answer the questions that came directly from the text (32.6% and 37.1% of students were able to answer the first and second items respectively) or inferential type of questions (the third item, with 34.6% of students able to answer correctly). Thus, around one-third 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	32.6%	3.4%
5b	Short answer, explicit	37.1%	3.6%
5c	Inferential from at least two sentences	34.6%	3.3%

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 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 हुर्किंदै 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 6.8%.

Figure 16: Distribution of scores for the grade 2 reading comprehension subtask (60 words)
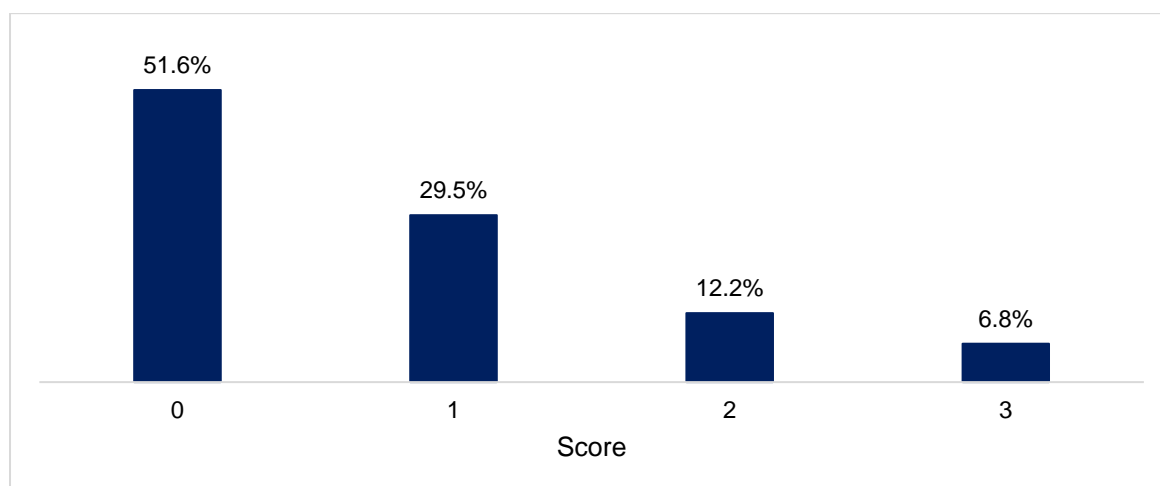
The average percentage score for this subtask was 24.7%. This result indicates that the average correct answer per student was less than one question out of three. In general, students found the second item in the subtask slightly more difficult than the first and third items (*Table 9*).

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	25.1%	3.1%
6b	Short answer, explicit	23.9%	2.9%
6c	Inferential from two or more sentences	25.1%	2.7%

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 31.7%. This result shows that students were able to answer 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 (सम्बन्धितलाई बिदा)	दशमी २ 16	एकादशी ३ 17 पुत्रदा 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 षटतिला एकादशी	द्वादशी १८ 01	त्रयोदशी १९ 02
चतुर्दशी २० 03	औषी २१ 04	प्रतिपदा २२ 05 सोनाम ल्होसार सम्बन्धितलाई 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 49.1% of the students were not able to answer a single question from this subtask. Very few students (27.9%) 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**
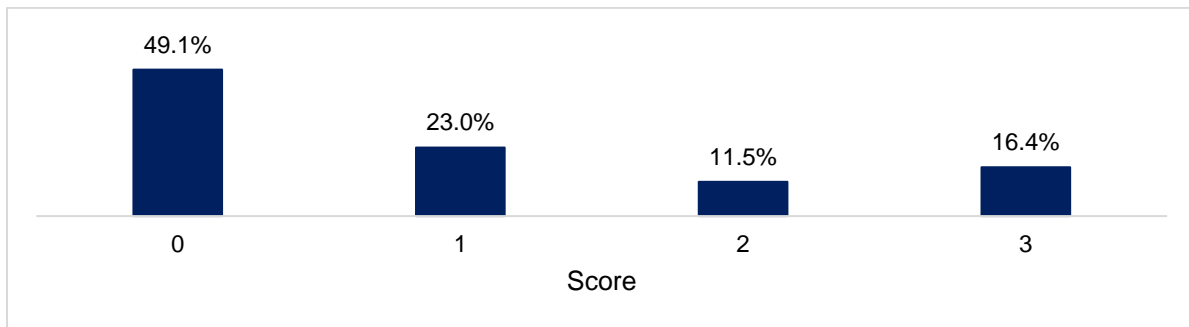
As shown in *Table 10*, the difficulty level of all three questions in this subtask was nearly the same. The first question was answered correctly by 27.1% of students, while 32.4% and 35.7% of students were able to answer the second and third questions correctly, respectively.

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	27.1%	3.7%
7b	Festival and date	32.4%	3.5%
7c	Number of specific days (e.g., Saturdays) in a month	35.7%	3.8%

3.1.3 Reading Achievement by Sex (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 less 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.) Similar results held for all subtasks: the achievement difference was not significantly different in any case (*Table 11*).

Figure 19. **Average percentage scores of grade 2 students, by sex**
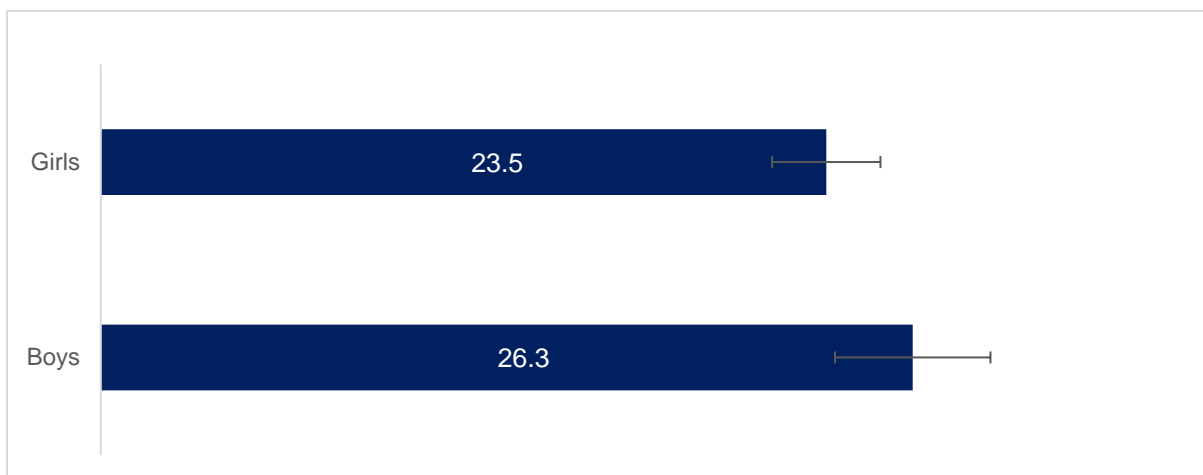
Table 11: Average percentage scores of grade 2 students, by subtask and sex

Subtask	Average percent score (boys)	Average percent score (girls)	Difference (boys – girls)
St1: Letter/matra identification	48.9	44.1	4.8
St2: Word and sentence identification	40.0	35.5	4.5
St3: Vocabulary	34.3	31.4	2.9
St4: Dictation	7.9	6.5	1.4
St5: Listening comprehension	36.2	33.7	2.5
St6: Reading comprehension	26.5	23.3	3.2
St7: Calendar reading	33.7	30.2	3.5

3.2 Grade 3 Findings

3.2.1 Overall Reading Achievement (Grade 3)

Similar to grade 2, students in grade 3 performed highest on word and sentence identification, with an average score of 42.2%, indicating that students were able to solve about half of the questions from the word and sentence identification 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 13.5% and for word separation was 21.6%. Similarly, students struggled with vocabulary and calendar reading. The average scores for these subtasks were 32.4% and 33% respectively, which indicates that the students answered only one-third of the questions correctly, on average. Details on the average scores of grade 3 students on each subtask are presented in *Figure 20.*

Figure 20: Average percentage scores, by subtask (grade 3)
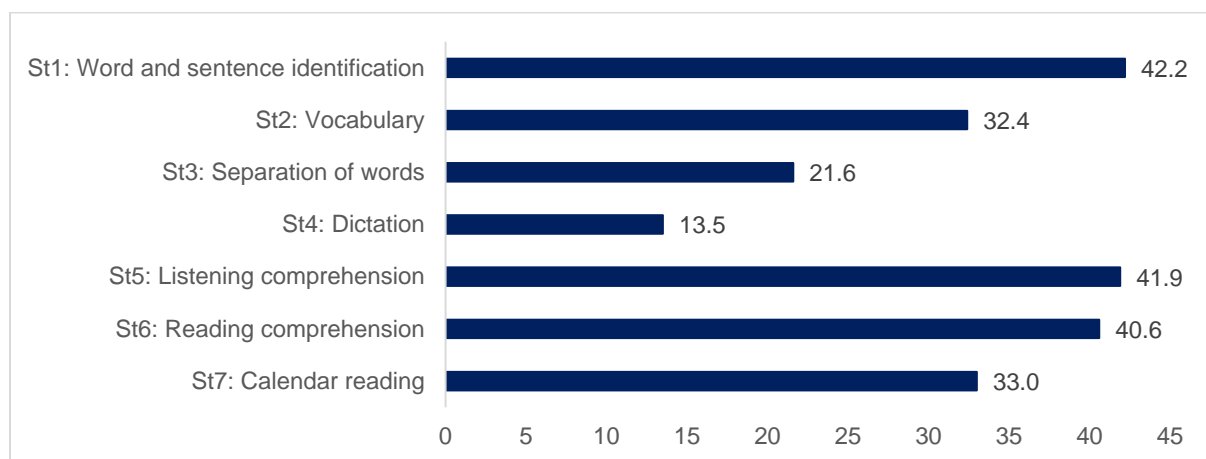
On average, grade 3 students were able to answer 6 out of 21 questions correctly in the assessment. The distribution of the average percentage scores is presented in *Figure 21,* with the percentage scores categorized into five groups. This figure shows that about three-fourths of the total students achieved 40% or less. There were very few students (5.2%) who were able to answer 81% or more questions correctly.

Figure 21:   Distribution of overall average percentage scores (grade 3)
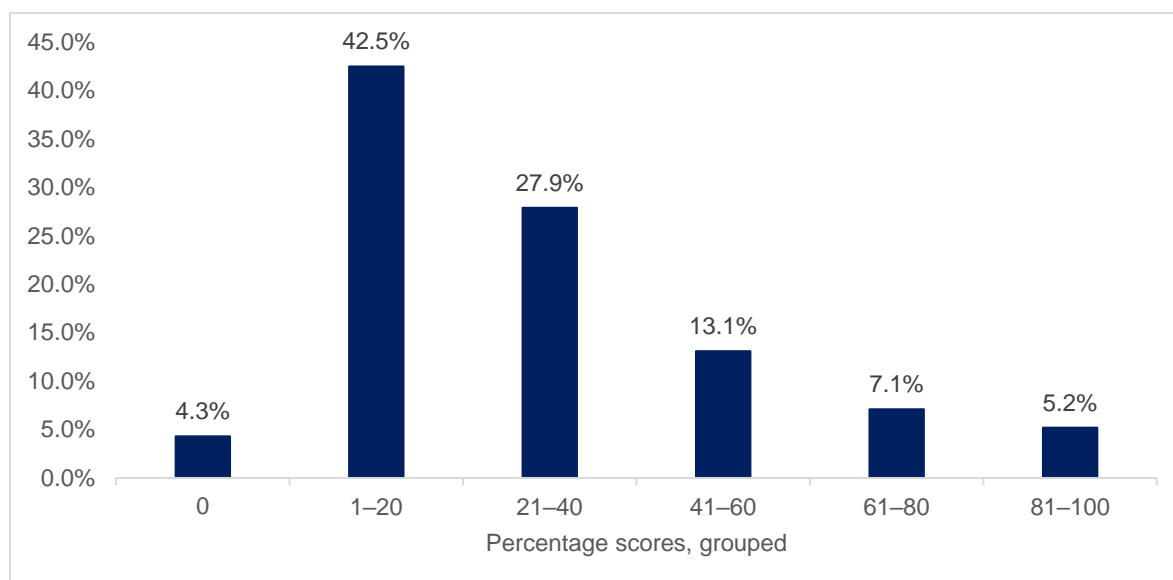
3.2.2 Reading Achievement by Subtask (Grade 3)

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 22.*

Figure 22:   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 23*) also indicates that about 60% of students were able to respond correctly to zero items or to only one item in the subtask. Only about one-fifth (18.5%) of the students were able to solve all of the questions in the subtask.

Figure 23:   Distribution of scores for the grade 3 word and sentence identification subtask
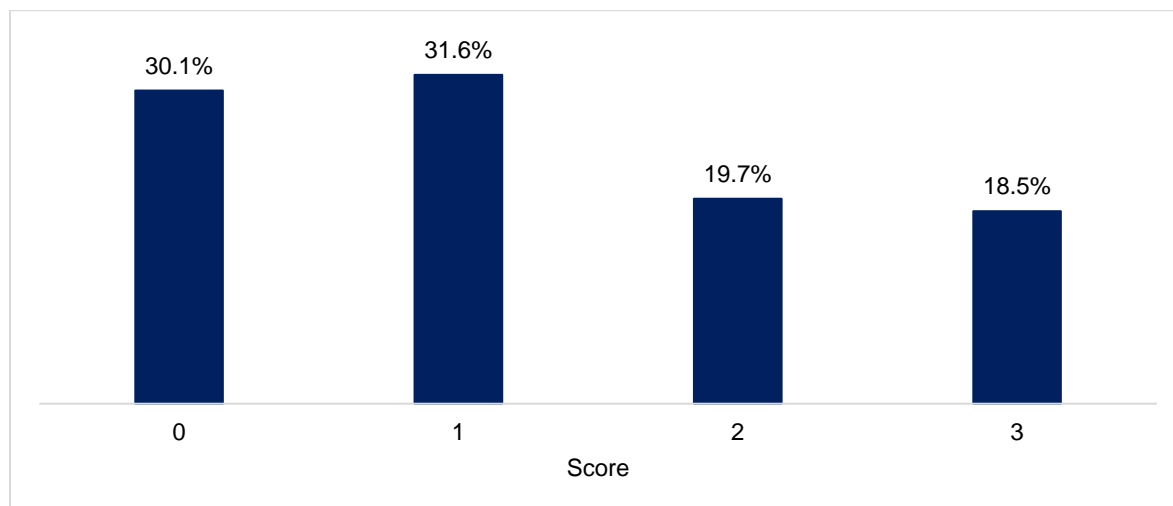
The overall average score on the word and sentence identification subtask was 42.2%. The result indicates that out of three questions, students were able to solve one item correctly on average. Item-level disaggregation (*Table 12*) indicates that the first item was comparatively easy, as it was correctly solved by 48.8% of students. It is surprising that more students were able to identify a five-word sentence (41.5%) than a four-word sentence (36.3%). This result might have been due to the choices of distractors for the questions.

Table 12: 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	48.8%	3.0%
1b	Identify four-word sentence	36.3%	2.9%
1c	Identify five-word sentence	41.5%	2.8%

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 24.*

Figure 24:   Student stimulus for the grade 3 vocabulary subtask

उपकार्य २

शब्दभण्डार

(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 25*) shows that 39.8% of the students were unable to respond correctly to any items. Another 32.5% of students solved only one question in this subtask. This result shows that many students struggled with vocabulary.

Figure 25:   Distribution of scores for the grade 3 vocabulary subtask
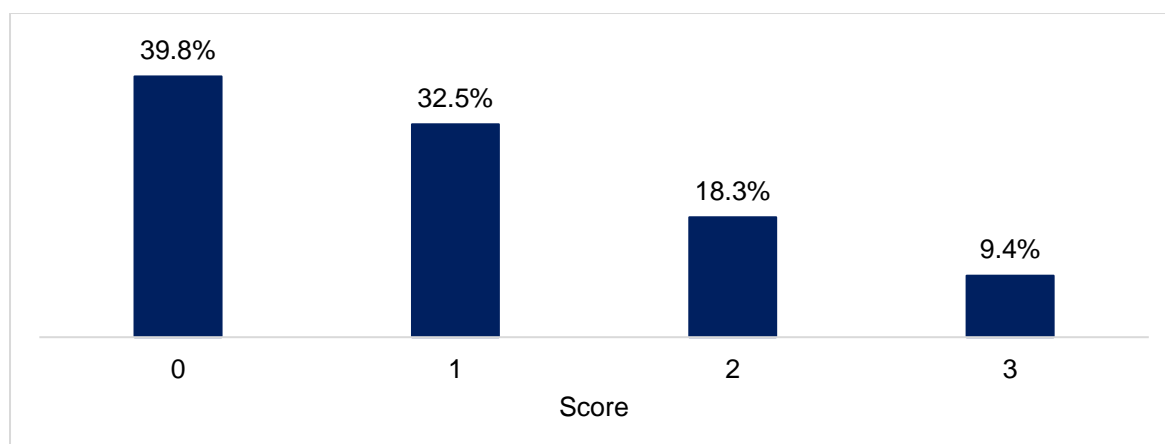
The overall average percentage on the vocabulary subtask was 32.4%. The disaggregated average scores by subtask (*Table 13*) indicate that fewer than one-third of the students (29.1%) were able to respond correctly to items related to defining a word (29.1%) or choosing a synonym (29.8%). The average percentage score for students who responded correctly to antonym-related items was slightly higher, at 38.3%.

Table 13: Average item scores for the grade 3 vocabulary subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
2a	Define a word	29.1%	3.0%
2b	Antonyms	38.3%	4.0%
2c	Synonyms	29.8%	3.3%

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 26.*

Figure 26:   Student stimulus for the grade 3 word separation subtask

उपकार्य ३

जोडिएका शब्द छुट्याउने

(क) पोखरीठुलोछ ।
....। |

(ख) बारीमागोलभैंडापाकेकोछ ।
.... |

(ग) दराजमाकिताबमिलाएरराखिएकाछन् ।
.... |

Figure 27 indicates that two-thirds (65.1%) of students were not able to solve a single question correctly. About four-fifths (81.1%) of the students answered zero or one question from this subtask correctly.

Figure 27: Distribution of scores for the grade 3 word separation subtask
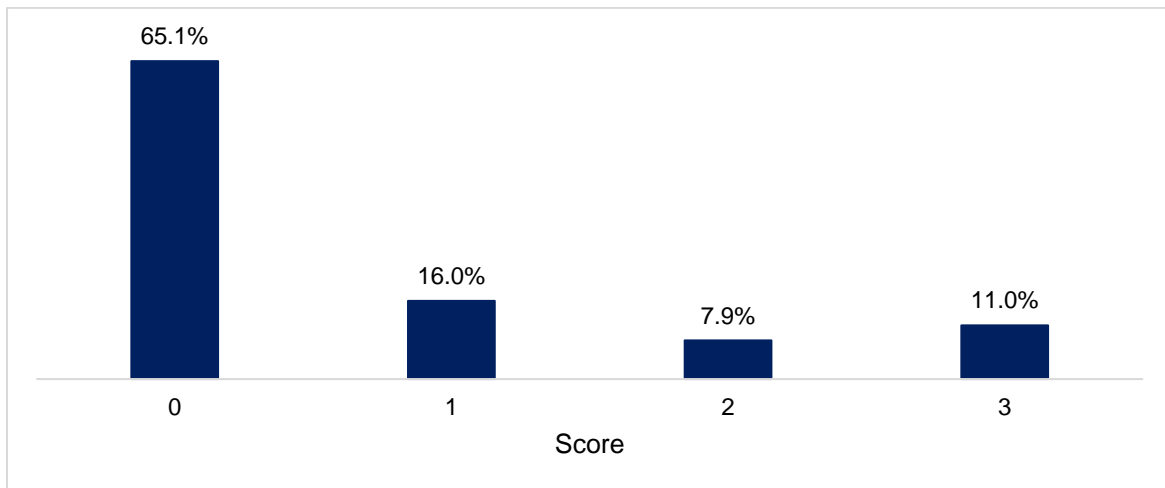
The overall average score on the word separation subtask was 21.6%. This result indicates that out of three questions, students were able to solve less than one question correctly on average. The disaggregated scores by subtask (***Table 14***) indicate that the first question was easier than the other two items. The first item—to separate a three-word sentence—was correctly solved by 29.5% of students, whereas the second and third items—separation of four-word and five-word sentences—were solved by only 17.6% and 17.8% of total students, respectively.

Table 14: Average item scores for the grade 3 word separation subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
3a	Three-word sentence	29.5%	3.0%
3b	Four-word sentence	17.6%	3.4%
3c	Five-word sentence	17.8%	2.9%

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 28*** is the student stimulus for dictation.

Figure 28:   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 13.5%, or less than one item out of three completed without errors. As indicated in *Figure 29,* a very minimal percentage of students (5.1%) of students was able to respond correctly to all three items on the dictation subtask.

Figure 29:   Distribution of scores for the grade 3 dictation subtask
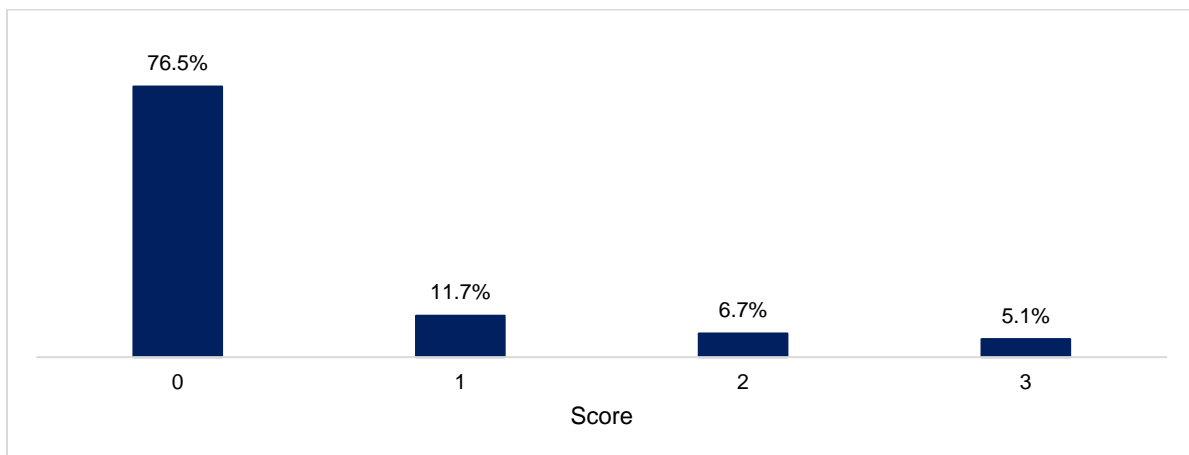
Table 15 also presents the average percentage of students who could correctly answer the different items in the subtask. The first question was dictation of a three-word sentence, and 19.9% of students were able to respond correctly. The second and third questions were completed correctly by lower proportions of students. Only 13.2% were able to correctly complete the second item, and the third item was most difficult of all, with only 7.4% of students able to complete this item correctly.

Table 15: Average item scores for the grade 3 dictation subtask

Subtask	Description	Percentage of students who answered correctly	Standard error
4a	Three-word sentence	19.9%	2.5%
4b	Four-word sentence	13.2%	2.1%
4c	Five-word sentence	7.4%	1.7%

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 30.*

Figure 30:   Student stimulus for the grade 3 listening comprehension subtask

उपकार्य ५

श्रुतिबोध

प्रज्ञा र बिशाखा विद्यालय जाँदै थिए। उनीहरूले बाटोमा झोला देखे। प्रज्ञाले हतपत झोला टिप्न खोजिन्। बिशाखाले प्रज्ञालाई झोला 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 झोला टिप्न दिइनन् ?

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

The analysis of each item in the subtask (*Figure 31*) identified that 21.6% of the students were able to solve all three questions from this subtask, while just over one-third (35.2%) could not answer any of the questions.

Figure 31:   Distribution of scores for the grade 3 listening comprehension subtask (30 words)
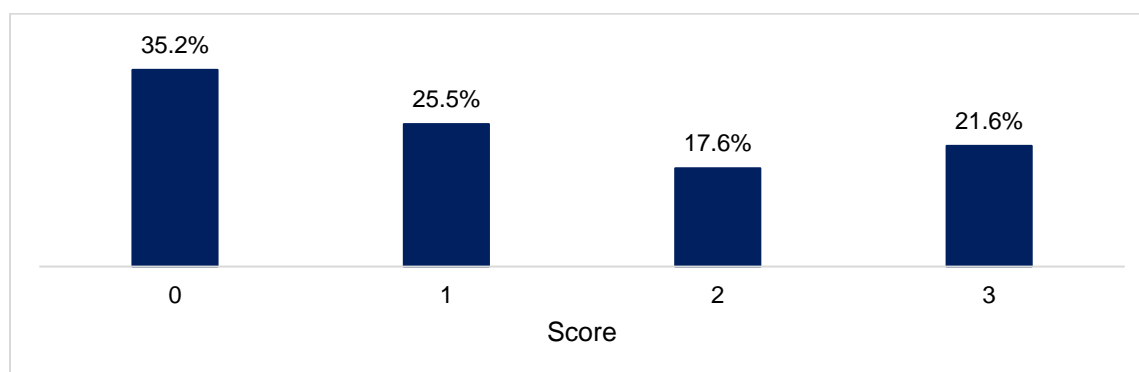
The overall average score for the listening comprehension subtask was 41.9%. The result shows that students performed better on the listening comprehension task than on the other tasks. Looking at disaggregated results for the items in the subtask (***Table 16***), slightly fewer than half (46.5%) of total students were able to solve the first question, whereas the second question was solved by 41.4%. At the same time, only 37.7% of students were able to solve the third, inferential question.

Table 16: 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	46.5%	3.2%
5b	Short answer, explicit	41.4%	3.1%
5c	Inferential from at least two sentences	37.7%	3.4%

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 32*** is the student stimulus for reading comprehension.

Figure 32: **Student stimulus for the grade 3 reading comprehension subtask (60 words)**

उपकार्य ६

पठनबोध

लाक्पाको 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 फँडानी 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) लाक्पाको 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 40.6%. This result indicates that the average correct answer per student was slightly more than one question out of three. The distribution in *Figure 33* shows that over one-third (35.3%) 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 18.8%.

Figure 33: **Distribution of scores for the grade 3 reading comprehension subtask (60 words)**
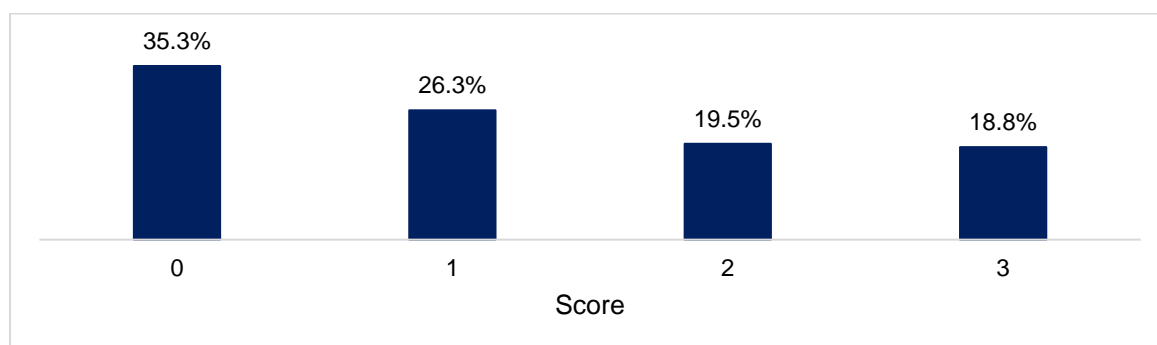
Looking at the average scores on each of the items within the subtask (*Table 17*), less than half of the students were able to answer the first and second questions (42.6% and 45.8% respectively), whereas the third, inferential question was solved by only one-third (33.6%) of

the students. This finding shows that students had more difficulty with the higher level (inferential) comprehension question.

Table 17: 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	42.6%	4.1%
6b	Short answer, explicit	45.8%	3.9%
6c	Inferential from two or more sentences	33.6%	3.5%

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 34* contains the student stimulus for calendar reading.

Figure 34: Student stimulus for the grade 3 calendar reading subtask

उपकार्य ७

पात्रो (क्यालेन्डर) पठन

(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 35 shows that 40.6% of the students were not able to answer a single question from this subtask. Very few students (11.0%) were able to answer all three questions.

Figure 35: Distribution of scores for the grade 3 calendar reading subtask
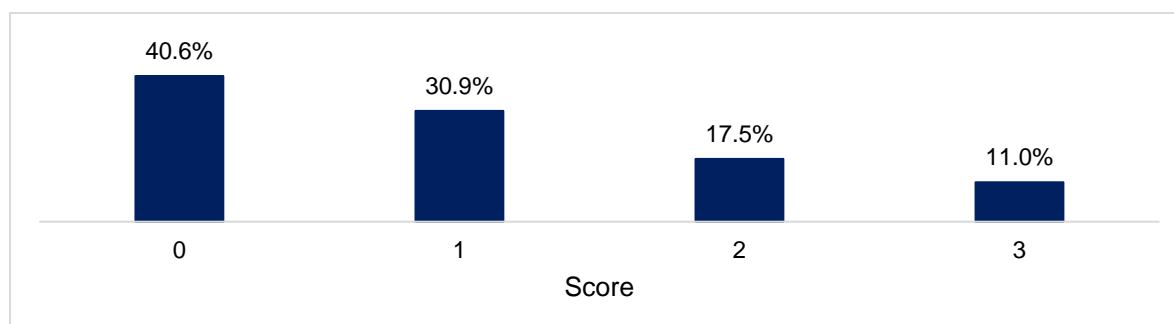
The overall average percentage for the subtask was 33.0%. This result shows that students were, on average, able to answer one question out of three correctly in this subtask. The analysis of average scores on each of the items in the subtasks (***Table 18***) indicates that, as with the grade 2 results, 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 18: 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	42.1%	4.0%
7b	Festival and date	26.8%	3.3%
7c	Last day of previous month	30.1%	3.3%

3.2.3 Reading Achievement by Sex (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 36,*** the average percentage score for girls was slightly less than for boys. The difference, however, was not statistically significant.

Figure 36:   Average percentage scores of grade 3 students, by sex
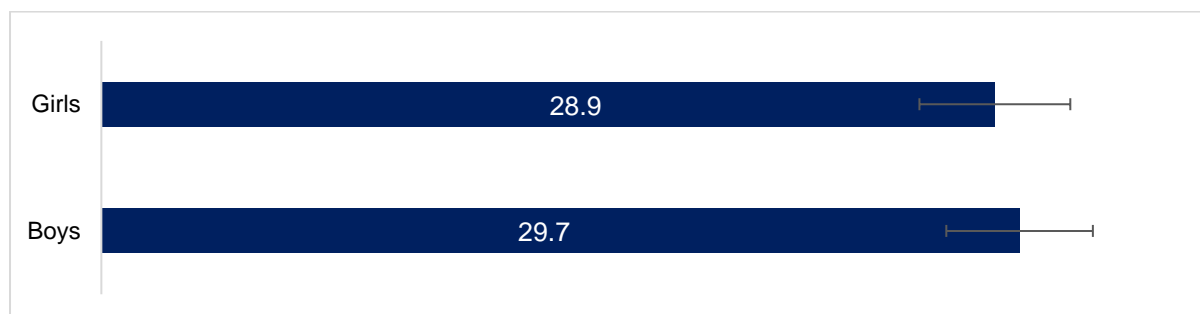
A similar result was found throughout the subtasks in the assessment, with no statistically significant achievement differences for any subtasks (***Table 19***).

Table 19: 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	43.4	41.2	2.2
St2: Vocabulary	33.4	31.6	1.8
St3: Separation of words	22.6	20.8	1.8
St4: Dictation	12.4	14.3	−1.9
St5: Listening comprehension	43.0	40.9	2.1
St6: Reading comprehension	41.8	39.7	2.1
St7: Calendar reading	32.5	33.4	−0.9

4   Extrapolation of Reading Achievement

As described in the Background section, alongside the CB-EGRA, EGRP II administered a mini-EGRA to a subset of the sampled children in order to allow for statistical equating and derivation of ORF rates from the CB-EGRA results. Program analysts used both sets of data to develop a set of simple linear regression models for this purpose.[7] The COVID-19-related baseline study in Province 2 adopted the same set of models to measure the reading fluency and reading comprehension for this more limited study population:

Average CB-EGRA percentage score (grade 2) = 19.901 + 0.911 × ORF

Average CB-EGRA percentage score (grade 2) = 24.003 + 9.201 × Average comprehension

Average CB-EGRA percentage score (grade 3) = 22.399 + 0.817 × ORF

Average CB-EGRA percentage score (grade 3) = 28.149 + 7.674 × average comprehension

4.1.1 Equivalent Scoring

Using the models above, we created equivalent CB-EGRA scores for emergent and fluent reader benchmarks (***Table 20***). 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. These scores will also become benchmark equivalencies for all future CB-EGRAs.

Table 20: 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

---

[7] For full details of the equating process, refer to *Baseline Report Vol. 1: Student Reading Performance in the Early Grades* (Neupane et al., 2021).

different for grades 2 and 3 and the results are, therefore, not directly comparable between the grades.

4.1.2 Students Who Met the Reading Benchmark

The Government of Nepal has set 45 words and 80% comprehension as Nepal's national reading benchmark (MOE 2017). The values extrapolated from the CB-EGRA results in this baseline evaluation for the Province 2 intervention were further analyzed to identify the percentage of students who met the reading benchmark. As shown in *Table 21,* 8.3% of grade 2 children and 13.4% of grade 3 children in the EGRP II baseline sample for the home- and community-based schooling intervention met the reading benchmark.

Table 21: Percentage of learners from the home- and community-based schooling intervention who met the reading benchmark (45 words and 80% comprehension), by grade

Grade	Percentage of learners who met the benchmark	Standard error
2	8.3%	2.0
3	13.4%	2.7

4.1.3 Comparing EGRP II Baseline Findings (Province 2 Intervention) with the 2020 NARN, 2020 EGRP endline, 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, 2020 EGRP endline, 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, EGRP II's 2021 baseline, the 2020 NARN, and the 2020 EGRP endline all 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 13.4%, with 95% confidence that the true population percentage was between 8.1% and 18.7%. 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%. Comparing with the EGRP endline, we note that 9.4% of grade 3 students met the reading benchmark with a 95% confidence interval of between 7.1% and 11.8%.

Figure` 37 demonstrates that the confidence intervals (the black lines at the end of each blue bar) for the EGRP II baseline, the 2020 NARN, and the 2020 EGRP baseline estimates overlap. Therefore, while the estimates have a difference of between 4.0 and 5.0 percentage points, we cannot be certain that the EGRP II baseline percentage is higher than the NARN or the EGRP endline 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` 37:   Percentage of grade 3 students who met the reading benchmark in the 2021 EGRP II baseline (Province 2 intervention), 2020 NARN, and EGRP endline
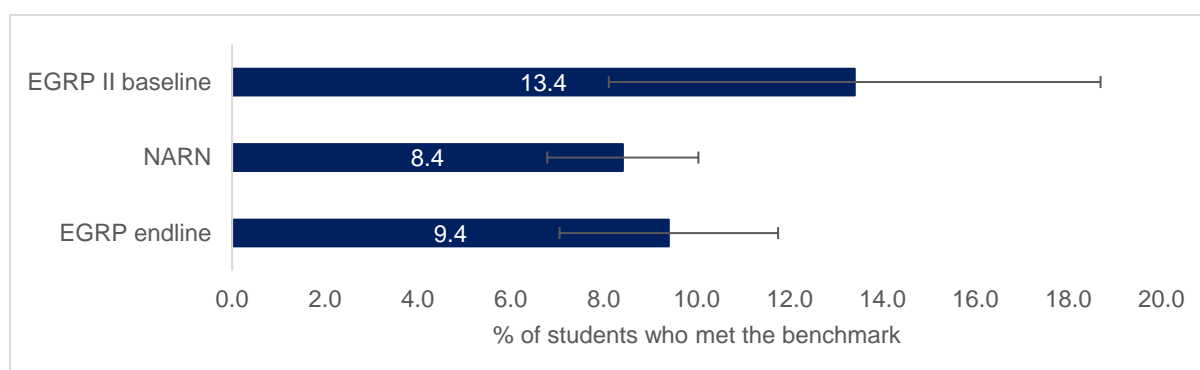
Another important caveat to keep in mind is that the NARN, EGRP endline, and the EGRP II baseline were administered to different samples of schools and children. The assessments were also conducted in different years: the NARN and EGRP endline in early 2020, before the COVID-19 pandemic; and the EGRP 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 24.8% in the EGRP II baseline. Similarly, average grade 3 CB-EGRA scores ranged from 66% to 68% in the past, with an average of 29.3% in the EGRP II baseline. The differences are captured in *Figure 38* below.

Figure 38:   Comparison of average CB-EGRA scores between previous CB-EGRA assessments and the EGRP II baseline (Province 2 intervention)

CB-EGRAs from previous years:	2021 EGRP II baseline:
Grade 2: 64%–66% Grade 3: 66%–68%	Grade 2: 24.8% Grade 3: 29.3%
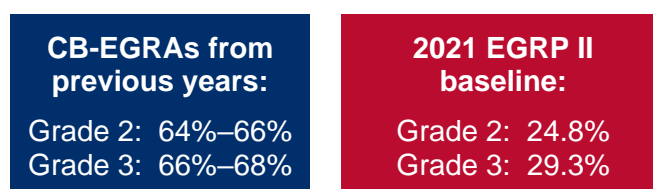
These differences in average outcomes on the CB-EGRA could be due to factors such as learning disruptions caused by the COVID-19 pandemic during the 2020–2021 academic year, as well as differences in the samples for the various assessments. At the same time, the drop could also be due in part to how 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.

4.1.4 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 22*** provides only the percentage of students categorized as emergent or fluent readers.

Table 22: Categories of readers, by grade

Grade	Emergent reader	Fluent reader
2	18.2% (2.3)	8.3% (2.0)
3	17.2% (1.8)	13.4% (2.7)

Note. The numbers in parentheses represent standard errors.

4.1.5 Measuring the Value-Added Impact of the Program Response to COVID-19

This section of the report addresses the EGRP II approach to respond to research question 3: What is the value added of the COVID-19 response component above the EGRP II support for the NEGRP minimum package?

To measure the value added of the COVID-19 response, we will be able to compare the baseline and endline results for the region receiving the COVID-19 response programming to the baseline and endline results for the rest of the EGRP II programming area (covered in Volume 1 of these baseline reports). This quasi-experimental design (QED) will compare the gains between two timepoints of the two program regions (***Figure 39***).

Figure 39:   Quasi-experimental design to measure the impact of COVID-19 response
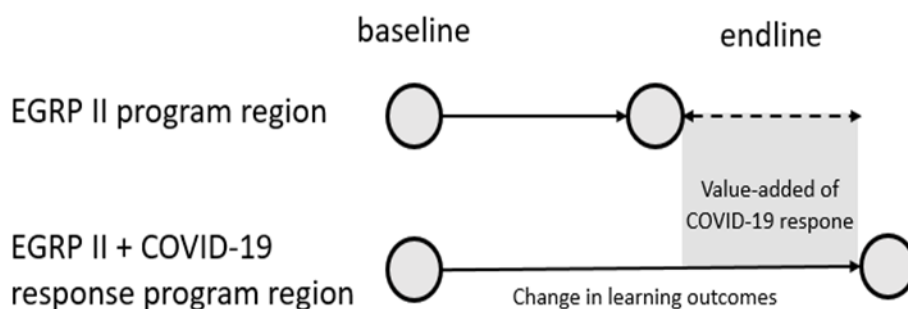
As *Figure 39* illustrates, any value-added gain of the COVID-19 intervention over the EGRP II programming will be attributable to the COVID-19 intervention.

Critical for a QED is to establish equivalence at baseline. This process assesses the difference in baseline estimates between two programs to determine whether any differences are small enough that we are confident we are comparing "apples to apples." If one program has a baseline average much higher or lower than the other program, for example, it is possible that the gains could be attributed instead to different baseline advantages or disadvantages. To determine whether we could rule out that possibility, we assessed the differences in the baseline estimates using a baseline balance test. We calculated the difference between the two baseline estimates in terms of effect size in standard deviations (Cohen's *d*) and applied this value to the metric presented in *Table 23.*

Table 23: Potential effect size differences at baseline

0 ≤ difference ≤ 0.05	0.05 ≤ difference ≤ 0.25	difference > 0.25
Satisfies the baseline equivalence requirement	Requires statistical adjustment to satisfy the baseline equivalence requirement	Does not satisfy the baseline equivalence requirement

Source: Institute of Educational Sciences (2020).

The calculation of the difference is:

$$\frac{difference\ between\ baseline\ estimates}{pooled\ standard\ deviation}$$

Table 23 shows that an effect size difference of less than 0.05 is optimal, and a difference between 0.05 and 0.25 indicates the need for a statistical adjustment (such as adding control covariates into a difference regression model).

Applying this process to the percentage of students achieving reading with fluency across both baseline evaluations, we found small effect size differences of 0.03 and 0.02 standard deviations for grades 2 and 3, respectively (far-right column of *Table 24*).

Table 24: Comparison of effect size differences between full baseline and COVID-19 response baseline

Grade	Location	Percentage of fluent readers	Standard deviation	Sample size (students)	Pooled standard deviation	Estimated difference	Difference / pooled standard deviation
2	EGRP II baseline region	7.40%	26.20%	826	26.9%	−0.90%	0.03
	COVID-response component region (Province 2)	8.30%	27.50%	920			
3	EGRP II baseline region	12.60%	33.20%	752	33.7%	−0.80%	0.02
	COVID-response component region (Province 2)	13.40%	34.10%	899			

Because these effect size differences are less than 0.05, we can conclude that we have two comparable groups and can assess the value-added impact of the COVID-19 program response at endline.

5   Summary and Conclusions

The study described in this analysis report was intended to establish a baseline for the EGRP II home- and community-based schooling intervention by assessing students' reading performance. For the study, a scientific sampling technique was used to select 47 schools from four districts from Province 2. 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 a subsample of 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 helpful 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 four research questions were asked in this baseline study. The summary and conclusion of the study are presented as responses to each research question.

Research Question 1: How do grade 2 and 3 students from the home- and community-based schooling intervention districts perform in reading skills?

The overall reading achievement measure using the average percentage CB-EGRA score for grade 2 was 24.8%, and for grade 3 it was 29.3%. For both grades, a total of seven subtasks and 21 items were used to assess the students' reading ability. Thus, this finding means that on average, a child from grade 2 was able to correctly respond to about five items and a child from grade 3 was able to correctly respond to about seven items. 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 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 letter, word, and sentence identification.

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

The reading performance was not significantly different between boys and girls in this baseline study. The average CB-EGRA performance of grade 2 boys was 26.3% and that of girls was 23.5%. Although there was a difference of 2.8 percentage points between girls and boys, this difference was not statistically significant. Similarly, boys from grade 3 scored 29.7% and girls scored 28.9%, a difference of 0.8 percentage points that also was not

statistically significantly different. Thus, it can be concluded there was no significant association between student sex and learning outcomes.

Research Question 3: What is the value added of the COVID-19 response component above the EGRP II support for the NEGRP minimum package?

This QED will compare the gains between two timepoints of the two program intervention regions. This research question will be answered after the two endline evaluations in 2022.

Research Question 4: 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 the 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 less than one-fourth of students in both grades in the study sample were emergent readers (18.2% from grade 2 and 17.2% 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 evaluation for the home- and community-based schooling intervention in Province 2.

Sample Size and Representativeness

The sample for this baseline study was intended to include 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 four selected 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 adopted 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 utilized 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

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

Institute of Educational Sciences. 2020. *What Works Clearinghouse™ Standards Handbook (Version 4.1).* Washington, DC: U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Ministry of Education. 2017. *Reading Benchmark.* Kathmandu: Ministry of Education.

Neupane, Sagar Mani, Swadesh Maharjan, Simon King, Jodie Fonseca, and Wendi Ralaingita. 2021. *USAID's Early Grade Reading Program II (EGRP II) in Nepal. Baseline Report Vol. 1, Student Reading Performance in the Early Grades.* Prepared for USAID under Cooperative Agreement No. 72036720CA00001. Research Triangle Park, North Carolina, USA: RTI International.

Annex A: EGRP II MEL Indicator Reporting

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

IND 01_ES. 1-1: Percent of learners who attain a minimum grade-level proficiency in reading at the end of grade 3 in targeted 16 local governments of Province Two.

Overall: 13.4% (Numerator: 1,117 Denominator: 13,521)

Male: 11.9% (Numerator: 657, Denominator: 5,501)

Female: 14.6% (Numerator: 1,025, Denominator: 7,018)