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

Endline Report: Program Impact on Student Reading Performance

in the Early Grades

A person writing on a piece of paper

Description automatically generated

|  |  |
| --- | --- |
| Logo  Description automatically generated | Logo  Description automatically generatedGON seal  **Government of Nepal,** **Ministry of Education, Science, and Technology** |
|  |  |

**April 30, 2021**

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

**May 2022**

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

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

Endline Report: Program Impact on Student Reading Performance in the Early Grades

Cooperative Agreement No. 72036720CA00001

Prepared for

Jayakrishna Upadhyay

Agreement Officer’s Representative

Education Development Office

United States Agency for International Development/Nepal

GPO Box No. 295

U.S. Embassy, Maharajgunj

Kathmandu, Nepal

Email: jupadhyay@usaid.gov

Prepared by

Swadesh Maharjan, Buddhi Chaudhary, Manoj Bhandari, Simon King, Jodie Fonseca, and Wendi Ralaingita

RTI International

P.O. Box 12194

Research Triangle Park, NC 27709-2194

USA

<http://www.rti.org/>

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

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

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

**Table of Contents**

[List of Tables ii](#_Toc103691227)

[List of Figures iii](#_Toc103691228)

[Acronyms and Abbreviations v](#_Toc103691229)

[Acknowledgments vii](#_Toc103691230)

[Executive Summary 1](#_Toc103691231)

[1 Background 8](#_Toc103691232)

[1.1 EGRP II Program Description 8](#_Toc103691233)

[1.2 Evaluation Context 12](#_Toc103691234)

[2 Study Design 13](#_Toc103691235)

[2.1 Research Questions 13](#_Toc103691236)

[2.2 Measuring Impact 13](#_Toc103691237)

[2.3 Sample Design 14](#_Toc103691238)

[2.4 Study Instruments 1](#_Toc103691239)

[2.5 Study Quality Assurance 4](#_Toc103691240)

[3 Study Findings 6](#_Toc103691241)

[3.1 Grade 2 Findings 6](#_Toc103691242)

[3.1.1 Overall Reading Achievement (Grade 2) 6](#_Toc103691243)

[3.1.2 Reading Achievement by Subtask (Grade 2) 9](#_Toc103691244)

[3.1.3 Reading Achievement by Sex and L1 (Grade 2) 25](#_Toc103691245)

[3.2 Grade 3 Findings 27](#_Toc103691246)

[3.2.1 Overall Reading Achievement (Grade 3) 27](#_Toc103691247)

[3.2.2 Reading Achievement by Subtask (Grade 3) 30](#_Toc103691248)

[3.2.3 Reading Achievement by Sex and Language (Grade 3) 46](#_Toc103691249)

[3.3 Extrapolation of Reading Achievement 48](#_Toc103691250)

[3.3.1 Equivalent Scoring 48](#_Toc103691251)

[3.3.2 Comparing EGRP II Baseline and Endline Findings with the 2020 NARN and CB‑EGRA Scores from Previous Years 49](#_Toc103691270)

[3.3.3 Reading Ability Categories and the National Early Grade Reading Benchmark 51](#_Toc103691271)

[4 Summary and Conclusions 53](#_Toc103691273)

[5 Study Limitations 56](#_Toc103691274)

[References 58](#_Toc103691275)

[Annex A: EGRP II MEL Indicator Reporting 59](#_Toc103691276)

List of Tables

[Table 1: Program impact on student reading proficiency levels, by grade 2](#_Toc103691301)

[Table 2: Program impact on student reading proficiency levels, overall intervention, by sex and grade 2](#_Toc103691302)

[Table 3: Program impact on student reading proficiency levels, COVID-19 response intervention, by sex and grade 3](#_Toc103691303)

[Table 4: Value-added impact of the COVID-19 response intervention on student reading proficiency levels, by grade 6](#_Toc103691304)

[Table 5: EGRP II-supported interventions 11](#_Toc103691305)

[Table 6: Baseline and endline study sample sizes 1](#_Toc103691306)

[Table 7: Description of grade 2 CB-EGRA assessment tool 2](#_Toc103691307)

[Table 8: Description of grade 3 CB-EGRA assessment tool 3](#_Toc103691308)

[Table 9: Change in average item scores for the grade 2 letter/matra identification subtask 11](#_Toc103691309)

[Table 10: Average item scores for the grade 2 word and sentence reading subtask 13](#_Toc103691310)

[Table 11: Average item scores for the grade 2 vocabulary subtask 15](#_Toc103691311)

[Table 12: Average item scores for the grade 2 dictation subtask 17](#_Toc103691312)

[Table 13: Average item scores for the grade 2 listening comprehension subtask 20](#_Toc103691313)

[Table 14: Average item scores for the grade 2 reading comprehension subtask 22](#_Toc103691314)

[Table 15: Average item scores for the grade 2 calendar reading subtask 25](#_Toc103691315)

[Table 16: Change in average item scores for the grade 3 word and sentence reading subtask 32](#_Toc103691316)

[Table 17: Average item scores for the grade 3 vocabulary subtask 34](#_Toc103691317)

[Table 18: Average item scores for the grade 3 word separation subtask 36](#_Toc103691318)

[Table 19: Average item scores for the grade 3 dictation subtask 38](#_Toc103691319)

[Table 20: Average item scores for the grade 3 listening comprehension subtask (30 words) 41](#_Toc103691320)

[Table 21: Average item scores for the grade 3 reading comprehension subtask 43](#_Toc103691321)

[Table 22: Average item scores for the grade 3 calendar reading subtask 46](#_Toc103691322)

[Table 23: Equivalent CB-EGRA scores for emergent and fluent benchmarks (in cwpm) 48](#_Toc103691323)

[Table 24: Categories of readers, by grade 51](#_Toc103691324)

List of Figures

[Figure 1: Program impact on student reading proficiency levels, by grade and language, overall intervention 5](#_Toc103691359)

[Figure 2: Implementation phases for community- and home-based learning in Madhesh Province 10](#_Toc103691360)

[Figure 3: Quasi-experimental design to measure the impact of the COVID-19 response activities 13](#_Toc103691361)

[Figure 4: Screen shots of data collection monitoring and real-time visualization systems, endline evaluation 5](#_Toc103691362)

[Figure 5: Change in average percentage scores by subtask in the overall intervention (grade 2) 7](#_Toc103691363)

[Figure 6: Change in average percentage scores by subtask in the COVID-19 response intervention (grade 2) 8](#_Toc103691364)

[Figure 7: Distribution of overall average percentage scores (grade 2) 9](#_Toc103691365)

[Figure 8: Student stimulus for the grade 2 letter/matra identification subtask 10](#_Toc103691366)

[Figure 9: Distribution of scores for the grade 2 letter/matra identification subtask 10](#_Toc103691367)

[Figure 10: Student stimulus for the grade 2 word and sentence reading subtask 12](#_Toc103691368)

[Figure 11: Distribution of scores for the grade 2 word and sentence reading subtask 12](#_Toc103691369)

[Figure 12: Student stimulus for the grade 2 vocabulary subtask 14](#_Toc103691370)

[Figure 13: Distribution of scores for the grade 2 vocabulary subtask 14](#_Toc103691371)

[Figure 14: Student stimulus for the grade 2 dictation subtask 16](#_Toc103691372)

[Figure 15: Distribution of scores for the grade 2 dictation subtask 16](#_Toc103691373)

[Figure 16: Student stimulus for the grade 2 listening comprehension subtask 18](#_Toc103691374)

[Figure 17: Distribution of scores for the grade 2 listening comprehension subtask 19](#_Toc103691375)

[Figure 18: Student stimulus for the grade 2 reading comprehension subtask 20](#_Toc103691376)

[Figure 19: Distribution of scores for the grade 2 reading comprehension subtask 21](#_Toc103691377)

[Figure 20: Student stimulus for the grade 2 calendar reading subtask 23](#_Toc103691378)

[Figure 21: Distribution of scores for the grade 2 calendar reading subtask 24](#_Toc103691379)

[Figure 22: Changes in proficiency levels of grade 2 students, by sex 26](#_Toc103691380)

[Figure 23: Changes in proficiency levels of grade 2 students, by language, overall intervention 27](#_Toc103691381)

[Figure 24: Change in average percentage scores by subtask in the overall intervention (grade 3) 28](#_Toc103691382)

[Figure 25: Change in average percentage scores by subtask in the COVID-19 response intervention (grade 3) 29](#_Toc103691383)

[Figure 26: Distribution of overall average percentage scores (grade 3) 30](#_Toc103691384)

[Figure 27: Student stimulus for the grade 3 word and sentence reading subtask 31](#_Toc103691385)

[Figure 28: Distribution of scores for the grade 3 word and sentence reading subtask 31](#_Toc103691386)

[Figure 29: Student stimulus for the grade 3 vocabulary subtask 33](#_Toc103691387)

[Figure 30: Distribution of scores for the grade 3 vocabulary subtask 33](#_Toc103691388)

[Figure 31: Student stimulus for the grade 3 word separation subtask 35](#_Toc103691389)

[Figure 32: Distribution of scores for the grade 3 word separation subtask 35](#_Toc103691390)

[Figure 33: Student stimulus for the grade 3 dictation subtask 37](#_Toc103691391)

[Figure 34: Distribution of scores for the grade 3 dictation subtask 37](#_Toc103691392)

[Figure 35: Student stimulus for the grade 3 listening comprehension subtask 39](#_Toc103691393)

[Figure 36: Distribution of scores for the grade 3 listening comprehension subtask (30 words) 40](#_Toc103691394)

[Figure 37: Student stimulus for the grade 3 reading comprehension subtask (60 words) 41](#_Toc103691395)

[Figure 38: Distribution of scores for the grade 3 reading comprehension subtask 42](#_Toc103691396)

[Figure 39: Student stimulus for the grade 3 calendar reading subtask 44](#_Toc103691397)

[Figure 40: Distribution of scores for the grade 3 calendar reading subtask 45](#_Toc103691398)

[Figure 41: Changes in proficiency levels of grade 3 students, by sex 47](#_Toc103691399)

[Figure 42: Changes in proficiency levels of grade 3 students, by language, overall intervention 48](#_Toc103691400)

[Figure 43: Percentage of grade 3 students who met the reading benchmark in the 2020 NARN, 2021 EGRP II baseline, and 2022 EGRP II endline 50](#_Toc103691401)

[Figure 44: Comparison of average CB-EGRA scores between previous CB‑EGRA assessments and the EGRP II baseline and endline 50](#_Toc103691402)

Acronyms and Abbreviations

CB-EGRA Classroom-Based Early Grade Reading Assessment

CEHRD Center for Education and Human Resource Development

COVID-19 Coronavirus Disease 2019 (SARS-CoV-2)

cwpm Correct Words Per Minute

DIBELS Dynamic Indicators of Basic Early Literacy Skills one use, citation

EGR Early Grade Reading

EGRA Early Grade Reading Assessment

EGRP II Early Grade Reading Program II

ERO Education Review Office

GON Government of Nepal

IC Integrated Curriculum

IEMIS Integrated Education Management Information System

IND Indicator Code

L1, L2 First Language, Second Language

LEU Local Education Unit two uses

MEL Monitoring, Evaluation, and Learning

MOE Ministry of Education (former name of MOEST) in citations only

MOEST Ministry of Education, Science, and Technology

N/A Not Applicable

NARN National Assessment for Reading and Numeracy

NEGRP National Early Grade Reading Program

ORF Oral Reading Fluency

*p*-value Probability of occurrence by chance (example: \*\*\**p* <.01)

RTI RTI International (registered trademark and trade name of Research Triangle Institute)

St Subtask

TPD Teacher Professional Development

TPS Teacher Professional Support

USAID United States Agency for International Development

Acknowledgments

The authors would like to acknowledge the contributions of all stakeholders and agencies that made it possible to complete this baseline assessment. We appreciate the kind support of officials from the Education Review Office, who not only provided the data collection tools from their item bank but also rigorously supported the full baseline and endline evaluation processes, from the master training of trainers to data collection in schools. The Early Grade Reading Program II (EGRP II) monitoring, evaluation, and learning coordinators, who led the entire process in their regions, also played crucial roles in administering the assessment. EGRP II’s central and regional office teams as well as district coordinators and local-level program officers—who monitored the classroom-based Early Grade Reading Assessment (CB-EGRA) data collection process and also collected mini-EGRA data—were equally invaluable for this study. Last but not least, we appreciate the efforts of all the teachers who worked as CB-EGRA assessors, and the head teachers who facilitated data collection in their schools despite the hardships caused by the coronavirus disease 2019 (COVID-19) pandemic.

Executive Summary

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

To understand the overall impact over the program period, EGRP II conducted a baseline study in February and March 2021, which was followed by an endline study in February and March 2022. This report discusses endline EGRP II impact, including that of a home- and community-based schooling intervention in the eight districts of Madhesh Province.[[2]](#footnote-3) The content and analysis presented in this report draws on the baseline reports, titled *Baseline Report Vol. 1, Student Reading Performance in the Early Grades* and *Baseline Report Vol. 2, COVID-19 Response: The Home- and Community-Based Schooling Intervention.*

The evaluation aimed to answer the following research questions: (1) What is the program impact on the reading skills of students in grades 2 and 3? (2) What is the program impact on the reading skills of boys and girls? (3) What is the program impact for students who speak Nepali as a first language (L1) and those who speak Nepali as a second language (L2)? (4) What is the value-added of the additional intervention in Madhesh Province?

The classroom-based Early Grade Reading Assessment (CB-EGRA) instrument was the key tool used in the evaluation. It was developed by the Education Review Office (ERO) and is a group-administered instrument used to measure the reading abilities of early grade students. During the baseline evaluation, a small pilot study was conducted to link student scores on the CB-EGRA to their scores on typical EGRA subtasks.

The following statistical models were developed to extrapolate children’s oral reading fluency and comprehension scores from their CB-EGRA scores during the baseline.

* Average grade 2 CB-EGRA percentage score = 19.9 + 0.91 × average oral reading fluency (ORF)
* Average grade 2 CB-EGRA percentage score = 24.003 + 9.201 × average comprehension
* Average grade 3 CB-EGRA percentage score = 22.4 + 0.82 × average ORF
* Average grade 3 CB-EGRA percentage score = 28.149 + 7.674 × average comprehension

To answer the first research question, ***Table 1*** shows the program impact on the percentage of students achieving different levels of reading proficiency. EGRP II identified a 7.5 percentage point increase in the proportion of fluent readers in grade 3​ but did not identify significant improvements in grade 2, although there was no further learning loss in that grade despite significant disruptions caused by the COVID-19 pandemic.

Table 1: Program impact on student reading proficiency levels, by grade

| **Grade** | **Benchmark** | **Time point** | **Percent of students** | **Difference**  **(percentage points)** | **Effect size** |
| --- | --- | --- | --- | --- | --- |
| 2 | Low  (<15 cwpm) | Baseline | 65% | –0.4 | 0.01 |
| Endline | 64% |
| Emergent  (15–44 cwpm) | Baseline | 28% | –0.7 | 0.02 |
| Endline | 27% |
| Fluent  (45+ cwpm) | Baseline | 7% | 1.1 | 0.04 |
| Endline | 8% |
| 3 | Low  (<15 cwpm) | Baseline | 60% | –7.9 | 0.16 |
| Endline | 52% |
| Emergent  (15–44 cwpm) | Baseline | 28% | 0.4 | 0.01 |
| Endline | 28% |
| Fluent  (45+ cwpm) | Baseline | 13% | 7.5\* | 0.20 |
| Endline | 20% |

*Note.* cwpm = correct words per minute.  
\**p* <.05.

To answer the second research question, ***Table 2*** shows the program impact on the percentage of students achieving different levels of reading proficiency disaggregated by the sex of students in the overall intervention. The words in parentheses in the “Impact” column indicate whether the program impact was favorable for boys or girls. As the table indicates, there were some differences in scores by sex, but none of the differences was statistically significant.

Table 2: Program impact on student reading proficiency levels, overall intervention, by sex and grade

| **Grade** | **Subtask** | **Sex** | **Baseline** | **Endline** | **Difference**  **(percentage points)** | **Impact**  **(percentage points)** |
| --- | --- | --- | --- | --- | --- | --- |
| 2 | Low  (<15 cwpm) | Boys | 66.8% | 64.9% | –2.0 | 3.1 (boys) |
| Girls | 62.8% | 63.9% | 1.2 |
| Emergent  (15–44 cwpm) | Boys | 25.5% | 27.7% | 2.2 | 5.6 (boys) |
| Girls | 30.1% | 26.7% | –3.4 |
| Fluent  (45+ cwpm) | Boys | 7.7% | 7.5% | –0.2 | 2.4 (girls) |
| Girls | 7.2% | 9.4% | 2.2 |
| 3 | Low  (<15 cwpm) | Boys | 65.2% | 55.1% | –10.1 | 4.5 (boys) |
| Girls | 55.1% | 49.5% | –5.6 |
| Emergent  (15–44 cwpm) | Boys | 24.3% | 29.1% | 4.9 | 8.4 (boys) |
| Girls | 30.5% | 26.9% | –3.5 |
| Fluent  (45+ cwpm) | Boys | 10.5% | 15.8% | 5.3 | 3.9 (girls) |
| Girls | 14.4% | 23.6% | 9.1 |

*Note.* cwpm = correct words per minute.

In contrast, ***Table 3*** shows that there were statistically significant differences in outcomes for both boys and girls in the COVID-19 intervention. In grade 2, there were significant decreases in the percentages of boys and girls in the low proficiency category, and significant increases in the percentages of boys and girls in the emergent and fluent categories. In all three cases, the impacts were larger for girls. In grade 3, there were also significant decreases in the percentages of boys and girls in the low proficiency category, and significant increases in the percentages of boys and girls in the emergent and fluent categories. The improvements in this grade favored boys, with a larger percentage of boys than girls moving into the fluent category.

Table 3: Program impact on student reading proficiency levels, COVID-19 response intervention, by sex and grade

| **Grade** | **Subtask** | **Sex** | **Baseline** | **Endline** | **Difference**  **(percentage points)** | **Impact**  **(percentage points)** |
| --- | --- | --- | --- | --- | --- | --- |
| 2 | Low  (<15 cwpm) | Boys | 71.0% | 51.5%\*\* | –19.5 | 5.4 (girls) |
| Girls | 75.6% | 50.8%\*\*\* | –24.8 |
| Emergent  (15–44 cwpm) | Boys | 19.7% | 32.0%\* | 12.3 | 2.9 (girls) |
| Girls | 16.9% | 32.1%\*\*\* | 15.2 |
| Fluent  (45+ cwpm) | Boys | 9.3% | 16.5% | 7.2 | 2.4 (girls) |
| Girls | 7.5% | 17.1%\*\* | 9.6 |
| 3 | Low  (<15 cwpm) | Boys | 67.9% | 37.1%\*\*\* | –30.7 | 6.6 (boys) |
| Girls | 70.5% | 46.4%\*\*\* | –24.1 |
| Emergent  (15–44 cwpm) | Boys | 20.2% | 29.1%\*\* | 8.9 | 4.1 (girls) |
| Girls | 14.9% | 28.0%\*\* | 13.1 |
| Fluent  (45+ cwpm) | Boys | 11.9% | 33.7%\*\*\* | 21.8 | 10.8 (boys) |
| Girls | 14.6% | 25.6%\* | 11.0% |

*Note.* cwpm = correct words per minute.  
\**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

To answer the third research question, EGRP II examined the changes in the proportions of students in different reading proficiency categories by home language between baseline and endline in the overall intervention (***Figure 1*).** This analysis identified statistically significant increases in the proportion of fluent readers among Nepali L2 students, with a 4 percentage point increase in grade 2 and an 11 percentage point increase in grade 3. There were no significant increases between baseline and endline for Nepali L1 students. Similar analysis disaggregating by language groups for the COVID-19 response intervention cannot be provided because that sample consisted entirely of L2 learners.

Figure 1: Program impact on student reading proficiency levels, by grade and language, overall intervention

|  |
| --- |
|  |
| \*\**p* < .01. |

To answer the fourth research question, ***Table 4*** shows that the community- and home-based learning (COVID-19 response) interventions that EGRP II designed and supported in Madhesh Province were associated with a strong boost in student reading proficiency in both grades compared to the overall intervention. Using difference-in-differences analysis, EGRP II identified a significant reduction in the proportion of low-proficiency readers (–22.1 percentage points in grade 2 and –18.9 percentage points in grade 3) and a significant increase in the proportion of emergent readers in both grades (14.6 percentage points in grade 2 and 10.9 percentage points in grade 3) for the students in the COVID-19 response intervention when compared to students in the overall intervention. EGRP II also identified an increase in the proportion of fluent readers favoring the students in the COVID-19 response intervention, although this increase was not statistically significant.

Table 4: Value-added impact of the COVID-19 response intervention on student reading proficiency levels, by grade

| **Grade** | **Proficiency category** | **Group** | **Percentage of students in each proficiency category** | | | **Impact**  **(percentage points)** |
| --- | --- | --- | --- | --- | --- | --- |
| **Baseline** | **Endline** | **Baseline-endline difference**  **(percentage points)** |
| 2 | Low  (<15 cwpm) | Overall intervention | 65% | 64% | 0 | –22.1\* |
| COVID-19 response intervention | 74% | 51% | –22 |
| Emergent  (15–44 cwpm) | Overall intervention | 28% | 27% | –1 | 14.6\* |
| COVID-19 response intervention | 18% | 32% | 14 |
| Fluent  (45+ cwpm) | Overall intervention | 7% | 8% | 1 | 7.5 |
| COVID-19 response intervention | 8% | 17% | 9 |
| 3 | Low  (<15 cwpm) | Overall intervention | 60% | 52% | –8 | –18.9\* |
| COVID-19 response intervention | 69% | 43% | –27 |
| Emergent  (15–44 cwpm) | Overall intervention | 28% | 28% | 0 | 10.9\* |
| COVID-19 response intervention | 17% | 28% | 11 |
| Fluent  (45+ cwpm) | Overall intervention | 13% | 20% | 8 | 8.0 |
| COVID-19 response intervention | 13% | 29% | 16 |

\**p* < .05.

The findings from the learning evaluation suggest that overall EGRP II interventions were associated with an improvement in reading proficiency in grade 3 but not grade 2, while EGRP II may have assisted grade 2 learners not to fall backward during the pandemic. At the same time, the overall intervention did not appear to have a differential impact between boys and girls, although there were positive changes for both boys and girls in the COVID-19 response intervention. In addition, there was a positive impact for L2 learners when compared to L1 learners in the overall intervention. The findings related to the community- and home-based schooling activity pointed to clear benefits for grade 2 and 3 students when they were provided with targeted support for catch-up learning.

# Background

## EGRP II Program Description

This subsection places the endline evaluation findings in the context of the overall program interventions. EGRP II was a 2-year, USAID-funded program of technical assistance to the GON that was implemented from June 1, 2020, to May 31, 2022. EGRP II’s support to the GON was 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 aimed 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 EGR service delivery (Objec­tive 2), improving teacher professional support (TPS) (Objective 3), and assisting with the COVID-19 response in the education sector (Objective 4).

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

* **Level 1** included the 10 districts that were scheduled to begin in-school implementation in 2020–2021, as well as the eight districts that were scheduled 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** consisted of the next four NEGRP rollout districts: Dhanusha, Rasuwa, Tanahun, and Taplejung.
* **Level 3** included 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 provided intensive support for implementation of the NEGRP minimum package in the 22 Level 1 and 2 districts, and continued technical assistance for the 16 districts that were targeted under EGRP (Level 3).

Through Objective 1, EGRP II provided technical support to the GON on developing and distributing a 10-day training package for IC teacher professional development (TPD), including a trainer’s guide and a packet of training resource materials covering all subjects in grades 1–3. EGRP II also supported IC orientations and cluster-level TPD trainings that rolled out following a first round of master trainings of trainers (MTOTs) funded by the GON in late 2020. To help the GON add to the cadre of TPD master trainers, the program team assisted the Center for Education and Human Resource Development (CEHRD) to plan for and implement a central-level master training and seven provincial-level trainings of trainers for the IC TPD package. These trainings reached a total of 203 participants, who were then prepared to cascade the training further at district and cluster levels.

To support policy making and use of data for decision making related to EGR, EGRP II provided technical assistance to the ERO on data analysis, reporting, and dissemination for the 2020 National Assessment of Reading and Numeracy (NARN)—and spearheaded dissemination of the EGRP endline findings—at national, provincial, district, and local levels. Through briefing papers, meetings, and written comments, the team shared inputs on EGR practices in the new School Education Sector Plan, which the GON plans to roll out from mid-2022 through 2030, to promote effective alignment of the new plan with the NEGRP minimum package. In addition, EGRP II supported a task team made up of MOEST and Central Level Agency representatives to revise Nepal’s national EGR benchmarks to align more appropriately with student skill levels.

Under Objective 2, EGRP II conducted orientation and coordination meetings with district and local government meetings. The program also supported three rounds of local capacity-development workshops, with Round 1 targeting all 396 palikas[[3]](#footnote-4) and Round 2 aiming to support the 251 palikas in the 22 NEGRP expansion districts. The first two rounds of workshops emphasized use of Integrated Education Management Information System (IEMIS) data for decision making, Municipal Education Plan (MEP) development, NEGRP and IC implementation, and TPS rollout at the local level. The third round of capacity-development efforts focused on supporting district governments to form technical task teams (3Ts) that then provided support to a targeted set of local governments on finalizing their MEPs, IEMIS education profiles, and budgets. In addition, in early 2021, EGRP II successfully completed distribution of more than half a million supplementary reading materials (SRMs) to 2,927 schools in 22 NEGRP expansion districts.

As a result of these combined efforts over the two program years, by EGRP II’s end, 67.2% of the 396 target palikas had either a final draft or an approved version of their MEP and IEMIS education profile in place. Moreover, 80% of palikas had allocated part of their budgets to EGR activities, and 76% to education-in-emergencies activities. Almost all (99%) of palikas had used IEMIS data for education sector planning and budgeting, while 96.5% of local governments had taken steps to validate the IEMIS data shared by schools.

Through Objective 3, EGRP II coordinated closely with CEHRD to revise Nepal’s TPS approach in response to research conducted under EGRP as well as the evolving decentralization of governance in the country. As a result, a revised TPS Management Procedure, TPS Guideline, and TPS training manual were developed. EGRP II also supported local governments to plan and budget for TPS provision through the local capacity-development activities described above as part of Objective 2. EGRP II supported CEHRD and the Education Training Centers to conduct TPS MTOTs to cover the 22 Level 1 and 2 districts that had not previously received TPS training through EGRP. A total of 80 TPS master trainers were reached through this effort and prepared to continue rolling out cluster-level TPS training if it is budgeted in upcoming GON fiscal years.

EGRP II supported joint monitoring visits to 1,307 schools in total across the Level 1–3 districts to promote effective professional support at the school level. The program also provided general orientation on the revised TPS guidelines to all palikas, and then worked with district and local stakeholders to identify one TPS sample palika in each district that would act as a hub of best practices for the other palikas in their district. To roll out the sample palika approach, EGRP II provided TPS capacity development to the 38 sample palikas and then began assisting them to conduct teacher learning groups before the omicron variant forced a return to widespread school closures in early 2022.

As a result of these TPS activities, by the end of April 2022, 87% of EGRP II’s target palikas had formally selected their TPS options and 96% had formed TPS expert groups to advise them, while TPS training for the expert groups had been provided in about one-quarter of the palikas.

Efforts under Objective 4 focused on supporting the GON’s response to COVID-19 in the education sector through a partnership with Open Learning Exchange Nepal to develop new digital early grade learning content for upload to CEHRD’s online learning portal. In total, 119 new digital learning lessons in the subjects of Nepali, mathematics, science, social studies, and Nepali Sign Language, as well as lessons for children with dyslexia, were finalized and handed over to CEHRD, although they had not yet been uploaded to the CEHRD portal by the end of EGRP II. The EGRP II team members also provided ongoing COVID-19 support in the education sector at the national, provincial, district, and local levels.

In addition, as previously mentioned, a robust, well-targeted community- and home-based learning approach was rolled out in the eight districts of Madhesh Province. This intervention had two phases, with Phase 1 starting in Year 1 and covering 219 schools and Phase 2 covering an additional 268 schools in Year 2 (***Figure 2***). In total, 487 schools participated across the two phases, with children divided into approximately three learning clusters per school (1,459 clusters in total) that met with a teacher outside of school hours for remedial learning activities aligned with the IC. In total, 82,245 grade 1–3 students (52.5% girls) benefited from the catch-up learning activities.

Figure 2: Implementation phases for community- and home-based learning in Madhesh Province

Activities related to the community- and home-based learning intervention included initial events to kick off activities in the target palikas and schools, plus orientations on the activity’s overall approach with school and community stakeholders. In addition, EGRP II provided training for participating teachers, focusing on how to set up and teach in the learning clusters and use appropriate multigrade, multilevel instructional approaches for groups of children of mixed ages and abilities. EGRP II also procured and distributed digital tablets preloaded with early grade learning content, stationery packs (notebooks, drawing pads, colored pencils, etc.), decodable readers, and sets of teaching and learning materials for the learning clusters.

The EGRP II team provided small grants to cover activity expenses in participating schools during Phase 1, switching to a travel allowance payment approach directly with participating teachers and head teachers in Phase 2 due to the administrative burdens of providing small grants during times of COVID-19 disruptions. The team also supported head teachers and local government officials to conduct periodic quality monitoring visits to the clusters, which identified modest but steady improvements in quality over time. The homeschooling activity ended successfully in mid-March 2022, with strong buy-in from participating local governments to continue providing catch-up learning to students in need after EGRP II support ended.

***Table 5*** breaks down the overall interventions that EGRP II supported across all program districts, plus the interventions for the COVID-19 response in Madhesh Province, and the approximate timing of the interventions in relation to baseline and endline data collection.

Table 5: EGRP II-supported interventions

|  | **Activities implemented** | **Baseline** | **Activities implemented** | **Endline** |
| --- | --- | --- | --- | --- |
| **Overall intervention** | SRM distribution to all schools in Level 1 and 2 districts, IC orientations at district and cluster level, Round 1 of local capacity-development workshops for local governments across Levels 1–3, joint monitoring visits, distribution of flexible learning materials for COVID-19 response across Levels 1–3 districts | February–March 2021 | IC orientations at district and cluster level, TPS orientations for local governments, formation of local government TPS expert groups, joint monitoring visits, TPS training for sample palikas, distribution of flexible learning materials for COVID-19 response across Levels 1–3 districts | February–March 2022 |
| **COVID-19 response intervention** | Start-up of Phase 1 of community and home-based schooling, including target palika and school selection; orientations for local governments and schools; formation of three clusters of grades 1–3 students in each target school; formation of cluster management committees consisting of parents and community leaders; teacher training for cluster teachers | Four months of implementation of community- and home-based learning activities in the 219 Phase 1 schools, with approximately 15 hours of catch-up learning in the clusters per week  Start-up and implementation of Phase 2 of the community and home-based learning activity. Previous activities continued in the 219 Phase 1 schools. For the 286 additional Phase 2 schools, activities included target palika and school selection, orientations for local governments and schools, formation of three clusters of grades 1–3 students in each target school, formation of cluster management committees consisting of parents and community leaders, teacher training for cluster teachers, provision of tablets to teachers, provision of stationery and decodable readers to students, and quality monitoring by local government officials and head teachers. All 487 schools received approximately 15 hours per week of catch-up learning in the clusters for 7.5 months. |

## 

## Evaluation Context

As a method to determine the project’s impact over the program period, EGRP II conducted a baseline study in February and March 2021 and undertook an endline in February and March 2022. Although the project started in June 2020, EGRP II timed the baseline study to align with the end of the academic year in Nepal. The public school year usually ends in February–March, although the pandemic resulted in some disruptions to the usual timing and the 2020–2021 school year ultimately was extended for a few months beyond March 2021. Subsequently, the 2021–2022 school year returned to roughly the normal timing and ended in March 2022.

To assess student reading ability, the CB-EGRA was conducted by trained teachers in the sampled schools (see Section 2.3 below for the sample design). 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, at baseline, EGRP II simultaneously conducted a subsample-based mini-EGRA consisting of an oral reading passage and related comprehension subtasks. The aim was to use a statistical model to produce equivalence scores between skills measured by the CB-EGRA and the EGRA-measured skills of reading fluency and comprehension. By describing this statistical model, EGRP II 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.

# Study Design

## Research Questions

The EGRP II endline study was designed to answer four questions.

1. What is the program impact on the reading skills of students in grades 2 and 3?
2. What is the program impact on the reading skills of boys and girls?
3. What is the program impact for students who speak Nepali as an L1 and those who speak Nepali as an L2?
4. What is the value-added of the COVID-19 response intervention in Madhesh Province?

## Measuring Impact

Measuring the impact on learning outcomes for the EGRP II-supported regions required a simple difference in averages at two time points. This design served as an effective method to report on standard USAID learning outcome indicators.

To measure the value-added for the COVID-19 response, the evaluation compared the impact of the community- and home-based learning activity to the impact of the overall activities implemented in other EGRP II locations. This quasi-experimental design compared the gains between two time points in the two program areas (***Figure 3***) using a difference in differences analysis.[[4]](#footnote-5)

Figure 3: Quasi-experimental design to measure the impact of the COVID-19 response activities

Chart, bubble chart

Description automatically generated

As the figure illustrates, the value-added gain for the COVID-19 response intervention in Madhesh Province above the overall EGRP II activities in other target areas can be attributed to the community- and home-based learning intervention.

## Sample Design

As indicated in the background section above, EGRP II worked in 38 program districts covering 396 palikas and supporting approximately 13,500 schools. As such, an estimated 328,929 students from grade 2 and 333,968 from grade 3 made up the population for the study. Using a 95% confidence level, 45 schools were sampled at random for the study of overall intervention impacts and 47 schools for the study of the impacts of the COVID-19 response intervention. The baseline and endline samples of students for both intervention areas are presented in ***Table 6.***

Table 6: Baseline and endline study sample sizes

| Sample source | No. of schools | No. of students assessed with CB-EGRA | | | | | | | No. of students assessed with mini-EGRA | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade 2 | | | Grade 3 | | | *Grand totals* | Grade 2 | | | Grade 3 | | | *Grand totals* |
| B | G | T | B | G | T | B | G | T | B | G | T |
| **Baseline** | | | | | | | | | | | | | | | |
| **Overall intervention** | 45 | 373 | 379 | 752 | 361 | 465 | 826 | *1,578* | 96 | 130 | 226 | 91 | 136 | 227 | *453* |
| **COVID-19 response intervention** | 47 | 410 | 510 | 920 | 401 | 498 | 899 | *1,819* | Not applicable (N/A) | | | | | | *–* |
| ***Baseline grand totals*** | *92* | *783* | *889* | *1,672* | *762* | *963* | *1,725* | *3,397* | *96* | *130* | *226* | *91* | *136* | *227* | *453* |
| **Endline** | | | | | | | | | | | | | | | |
| **Overall intervention** | 45 | 330 | 357 | 687 | 315 | 389 | 704 | *1,391* | *N/A* | | | | | | |
| **COVID-19 response intervention** | 47 | 381 | 528 | 909 | 385 | 537 | 922 | *1,831* |
| ***Endline grand totals*** | *92* | *711* | *885* | *1,596* | *700* | *926* | *1,1626* | *3,222* |

*Note.* B = boys. G = girls. T = total.

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

At baseline, from the 47 schools in the COVID-19 response sample, EGRP II selected 920 students from grade 2 (boys: 44.6%; girls: 55.4%) and 899 students from grade 3 (boys: 44.6%, girls: 55.4%). 100% of the sampled students were learners with Nepali as their L2.

At endline, from the 45 schools in the overall sample, EGRP II selected 687 students from grade 2 (boys: 48.0%; girls: 52.0%) and 704 students from grade 3 (boys: 44.7%, girls: 55.3%). Of the grade 2 sample, 45.1% were learners with Nepali as L1 and 54.9% were learners with Nepali as L2. In grade 3, the sample consisted of 45.6% learners with Nepali as L1 and 54.4% with Nepali as L2. Overall, most of the sampled students (54.6%) had Nepali as their L2.

At endline, from the 47 schools in the COVID-19 response sample, EGRP II selected 909 students from grade 2 (boys: 41.9%; girls: 58.1%) and 922 students from grade 3 (boys: 41.8%, girls: 58.2%). As with the baseline, 100% of the sampled students at endline were learners with Nepali as their L2.

Although there were some changes in the sex and language composition of the sample between baseline and endline – for example, a greater percentage of girls in the COVID-19 response sample at endline – the differences were not statistically significant.

In this assessment, EGRP II used the approach followed by ERO to conduct the CB-EGRA (ERO 2017), which aims to sample an average of 18 students 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 of was calculated for the overall sample. 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, 45 schools were sampled for the study. Different sociocultural and geographical attributes were considered when selecting the sample districts and municipalities. One district was selected from each province to obtain a balance with regard to the language majority, level of EGRP II’s interventions, and topographical distribution. From each province, one palika was selected randomly and, to balance the sampling weight, the number of schools was adjusted to be sampled randomly from each palika. Initial student selection within each school, for administration of the CB‑EGRA, was also random. While adjusting the number of sampled schools from each palika, a sample number was selected so that the ratio of sample weights among the cluster would not exceed 10.

## Study Instruments

As described in Section 1.2, the CB-EGRA was used to collect students’ reading proficiency data for the baseline and endline. The CB-EGRA had a total of seven subtasks and each subtask included 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, instructing the whole class at once on each subtasks. Two separate CB-EGRA assessment tools were used for grade 2 and grade 3. ***Table 7*** and ***Table 8*** provide the details of the tools that were used for each grade in the study.

Table 7: Description of grade 2 CB-EGRA assessment tool

| No. | Subtask name | Items | Type | No. of distractors for each item | Example? | Time (minutes) | | Subtask weight |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Example/ teacher instruction | Assessment |
| 1 | Letter/matra identification | 3 | Multiple choice | 5 | Yes | 2 min | 3 min | 1 |
| 2 | Word and sentence reading | 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** assessed students’ ability to identify the first letter or matra from the word that the teacher said.

**Subtask 2: Word and sentence** reading assessed students’ ability to identify the word or a sentence that the teacher read aloud.

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

**Subtask 4: Dictation** assessed students’ writing skills. For this subtask, students had to write the entire sentence correctly as the teacher dictated. The teacher read the sentence three times.

**Subtask 5: Listening comprehension** measured the number of comprehension questions that students answered correctly, based on a story of 25 words that the teacher read aloud two times.

**Subtask 6: Reading comprehension** measured the number of comprehension questions that students answered correctly after they had read a 60-word paragraph.

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

Table 8: 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 reading | 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** reading assessed students’ ability to identify the word or a sentence that the teacher read aloud.

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

**Subtask 3: Word separation** assessed the children’s ability to decode words. It measured how well children could separate the words in a sentence when all the words were joined together.

**Subtask 4: Dictation** assessed students’ writing skills. For this subtask, students had to write the entire sentence correctly as the teacher dictated. The teacher read the sentence three times.

**Subtask 5: Listening comprehension** measured the number of comprehension questions the students answered correctly, based on a story of 30 words that the teacher read aloud two times.

**Subtask 6: Reading comprehension** measured the number comprehension questions that students answered correctly after reading a 60-word passage.

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

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

## Study Quality Assurance

Quality assurance was prioritized throughout the study. For the baseline evaluation in February 2021, the Kathmandu-based EGRP II monitoring, evaluation, and learning (MEL) team, along with ERO technical personnel, had provided an in-person training of trainers to EGRP II technical leads and regional MEL coordinators. This 2-day training focused on the theoretical and practical aspects of the CB-EGRA and EGRA and the logistics that would be required while the trainees were collecting the data. For the endline, a virtual refresher training of trainers ~~sufficed~~ was conducted in February 2022 for the EGRP II technical leads and regional MEL coordinators.

The MEL team 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 of progress on baseline and endline data collection.

The EGRP II MEL coordinators, along with the Kathmandu-based team members, subsequently rolled out the CB-EGRA training at both baseline and endline to teachers from the sampled schools who would administer the CB-EGRA, while the EGRP II district coordinators were trained on mini-EGRA administration and quality monitoring for the baseline. 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. At baseline, mini-EGRA data collection was carried out by the district coordinators. 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 assessments.

***Figure 4*** shows screen shots of the assessment monitoring and real-time data visualization systems from the endline.

|  |
| --- |
| Figure 4: Screen shots of data collection monitoring and real-time visualization systems, endline evaluation |
|  |
| Chart, scatter chart  Description automatically generated |

# Study Findings

This section presents the findings from the EGRP II learning evaluation, calculated by comparing the baseline and endline results. For the CB-EGRA data for both grades 2 and 3, sample weights were calculated 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 then calculated based on the sample weights and subtask weights.[[5]](#footnote-6) IBM SPSS Statistics version 21 was used to analyze the data using the Complex Sample module. Using this approach to sample weighting affords confidence that the baseline and endline results represent the estimated population.

## Grade 2 Findings

### Overall Reading Achievement (Grade 2)

For the overall intervention, on average, grade 2 students were able to answer 6 out of 21 questions during both the baseline and endline assessments. There was a statistically significant increase at endline for only one of the subtasks: calendar reading. At both baseline and endline, students performed best on letter and matra identification and listening comprehension, and had the most difficulty with the dictation subtask.

A breakdown of the average subtask scores for grade 2 students in the overall intervention at baseline and endline is presented in ***Figure 5.***

Figure 5: Change in average percentage scores by subtask in the overall intervention (grade 2)

\*\*\**p* < .001.

For the COVID-19 response intervention, at baseline, grade 2 students were able to answer an average of roughly 5 out of 21 questions correctly, a figure that increased to nearly 8 out of 21 questions at endline. There were significant increases in the average scores for all subtasks, with the largest increases for the calendar reading, reading comprehension, and listening comprehension subtasks. Students scored the lowest on dictation at both baseline and endline and experienced the smallest increase by endline on this subtask.

A breakdown of the average scores for grade 2 students for each subtask at baseline and endline is presented in ***Figure 6.***

Figure 6: Change in average percentage scores by subtask in the COVID-19 response intervention (grade 2)

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

The distribution of composite scores in grade 2—in other words, the average percentage scores across all subtasks—is presented for the overall intervention and the COVID-19 response intervention in ***Figure 7.*** The overall average percentage scores are categorized into five different groups: 0, 1%–20%, 21%–40%, 41%–60%, 61%–80% and 81%–100%. These distributions showed little change between baseline and endline for the overall intervention, but a substantial movement of students in the COVID-19 response endline out of the lower distributions into higher distributions.

Figure 7: Distribution of overall average percentage scores (grade 2)

|  |
| --- |
|  |
|  |

### Reading Achievement by Subtask (Grade 2)

The following analysis provides details about the average grade 2 percentage scores for each of 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 three items in the subtask were multiple choice. There were five possible answers in each item, including one correct option and four distractors. ***Figure 8*** isa screen shot of the baseline and endline student stimulus for the grade 2 letter/matra identification subtask.

Figure 8: Student stimulus for the grade 2 letter/matra identification subtask

Table

Description automatically generated

The score distributions for the subtask for both the overall intervention and the COVID-19 response intervention are presented in ***Figure 9.*** These figures show that roughly half of students could not answer any questions or could only answer one question at baseline. However, for the COVID-19 response intervention at endline, there was a 5-point increase in the percentage of students who could answer two out of three questions correctly and a 10‑point increase in the percentage of students who could correctly answer all three questions.

Figure 9: Distribution of scores for the grade 2 letter/matra identification subtask

|  |
| --- |
|  |
|  |

***Table 9*** shows the change in the average item scores for the grade 2 letter/matra identification subtask for the overall intervention and the COVID-19 response intervention. As this table demonstrates, in general, students found the vowel/consonant and simple matra identification tasks easier than the mixed letter/matra task. There were increases in the scores at endline on identifying a vowel or consonant letter for both groups, while there were increases only on identifying a simple matra and mixed letter/matra for the COVID-19 response intervention. However, only the increase for identifying mixed a mixed letter/matra in the COVID-19 response intervention was statistically significant.

Table 9: Change in average item scores for the grade 2 letter/matra identification subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference (percentage points) |
| 1a | Identify vowel or consonant letter | Overall intervention | 58.6% (3.6%) | 65.8% (3.6%) | 7.2 |
| COVID-19 response intervention | 53.2% (3.6%) | 61.7% (4.5%) | 8.5 |
| 1b | Identify simple matra | Overall intervention | 60.3% (3.4%) | 60.7% (4.0%) | 0.3 |
| COVID-19 response intervention | 53.7% (3.4%) | 62.3% (4.4%) | 8.7 |
| 1c | Identify mixed letter/matra | Overall intervention | 32.1% (3.2%) | 20.9% (3.6%) | –11.2 |
| COVID-19 response intervention | 31.7% (3.2%) | 49.9% (3.5%) | 18.2\*\*\* |

*Note.* Standard errors in parentheses.  
\*\*\**p* < .001.

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 10*** provides a screen shot of the baseline and endline student stimulus for the grade 2 word and sentence reading subtask.

Figure 10: Student stimulus for the grade 2 word and sentence reading subtask

Table

Description automatically generated

The distribution of scores for this subtask (***Figure 11***) indicates that there were small shifts between baseline and endline for the overall intervention but without a clear pattern. For the COVID-19 response intervention, zero scores decreased by 13.4 percentage points, and the proportion of students answering two or three questions correctly between baseline and endline increased substantially.

Figure 11: Distribution of scores for the grade 2 word and sentence reading subtask

|  |
| --- |
|  |
|  |

***Table 10*** shows the change in the average item scores for the grade 2 word and sentence reading subtask for the overall intervention and the COVID-19 response intervention. Generally, and not unexpectedly, students found the one-word and three-word sentence items easier and the four-word sentence item more challenging. There were increases in both groups on the average scores for identifying one word, and in the COVID-19 response group for identifying three and four-word sentences.

Table 10: Average item scores for the grade 2 word and sentence reading subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 2a | Identify one word | Overall intervention | 62.4% (4.1%) | 65.6% (4.5%) | 3.2 |
| COVID-19 response intervention | 45.5% (4.1%) | 62.3% (3.4%) | 16.8\*\* |
| 2b | Identify three-word sentence | Overall intervention | 48.0% (2.8%) | 48.7% (3.7%) | 0.7 |
| COVID-19 response intervention | 40.2% (2.8%) | 52.9% (2.6%) | 12.7\*\* |
| 2c | Identify four-word sentence | Overall intervention | 25.7% (3.0%) | 23.4% (2.4%) | –2.3 |
| COVID-19 response intervention | 26.9% (3.0%) | 34.1% (2.9%) | 7.2 |

*Note.* Standard errors in parentheses.  
\*\**p* < .01.

**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 baseline and endline student stimulus is presented in ***Figure 12.***

Figure 12: Student stimulus for the grade 2 vocabulary subtask

Graphical user interface, table

Description automatically generated

The score distribution for the grade 2 vocabulary subtask ***(Figure 13***) shows that a considerable percentage of students in both the overall intervention and the COVID-19 response intervention scored zero at baseline. By endline, there were limited changes in the distribution for the overall intervention group, with a small decrease in the percentage of zero scores and an increase in the percentage of students answering two out of three questions correctly in the overall intervention. At the same time, there was a substantial reduction in zero scores and increases in the percentages of students scoring three or more correct in the COVID-19 response group.

Figure 13: Distribution of scores for the grade 2 vocabulary subtask

|  |
| --- |
|  |
|  |

***Table 11*** shows the change in the average item scores for the grade 2 vocabulary subtask for the overall intervention and the COVID-19 response intervention. Students in both groups generally found identifying antonyms and synonyms more challenging than defining a word. Average scores decreased between baseline and endline for the overall intervention group, although the changes were not significant. At the same time, average scores increased for the COVID-19 response group (significantly for defining a word and identifying antonyms), enabling that group to overtake the average scores in the overall group across all three subtasks.

Table 11: Average item scores for the grade 2 vocabulary subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 3a | Define a word | Overall intervention | 50.6% (3.4%) | 45.8% (4.2%) | –4.8 |
| COVID-19 response intervention | 39.4% (3.4%) | 52.5% (3.5%) | 13.1\*\* |
| 3b | Identify antonyms | Overall intervention | 25.7% (3.4%) | 23.8% (4.6%) | –1.9 |
| COVID-19 response intervention | 27.8% (3.4%) | 41.3% (4.5%) | 13.5\* |
| 3c | Identify synonyms | Overall intervention | 33.2% (3.4%) | 25.0% (3.5%) | –8.2 |
| COVID-19 response intervention | 30.8% (3.4%) | 35.0% (4.1%) | 4.3 |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\**p* < .01.

**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 baseline and endline student stimulus is presented in ***Figure 14.***

Figure 14: Student stimulus for the grade 2 dictation subtask

|  |
| --- |
| Text  Description automatically generated with medium confidence |
| Text  Description automatically generated |
| Table  Description automatically generated |

The score distribution for the grade 2 dictation subtask (***Figure 15***) shows that the majority of students scored zero at baseline and endline in both the overall intervention and the COVID‑19 response intervention, although the percentage of zero scorers had decreased by 8.8 percentage points in the latter group at endline.

Figure 15: Distribution of scores for the grade 2 dictation subtask

|  |
| --- |
|  |
|  |

*Table 12* shows the change in the average item scores for the grade 2 dictation subtask for the overall intervention and the COVID-19 response intervention. At both time points, students in both groups found all items very difficult, although the three-word sentence dictation task was slightly easier. By endline, students in the COVID-19 response group had demonstrated significant increases on the three- and four-word sentence tasks, and their average scores had overtaken those of the students in the overall intervention for all three tasks.

Table 12: Average item scores for the grade 2 dictation subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 4a | Three-word sentence | Overall intervention | 18.1% (2.4%) | 14.4% (2.5%) | –3.7 |
| COVID-19 response intervention | 10.7% (2.4%) | 19.2% (3.0%) | 8.6\* |
| 4b | Four-word sentence | Overall intervention | 8.3% (1.2%) | 6.6% (2.0%) | –1.7 |
| COVID-19 response intervention | 5.7% (1.2%) | 13.6% (2.6%) | 7.9\*\* |
| 4c | Four-word sentence, difficult words | Overall intervention | 1.9% (1.0%) | 4.3% (1.1%) | 2.3 |
| COVID-19 response intervention | 4.9% (1.0%) | 7.8% (1.8%) | 2.8 |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\**p* < .01.

**Subtask 5** assessed the listening comprehension ability of students. The teacher read a 25-word passage and asked three questions about it. The first question was in short-answer format and could be answered based on information provided explicitly in the first or second sentence of the paragraph. The second question’s answer was also found directly in the text. The third was an inferential question where students had to build answers from information in at least two sentences in the text. The items in the subtask were multiple choice, with five answer options, including one correct option and four distractors. The baseline and endline student stimulus is presented in ***Figure 16.***

Figure 16: Student stimulus for the grade 2 listening comprehension subtask

|  |
| --- |
| Text  Description automatically generated |
| Text  Description automatically generated |
| Table  Description automatically generated |

The score distribution for the grade 2 listening comprehension subtask (*Figure 17*) identified no consistent pattern of increases or decreases for the overall intervention. However, there was a substantial decrease in the percentage of students scoring zero in the COVID-19 response intervention between baseline and endline—dropping from two-fifths to under one-third of students scoring zero—combined with increases in the percentage of students scoring correctly one, two, or three items.

Figure 17: Distribution of scores for the grade 2 listening comprehension subtask

|  |
| --- |
|  |
|  |

***Table 13*** shows the change in the average item scores for the grade 2 listening comprehension subtask for the overall intervention and the COVID-19 response intervention. This analysis shows that all three items had generally the same level of difficulty at baseline. By endline, the percentage of students answering correctly in the COVID-19 response group had increased significantly for the two explicit, short-answer questions but not for the inferential question, although the trend was positive. There was a decrease in the percentage of students answering correctly for all three items in the overall intervention, although the differences were not significant.

Table 13: Average item scores for the grade 2 listening comprehension subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 5a | Short answer, explicit | Overall intervention | 54.2% (3.5%) | 53.9% (4.9%) | –0.3 |
| COVID-19 response intervention | 32.6% (3.5%) | 48.7% (4.1%) | 16.1\*\* |
| 5b | Short answer, explicit | Overall intervention | 49.8% (3.7%) | 44.6% (3.7%) | –5.2 |
| COVID-19 response intervention | 37.1% (3.7%) | 51.2% (4.3%) | 14.1\* |
| 5c | Inferential from at least two sentences | Overall intervention | 49.2% (3.4%) | 42.0% (5.5%) | –7.2 |
| COVID-19 response intervention | 34.6% (3.4%) | 45.0% (4.5%) | 10.4 |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\**p* < .01.

**Subtask 6** assessed reading comprehension ability. Students had to read a passage of 60 words and answer three questions based on the text. The first and second questions could be answered directly by referring to the text, and the third question was inferential and demanded that the student consider information from two or more sentences from the text. The items in the subtask were multiple choice. There were five answer options, with one correct option and four distractors. The student baseline and endline stimulus is presented in ***Figure 18.***

Figure 18: Student stimulus for the grade 2 reading comprehension subtask

|  |
| --- |
| Text, table  Description automatically generated |
| Table  Description automatically generated |

The score distribution for the grade 2 reading comprehension subtask (***Figure 19***) shows that around half of students in the overall intervention and the COVID-19 response intervention scored zero at baseline. There were decreases in zero scores for both groups by endline, combined with increases in the percentage of students answering one item correctly in the overall intervention, and answering two or three items correctly in the COVID-19 response intervention.

Figure 19: Distribution of scores for the grade 2 reading comprehension subtask

|  |
| --- |
|  |
|  |

***Table 14*** shows the change in the average item scores for the grade 2 reading comprehension subtask for the overall intervention and the COVID-19 response intervention. The average percentage correct on each item was around roughly one-quarter to one-third at baseline. By endline, on all three items there were small but statistically insignificant increases for the overall group, and significant increases for the COVID-19 response group.

Table 14: Average item scores for the grade 2 reading comprehension subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 6a | Short answer, explicit | Overall intervention | 20.4% (3.2%) | 26.5% (3.7%) | 6.2 |
| COVID-19 response intervention | 25.1% (3.2%) | 38.9% (4.3%) | 13.8\* |
| 6b | Short answer, explicit | Overall intervention | 29.1% (2.9%) | 30.5% (3.4%) | 1.4 |
| COVID-19 response intervention | 23.9% (2.9%) | 39.5% (4.0%) | 15.6\*\* |
| 6c | Inferential from two or more sentences | Overall intervention | 26.5% (2.8%) | 28.1% (3.8%) | 1.5 |
| COVID-19 response intervention | 25.1% (2.8%) | 42.0% (4.2%) | 16.9\*\* |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\**p* < .01.

**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 items in the subtask were multiple choice. There were four distractors in each item in addition to one correct option. ***Figure 20*** shows the baseline and endline student stimulus.

Figure 20: Student stimulus for the grade 2 calendar reading subtask

Calendar

Description automatically generated

The score distribution for the grade 2 calendar reading subtask in ***Figure 21*** shows that many students found the subtask difficult at baseline, but there were large decreases in the percentage of zero scorers in both groups between baseline and endline. By endline, nearly one-third of students in the COVID-19 response could answer all three questions correctly, up from 16.4% of students at baseline.

Figure 21: Distribution of scores for the grade 2 calendar reading subtask

|  |
| --- |
|  |
|  |

As shown in ***Table 15,*** for both interventions, the difficulty level of all three items in this subtask was similar. Both groups experienced significant increases in the percentage of correct scores for identifying the day and date and information about festivals, while there was a significant increase on counting the number of days in a month only for the COVID-19 response intervention.

Table 15: Average item scores for the grade 2 calendar reading subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 7a | Day and date | Overall intervention | 17.6% (3.8%) | 29.4% (3.7%) | 11.8\* |
| COVID-19 response intervention | 27.1% (3.8%) | 46.0% (4.4%) | 18.9\*\* |
| 7b | Festival and date | Overall intervention | 19.6% (3.6%) | 36.0% (3.6%) | 16.4\*\*\* |
| COVID-19 response intervention | 32.4% (3.6%) | 49.3% (3.8%) | 16.9\*\* |
| 7c | Number of specific days  (e.g., Saturdays) in a month | Overall intervention | 22.2% (3.8%) | 31.8% (3.8%) | 9.6 |
| COVID-19 response intervention | 35.7% (3.8%) | 54.1% (4.2%) | 18.4\*\* |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

### Reading Achievement by Sex and L1 (Grade 2)

Student reading achievement was disaggregated by the sex of students to understand whether scores varied between boys and girls in grade 2, as shown in ***Figure 22.*** There were no significant changes in the percentages of boys or girls in different proficiency levels between baseline and endline for the overall intervention. However, for the COVID-19 response intervention, there was a statistically significant decrease of 19.5 points in the percentage of boys in the low proficiency category and a statistically significant increase of 12.3 points in the percentage of boys in the emergent proficiency category. There was also a statistically significant decrease of 24.8 points in the percentage of girls in the low proficiency category and statistically significant increases in the percentages of girls in the emergent and fluent proficiency categories. These findings indicate benefits from the COVID-19 response intervention for both boys and girls, but suggest that there were greater positive outcomes for girls.

Figure 22: Changes in proficiency levels of grade 2 students, by sex

|  |
| --- |
|  |
|  |

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

Student reading achievement was also disaggregated by the home language of students to understand whether scores varied between Nepali L1 and L2 students in the overall intervention, as shown in ***Figure 23.*** The figureshows that there were no significant changes in outcomes for L1 students between baseline and endline, although there was an increase in the proportion of L1 students who were low readers. However, for L2 students, there was a significant decrease (–20 percentage points) in the percentage of students in the low reader category combined with significant increases in the emergent and fluent categories.

Figure 23: Changes in proficiency levels of grade 2 students, by language, overall intervention

|  |
| --- |
|  |

\*\**p* < .01.

## Grade 3 Findings

### Overall Reading Achievement (Grade 3)

For the overall intervention, on average, grade 3 students were able to answer nearly 7 out of 21 questions during the baseline and nearly 8 out of 21 questions at the endline. There was a statistically significant increase at endline for two of the subtasks: listening comprehension and calendar reading. Like the grade 2 results, at both baseline and endline, students performed best on word and sentence reading and listening comprehension, and had the most difficulty with the dictation subtask.

A breakdown of the average subtask scores for grade 3 students in the overall intervention at baseline and endline is presented in ***Figure 24.***

Figure 24: Change in average percentage scores by subtask in the overall intervention (grade 3)

\**p* < .05.

For the COVID-19 response intervention, at baseline, grade 3 students were able to answer an average of just over 6 out of 21 questions correctly, a figure that increased to just over 9 out of 21 questions at endline. There were significant increases in the average scores for all subtasks, with the largest increases for the dictation, word separation, and listening comprehension subtasks. Students experienced smaller increases on the vocabulary, calendar reading, and word and sentence reading subtasks.

A breakdown of the average scores for grade 3 students for each subtask at baseline and endline is presented in ***Figure 25***.

Figure 25: Change in average percentage scores by subtask in the COVID-19 response intervention (grade 3)

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

The distribution of composite scores in grade 3 is presented for the overall intervention and the COVID-19 response intervention in ***Figure 26.*** These distributions show some increases in the percentages of students scoring in the higher ranges between baseline and endline for the overall intervention, but a substantial movement of students in the COVID-19 response endline out of the lower distributions into higher distributions.

Figure 26: Distribution of overall average percentage scores (grade 3)

|  |
| --- |
|  |
|  |

### Reading Achievement by Subtask (Grade 3)

The following analysis provides details about the average grade 3 percentage scores for each of 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 baseline and endline student stimulus appears in ***Figure 27.***

Figure 27: Student stimulus for the grade 3 word and sentence reading subtask

Table

Description automatically generated

The score distributions for the subtask for both the overall intervention and the COVID-19 response intervention are presented in ***Figure 28***. The bars show that around one-quarter to one-third of students could not answer any questions at baseline. Subsequently, there was a small reduction in zero scores and an increase in the percentage of students who could answer all three questions at endline in the overall intervention. For the COVID-19 response intervention at endline, there was a larger reduction in zero scorers combined with a 2.8 point increase in the percentage of students who could answer two out of three questions correctly and a 10.9 point increase in the percentage of students who could correctly answer all three questions.

Figure 28: Distribution of scores for the grade 3 word and sentence reading subtask

|  |
| --- |
|  |
|  |

***Table 16*** shows the change in the average item scores for the grade 3 word and sentence reading subtask for the overall intervention and the COVID-19 response intervention. As this table demonstrates, there were statistically significant increases in the percentages of students answering correctly in all three tasks in the COVID-19 response intervention, and for the four-word sentence reading task in the overall intervention.

Table 16: Change in average item scores for the grade 3 word and sentence reading subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 1a | Identify word | Overall intervention | 62.7% (3.0%) | 66.3% (3.2%) | 3.6 |
| COVID-19 response intervention | 48.8% (3.0%) | 63.9% (3.1%) | 15.1\*\*\* |
| 1b | Identify four-word sentence | Overall intervention | 39.7% (2.9%) | 54.8% (4.6%) | 15.2\* |
| COVID-19 response intervention | 36.3% (2.9%) | 46.6% (3.8%) | 10.2\* |
| 1c | Identify five-word sentence | Overall intervention | 50.9% (2.9%) | 57.2% (3.4%) | 6.2 |
| COVID-19 response intervention | 41.5% (2.9%) | 50.9% (2.6%) | 9.4\* |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\*\**p* < .001.

**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 baseline and endline student stimulus is presented in ***Figure 29.***

Figure 29: Student stimulus for the grade 3 vocabulary subtask

Graphical user interface, table

Description automatically generated

The score distribution for the grade 3 vocabulary subtask ***(Figure 30)*** shows that a considerable percentage of students in both the overall intervention and the COVID-19 response intervention scored zero at baseline. By endline, there were limited changes in the distribution for the overall intervention group, but a substantial reduction in zero scores and increases in the percentages of students scoring one or more correct in the COVID-19 response group.

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

|  |
| --- |
|  |

***Table 17*** shows the change in the average item scores for the grade 3 vocabulary subtask for the overall intervention and the COVID-19 response intervention. Students in both groups generally found all three items in the subtask to be challenging. There were no significant changes in the percentages of students who answered any of the items correctly at endline in the overall intervention, although there were some slight increases for defining a word and for identifying antonyms. At the same time, average scores increased for the COVID-19 response group (significantly for identifying synonyms), enabling that group to overtake the average scores in the overall group across all three subtasks.

Table 17: Average item scores for the grade 3 vocabulary subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 2a | Define a word | Overall intervention | 28.9% (3.1%) | 29.9% (3.8%) | 1.0 |
| COVID-19 response intervention | 29.1% (3.1%) | 32.8% (3.5%) | 3.7 |
| 2b | Identify antonyms | Overall intervention | 23.6% (4.0%) | 31.3% (4.4%) | 7.7 |
| COVID-19 response intervention | 38.3% (4.0%) | 48.6% (3.9%) | 10.3 |
| 2c | Identify synonyms | Overall intervention | 29.2% (3.4%) | 27.2% (3.1%) | –2.0 |
| COVID-19 response intervention | 29.8% (3.4%) | 46.4% (3.3%) | 16.6\*\*\* |

*Note.* Standard errors in parentheses.  
\*\*\**p* < .001.

**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 baseline and endline student stimulus is presented in ***Figure 31.***

Figure 31: Student stimulus for the grade 3 word separation subtask

|  |
| --- |
| Text  Description automatically generated |
| A screenshot of a computer  Description automatically generated with low confidence |

The score distribution for the grade 3 word separation subtask ***(Figure 32***) shows that more than half of students in the overall intervention and nearly three-quarters of students in the COVID-19 response intervention scored zero at baseline. By endline, there were limited changes in the distribution for the overall intervention group, with a small decrease in the percentage of zero scores and an increase in the percentage of students answering one or three out of three questions correctly in the overall intervention. At the same time, there was a substantial reduction in zero scores and increases in the percentages of students scoring one or more correct in the COVID-19 response group.

Figure 32: Distribution of scores for the grade 3 word separation subtask

|  |
| --- |
|  |
|  |

***Table 18*** shows the change in the average item scores for the grade 3 word separation subtask for the overall intervention and the COVID-19 response intervention. Generally, students found the four- and five-word sentence separation tasks very difficult at baseline while the three-word task was somewhat easier. At endline, there were no significant increases on any of the items for the overall intervention, while there were large and significant increases on all three items for the COVID-19 response intervention.

Table 18: Average item scores for the grade 3 word separation subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 3a | Three-word sentence | Overall intervention | 41.8% (3.0%) | 44.6% (3.2%) | 2.8 |
| COVID-19 response intervention | 29.5% (3.0%) | 49.1% (4.4%) | 19.5\*\*\* |
| 3b | Four-word sentence | Overall intervention | 14.7% (3.5%) | 20.1% (3.1%) | 5.4 |
| COVID-19 response intervention | 17.6% (3.5%) | 34.4% (4.4%) | 16.8\*\* |
| 3c | Five-word sentence | Overall intervention | 14.8% (2.9%) | 20.3% (2.7%) | 5.5 |
| COVID-19 response intervention | 17.6% (2.9%) | 32.0% (4.0%) | 14.3\*\* |

Note. Standard errors in parentheses.  
\*\**p* < .01, \*\*\**p* < .001.

**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 33*** is the baseline and endline student stimulus for dictation.

Figure 33: Student stimulus for the grade 3 dictation subtask

|  |
| --- |
|  |
| Text  Description automatically generated |
| Table  Description automatically generated |

The score distribution for the grade 3 dictation subtask (***Figure 34***) shows that the majority of students scored zero at baseline and endline in both the overall intervention and the COVID‑19 response intervention. However, there were small increases in the percentages of students answering two or three out of the three items correctly in the overall intervention. There was a large reduction (–20.6 percentage points) in the percentage of zero scorers, with parallel increases in the percentages of students correctly answering two or three items correctly in the COVID-19 response intervention.

Figure 34: Distribution of scores for the grade 3 dictation subtask

|  |
| --- |
|  |
|  |

*Table 19* shows the change in the average item scores for the grade 3 dictation subtask for the overall intervention and the COVID-19 response intervention. At both time points, students in both groups found all items very difficult, although the three-word sentence dictation task was slightly easier for the overall intervention group. By endline, students in the COVID-19 response group had demonstrated large and significant increases on all three items, whereas there was only one statistically significant increase—on the four-word sentence item—for the overall intervention.

Table 19: Average item scores for the grade 3 dictation subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 4a | Three-word sentence | Overall intervention | 31.0% (2.6%) | 31.0% (4.0%) | 0.0 |
| COVID-19 response intervention | 19.9% (2.6%) | 41.0% (4.6%) | 21.2\*\*\* |
| 4b | Four-word sentence | Overall intervention | 10.0% (2.2%) | 18.8% (3.3%) | 8.7\* |
| COVID-19 response intervention | 13.2% (2.2%) | 31.5% (4.3%) | 18.3\*\*\* |
| 4c | Five-word sentence | Overall intervention | 4.7% (1.7%) | 6.7% (1.8%) | 1.9 |
| COVID-19 response intervention | 7.4% (1.7%) | 19.2% (2.9%) | 11.8\*\*\* |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\*\**p* <.001.

**Subtask 5** assessed the listening comprehension ability of students. The teacher read aloud 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 baseline and endline student stimulus is presented in ***Figure 35.***

Figure 35: Student stimulus for the grade 3 listening comprehension subtask

|  |
| --- |
|  |
|  |
| Text  Description automatically generated |
| Table  Description automatically generated |

The score distribution for the grade 3 listening comprehension subtask (***Figure 36***) identified no consistent pattern of increases or decreases for the overall intervention. However, there was a substantial decrease in the percentage of students scoring zero in the COVID-19 response intervention between baseline and endline—dropping from two-fifths to under one**-**third of students scoring zero—combined with increases in the percentages of students scoring one, two, or three items correctly.

Figure 36: Distribution of scores for the grade 3 listening comprehension subtask (30 words)

|  |
| --- |
|  |
|  |

***Table 20*** shows the change in the average item scores for the grade 3 listening comprehension subtask for the overall intervention and the COVID-19 response intervention. This analysis shows that all three items had generally the same level of difficulty at baseline. By endline, the percentage of students answering correctly in the COVID-19 response group increased significantly for all three items. There were decreases in the percentages of students answering correctly for all three items in the overall intervention, although the differences were not significant except for the second short answer item.

Table 20: Average item scores for the grade 3 listening comprehension subtask (30 words)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 5a | Short answer, explicit | Overall intervention | 64.3% (3.3%) | 73.3% (4.7%) | 8.9 |
| COVID-19 response intervention | 46.5% (3.3%) | 61.5% (4.2%) | 15.0\*\* |
| 5b | Short answer, explicit | Overall intervention | 56.0% (3.2%) | 69.7% (4.4%) | 13.7\* |
| COVID-19 response intervention | 41.4% (3.2%) | 56.2% (4.0%) | 14.8\*\* |
| 5c | Inferential from at least two sentences | Overall intervention | 44.0% (3.5%) | 51.8% (3.4%) | 7.8 |
| COVID-19 response intervention | 37.7% (3.5%) | 49.3% (4.0%) | 11.7\* |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\**p* < .01.

**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 37*** is the baseline and endline student stimulus for reading comprehension.

Figure 37: Student stimulus for the grade 3 reading comprehension subtask (60 words)

|  |
| --- |
| Text  Description automatically generated |
| Graphical user interface, application, table  Description automatically generated |

The score distribution for the grade 3 reading comprehension subtask (***Figure 38***) shows that roughly one-quarter to one-third of students in the overall intervention and the COVID-19 response intervention scored zero at baseline. There were decreases in zero scores for both groups by endline, combined with increases in the percentage of students answering one or three items correctly in the overall intervention, and answering two or three items correctly in the COVID-19 response intervention.

Figure 38: Distribution of scores for the grade 3 reading comprehension subtask

|  |
| --- |
|  |
|  |

***Table 21*** shows the change in the average item scores for the grade 3 reading comprehension subtask for the overall intervention and the COVID-19 response intervention. The average percentage correct on each item was around half at baseline for the two short-answer questions and lower for the inferential question. By endline, on all three items there were small but statistically insignificant increases for the overall group, and significant increases for the COVID-19 response group on the first short-answer question and on the inferential question.

Table 21: Average item scores for the grade 3 reading comprehension subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 6a | Short answer, explicit | Overall intervention | 53.4% (4.2%) | 61.1% (4.0%) | 7.8 |
| COVID-19 response intervention | 42.6% (4.2%) | 54.8% (3.7%) | 12.2\* |
| 6b | Short answer, explicit | Overall intervention | 54.2% (4.0%) | 57.7% (4.9%) | 3.5 |
| COVID-19 response intervention | 45.8% (4.0%) | 57.6% (5.5%) | 11.9 |
| 6c | Inferential from two or more sentences | Overall intervention | 27.0% (3.6%) | 31.9% (3.4%) | 4.9 |
| COVID-19 response intervention | 33.6% (3.6%) | 47.1% (4.1%) | 13.6\* |

*Note.* Standard errors in parentheses.  
\**p* < .05.

**Subtask 7** involved calendar reading. In this subtask, a month from the Nepali calendar was named 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 39*** shows the baseline and endline student stimulus.

Figure 39: Student stimulus for the grade 3 calendar reading subtask

Calendar

Description automatically generated

The score distribution for the grade 3 calendar reading subtask in ***Figure 40*** shows that a substantial percentage of students scored zero at baseline, but there were decreases in the percentage of zero scorers in both groups between baseline and endline, and especially large in the COVID-19 response group. By endline, more than one-quarter of students in the overall intervention could answer two or three items correctly, while there were increases in the percentages of students who could answer one, two, or three questions in the COVID-19 response.

Figure 40: Distribution of scores for the grade 3 calendar reading subtask

|  |
| --- |
|  |
|  |

As shown in ***Table 15,*** for both interventions, the difficulty level of all three items in this subtask was similar. Both groups experienced significant increases in the percentage of correct scores for identifying the date of the last day of the month, while there was a significant increase on identifying a festival and its date only for the COVID-19 response intervention. Neither group demonstrated significant increases between baseline and endline on identifying the last day of the previous month.

Table 22: Average item scores for the grade 3 calendar reading subtask

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subtask | Description | Group | Percentage of students who answered correctly | | |
| Baseline | Endline | Difference  (percentage points) |
| 7a | Day of last date of month | Overall intervention | 26.7% (4.1%) | 41.0% (3.8%) | 14.3\*\* |
| COVID-19 response intervention | 42.1% (4.1%) | 57.2% (.8%) | 15.1\*\* |
| 7b | Festival and date | Overall intervention | 23.8% (3.4%) | 28.8% (3.5%) | 5.0 |
| COVID-19 response intervention | 26.8% (3.4%) | 38.4% (4.1%) | 11.5\* |
| 7c | Last day of previous month | Overall intervention | 18.8% (3.4%) | 26.3% (3.7%) | 7.4 |
| COVID-19 response intervention | 30.1% (3.4%) | 34.7% (4.8%) | 4.7 |

*Note.* Standard errors in parentheses.  
\**p* < .05, \*\**p* < .01.

### Reading Achievement by Sex and Language (Grade 3)

Student reading achievement was disaggregated by the sex of students to understand whether scores varied between boys and girls in grade 3, as shown in ***Figure 41.*** Although there were some improvements between baseline and endline for both girls and boys in the overall intervention, the changes were not statistically significant. However, for the COVID-19 response intervention, there was a statistically significant decrease of 30.8 points in the percentage of boys in the low proficiency category and statistically significant increases in the percentages of boys the emergent and fluent proficiency categories. There was also statistically significant decrease of 24.1 points in the percentage of girls in the low proficiency category and statistically significant increases in the percentages of girls in the emergent and fluent proficiency categories. These findings indicate benefits from the COVID-19 response intervention for both boys and girls.

Figure 41: Changes in proficiency levels of grade 3 students, by sex

|  |
| --- |
|  |
|  |

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

***Figure 42*** shows that there were no significant changes in outcomes for grade 3 L1 students between baseline and endline. However, there was a large and significant reduction (–20 percentage points) in the percentage of low readers in the L2 category, combined with commensurate increases in the percentages of students in the emergent and fluent categories. Comparisons of outcomes by language cannot be made for the COVID-19 response intervention because that sample consisted entirely of L2 students.

Figure 42: Changes in proficiency levels of grade 3 students, by language, overall intervention

|  |
| --- |
|  |

\*\**p* < .01.

## Extrapolation of Reading Achievement

### Equivalent Scoring

As noted in Section 1.2, using statistical models, equivalent CB-EGRA scores were created for emergent and fluent reader oral reading fluency benchmarks (***Table 23***). These scores can be used to calculate the percentage of students at baseline meeting Nepal’s emergent and fluent benchmarks in line with EGRP II’s performance indicators, as described in ***Annex A.*** These scores will also become benchmark equivalencies for all future CB‑EGRAs. For further details about how the statistical equating between students’ scores on the CB-EGRA and their scores on the mini-EGRA, readers can refer to EGRP II’s *Baseline Report Vol. 2, COVID-19 Response: The Home- and Community-Based Schooling Intervention.*

Table 23: Equivalent CB-EGRA scores for emergent and fluent benchmarks (in cwpm)

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

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

### Comparing EGRP II Baseline and Endline Findings with the 2020 NARN and CB‑EGRA Scores from Previous Years

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

For example, EGRP II’s 2021 baseline, the 2022 endline, and the 2020 NARN 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 endline, the estimate of grade 3 students who met the reading benchmark of 45 cwpm with 80% comprehension was 20.1%, with 95% confidence that the true population percentage was between 15.4% and 24.8%. When comparing these values with the EGRP II baseline, the estimate for the baseline was 12.6% with a 95% confidence interval of between 7.9% and 17.3%. Similarly, for the 2020 NARN, the estimate for the NARN was 8.41% with a 95% confidence interval of between 6.8% and 10.0%.

***Figure 43*** demonstrates that the confidence intervals (the black lines at the end of each blue bar) for the EGRP II baseline and the endline estimates overlap. Therefore, while the two estimates have a difference of over 7.5 percentage points, it is not certain that the EGRP II endline percentage is higher than the baseline with any degree of statistical significance, due to the overlapping confidence intervals. Similarly, it is not certain that the EGRP II baseline percentage is 4 percentage points higher than the 2020 NARN, because the confidence intervals for the EGRP II baseline and the 2020 NARN estimates overlap. The confidence intervals for the EGRP II endline and the 2020 NARN estimates do not overlap however. Therefore, it is certain that the EGRP II endline percentage is higher than for the 2020 NARN. It can be concluded that average student performance as measured by the 2020 NARN and the EGRP II baseline was roughly similar but that the EGRP II endline showed progress since the 2020 NARN.

Figure 43: Percentage of grade 3 students who met the reading benchmark in the 2020 NARN, 2021 EGRP II baseline, and 2022 EGRP II endline

Another important caveat to keep in mind is that the NARN and the EGRP II baseline and endline were administered to different samples of schools and children. The assessments were also conducted in different years: the NARN in early 2020, before the COVID-19 pandemic; and the EGRP II baseline in the midst of the pandemic in early 2021; and the EGRP II endline in early 2022. Consequently, direct comparisons in the average scores should be interpreted with caution.

Use of the CB-EGRA to help teachers gauge children’s EGR skills is one of the core elements of the NEGRP. Annual rollout of CB-EGRAs began in 2017. Typical scores from previous years were substantially higher, on average, than the average scores from both the EGRP II baseline assessment in 2021 and endline in 2022. For instance, previous average grade 2 CB-EGRA scores ranged from 64% to 66%, while the average was 28.5% in the EGRP II baseline and 31.7% in the EGRP II endline. Similarly, average grade 3 CB-EGRA scores ranged from 66% to 68% in the past, with an average of 32.2% in the EGRP II baseline and 38.5% in the EGRP II endline. The differences are captured in ***Figure 44*** below.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CB-EGRAs from previous years:**  Grade 2: 64%–66%  Grade 3: 66%–68% |  | **2021 EGRP II baseline:**  ***Overall***  Grade 2: 28.5%  Grade 3: 32.2%  ***COVID-19 sample***  Grade 2: 24.8%  Grade 3: 29.3% |  | **2022 EGRP II endline:**  ***Overall***  Grade 2: 31.7%  Grade 3: 38.5%  ***COVID-19 sample***  Grade 2: 39.2%  Grade 3: 44.7% |

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 and 2021–2022 academic years, 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 and endline. 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 the both baseline and endline assessments 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 both the baseline and endline assessment.

### Reading Ability Categories and the National Early Grade Reading Benchmark

This section describes the performance of students in the EGRP II baseline and endline by using the fluent reading proficiency cut-off in the current national early grade reading benchmark and the reading proficiency categories established in the 2020 NARN.

In 2017, the GON set 45 correct words per minute with 80% comprehension as Nepal’s national reading benchmark (MOE 2017). In 2021 and 2022, the GON undertook a process to revise the benchmark to describe several categories of readers and better reflect the actual status of early grade learning in the country. The revised benchmarks were not yet approved by the GON at the time of preparing this report, and as such, the current benchmark is still used in this section.

Apart from the current benchmark, 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).

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

Table 24: Categories of readers, by grade

| Grade | Emergent reader | | Fluent reader (meets the benchmark of 45 cwpm) | |
| --- | --- | --- | --- | --- |
| Baseline | Endline | Baseline | Endline |
| 2 | 27.8% | 27.1% | 7.4% | 8.5% |
| 3 | 27.6% | 27.9% | 12.6% | 20.1% |

As shown in the table, there were only minor shifts in the percentages of students categorized as emergent readers between the EGRP II baseline and endline. There were also modest changes in the percentage of fluent readers who met the national benchmark. Specifically, 7.4% of grade 2 children and 12.6% of grade 3 children in the EGRP II baseline sample met the fluent reading benchmark. By endline, those figures had increased slightly for grade 2 (1.1 percentage points) and to a greater degree for grade 3 (7.5 percentage points).

# Summary and Conclusions

The study described in this analysis report was intended to investigate the program’s impact on student reading performance in the early grades over the program period. EGRP II established a baseline by assessing students’ reading performance in February and March 2021 and undertook an endline evaluation after a year in February and March 2022. For both studies, a scientific sampling technique was used to select 45 schools from seven districts. The government’s CB-EGRA tools for grades 2 and 3 were used for data collection. As a group-administered assessment, the CB-EGRA cannot measure ORF. However, standard indicators, such as under the Sustainable 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, students’ ORF and comprehension data were simultaneously collected, using a “mini-EGRA,” from all sampled schools, on a subsample basis during the baseline assessment. The team then developed a statistical model to equate the CB-EGRA scores with the mini-EGRA scores. This model was helpful for extrapolating the ORF and comprehension scores for the EGRP II baseline and endline studies. In addition, it will be useful to the GON at the national and subnational levels—including for district and palika officials—for identifying, reviewing, and reporting on key reading indicators, such as the number of children reaching the MOEST’s EGR current benchmark or the revised benchmarks, once approved.

The EGRP II endline study was designed to answer four research questions. The summary and conclusions from the study are presented as responses to each research question below.

Research Question 1: What is the program impact on the reading skills of students in grades 2 and 3?

The findings showed that reading achievements measured using the average percentage CB-EGRA score for grades 2 and 3 did not significantly improve from baseline to endline in the overall intervention. The average percentage CB-EGRA score for grade 2 was 28.5% at baseline and 28.7% at endline, and for grade 3 it was 32.2% at baseline and 37.9% at endline. For both grades, seven subtasks and 21 items were used to assess student reading ability. Thus, this finding means that on average, a child from grade 2 was able to correctly respond to about six items at both baseline and endline, and a child in grade 3 was able to correctly respond to about seven items at baseline and about eight items at endline. This study did not produce direct evidence to explain the minimal impact on the reading skills of students in grades 2 and 3, but school closures for more than 10 months in the 2020–2021 academic year and almost 4 months in the 2021–2022 academic year due to COVID-19 could be an important contributing factor.

On average, grade 2 students scored lowest on the dictation and reading comprehension subtasks, while they performed highest on letter/matra identification and listening comprehension in both the baseline and endline. Grade 2 students made significant progress on calendar reading (visual literacy), with the average score moving from 19.8% at baseline to 32.4% at endline.

Grade 3 students scored lowest on the vocabulary, word separation, and dictation and reading comprehension subtasks on average, while they performed highest on word and sentence reading and listening comprehension in both the baseline and endline. Grade 3 students made significant progress on listening comprehension, with the average score increasing from 54.8% at baseline to 64.9% at endline. There were also significant improvements on calendar reading (visual literacy), with the average score moving from 19.8% at baseline to 32.4% at endline.

Research Question 2: What is the program impact on the reading skills of boys and girls?

Reading performance was not significantly different between boys and girls at endline in the overall intervention in either grade 2 or grade 3. This leads to the conclusion that the overall intervention was not associated with differential impacts by sex.

However, in the COVID-19 response intervention, there were large and significant improvements for L2 boys and girls in both grades. In grade 2, there was a significant decrease in the percentage of L2 boys and girls in the low reading proficiency category (-19.5 percentage points and -24.8 percentage points, respectively). These decreases were combined with a significant increase (+12.3 percentage points) in the percentage of L2 boys in the emergent reader category and of girls in the emergent reader category (+15.2 percentage points) and the fluent reader category (+9.6 percentage points).

The outcomes were similar for the COVID-19 response intervention in grade 3. There was a significant decrease in the percentage of L2 boys and girls in the low reading proficiency category (-30.8 percentage points and -24.1 percentage points, respectively). These decreases were combined with significant increases in the percentage of L2 boys in the emergent reader category (+8.9 percentage points) and the fluent reader category (+21.8 percentage points). Likewise, there were significant increases in the percentages of girls in the emergent reader category (+13.1 percentage points) and the fluent reader category (+11.0 percentage points).

Research Question 3: What is the program impact for students who speak Nepali as an L1 and those who speak Nepali as an L2?

Students were categorized into two groups according to their home language. Students with Nepali as their home language (mother tongue) were categorized as L1 learners and students with languages other than Nepali as their home language were categorized as L2 learners. In the overall intervention, L2 students from grade 2 achieved an average CB-EGRA score of 18.4% at baseline and 27.5% at endline, a statistically significant increase of approximately 9 percentage points. In contrast, the average CB-EGRA score for grade 2 L1 students decreased by roughly 4 percentage points (baseline 33.3%, endline 29.2%), although the change was not statistically significant.

A similar result was found in grade 3, where the average CB-EGRA score of L2 students increased significantly by 13 percentage points (baseline 21.2%, endline 34.2%). Although the score for L1 learners in grade 3 increased by 2 percentage points (baseline 37.4%, endline 39.4%), that level of difference was not statistically significant.

As noted under research question 1, there was no statistically significant change in outcomes in grade 2 or 3 across the sample in the overall intervention. However, the evaluation findings show that there were statistically significant positive impacts for L2 learners in both grades. This boost was substantial enough to help close the gap in outcomes between L1 and L2 students by the endline despite the fact that the average scores were much lower for L2 students in the baseline.

Research Question 4: What is the value-added of the COVID-19 response intervention in Madhesh Province?

This evaluation identified remarkable “value added” impact at endline for students who participated in the COVID-19 response intervention in comparison to the overall intervention. The benefits of the community- and home-based schooling activity in Madhesh Province were evident in the significant improvements in student reading proficiency levels between the baseline and endline, with large percentages of students in grades 2 and 3 moving out of the low proficiency category and into the emergent and fluent levels. In addition, the significant improvements in average scores for each of the CB-EGRA sub-tasks in both grades suggest that the intervention was effective in boosting performance broadly across a set of skills viewed by the GON as integral to success in the early grades.

# Study Limitations

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

Sample Size and Representativeness

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

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

Assessment Method

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

Lack of Estimates for Nonreaders and Initial Readers

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

Equated Scores Are Estimates

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

References

Campbell, Laurie O., Glen W. Lambie, Claudia C. Sutter, Ashley R. Bickham, and Lindsay Pulse. 2018. *Measuring the Predictability of Istation Indicators of Progress Early Reading (ISIP-ER) Scores on Florida Standards Assessment (FSA) Scores.* Orlando, Florida: Morgridge International Reading Center, University of Central Florida. <https://www.istation.com/Content/downloads/studies/FSAReadingPredictability.pdf>.

College Board and ACT. 2018. *Guide to the 2018 ACT®/SAT® Concordance.* <https://collegereadiness.collegeboard.org/pdf/guide-2018-act-sat-concordance.pdf>.

Education Review Office (ERO). 2017. *Classroom-Based Early Grade Reading Assessment: A Framework.* Bhaktapur: ERO, MOEST, Government of Nepal. <https://www.ero.gov.np/upload_file/files/post/1595301167_1390229304_1587620785_1577711485_CBEGRA_-Framework-2017.pdf>.

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

Mathes, Patricia, Joseph Torgesen, and Jeannine Herron. 2016. *Istation’s Indicators of Progress (ISIP) Early Reading Technical Report: Computer Adaptive Testing System for Continuous Progress Monitoring of Reading Growth for Students Pre-K through Grade 3.* Dallas, Texas: Istation. <https://www.istation.com/Content/downloads/studies/er_technical_report.pdf>.

Muraki, Eiji, Catherine M. Hombo, and Yong-Won Lee. 2000. “Equating and Linking of Performance Assessments.” *Applied Psychological Measurement* 24, no. 4, 325–37. <https://doi.org/10.1177/01466210022031787>.

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

Neupane, Sagar Mani, Swadesh Maharjan, Buddhi Chaudhary, Simon King, Jodie Fonseca, and Wendi Ralaingita. 2021. *USAID’s Early Grade Reading Program II (EGRP II) in Nepal. Baseline Report Vol. 2, COVID-19 Response: The Home- and Community-Based Schooling Intervention.* Prepared for USAID under Cooperative Agreement No. 72036720CA00001. Research Triangle Park, North Carolina, USA: RTI International.

University of Oregon. n.d. DIBELS home page. Eugene, Oregon: DIBELS. <https://dibels.uoregon.edu/>.

Annex A: EGRP II MEL Indicator Reporting

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

Figure xxx:

|  |  |
| --- | --- |
| **IND 01\_ES. 1-1: Percent of learners targeted for United States Government assistance who attain a minimum grade-level proficiency in reading at the end of grade 2** | |
| **Baseline** | **Endline** |
| **Overall: 7.4%**  (Numerator: 24,394, Denominator: 328,929)  **Male: 7.7%**  (Numerator: 12,295, Denominator: 160,269)  **Female: 7.2%**  (Numerator: 12,099, Denominator: 168,660) | **Overall: 8.5%**  (Numerator: 27,861, Denominator: 328,817)  **Male: 7.5%**  (Numerator: 11,683, Denominator: 156,732)  **Female: 9.4%**  (Numerator: 16,178, Denominator: 172,085) |
| **IND 04\_Custom: Percent of grade 2 and 3 students classified as fluent readers using national benchmarks** | |
| **Baseline** | **Endline** |
| **Grade 2**  **Overall: 7.4%**  (Numerator: 24,394, Denominator: 328,929)  **Male: 7.7%**  (Numerator: 12,295, Denominator: 160,269)  **Female: 7.2%**  (Numerator: 12,099, Denominator: 168,660)  **Grade 3**  **Overall: 12.6%**  (Numerator: 42,045, Denominator: 333,969)  **Male: 10.5%**  (Numerator: 16,454, Denominator: 156,849)  **Female: 14.4%**  (Numerator: 25,591, Denominator: 177,120) | **Grade 2**  **Overall: 8.5%**  (Numerator: 27,861, Denominator: 328,817)  **Male: 7.5%**  (Numerator: 11,683, Denominator: 156,732)  **Female: 9.4%**  (Numerator: 16,178, Denominator: 172,085)  **Grade 3**  **Overall: 20.1%**  (Numerator:, Denominator: 307,874)  **Male: 15.8%**  (Numerator: 21,637, Denominator: 137,165)  **Female: 23.6%**  (Numerator: 40,278, Denominator: 170,709) |
| **IND 05\_Custom: Percent of grade 2 and 3 students classified as emergent readers using national benchmarks** | |
| **Baseline** | **Endline** |
| **Grade 2**  **Overall: 27.8%**  (Numerator: 91,562, Denominator: 328,929)  **Male: 25.5%**  (Numerator: 40,870, Denominator: 160,269)  **Female: 30.1%**  (Numerator: 50,692, Denominator: 168,660)  **Grade 3**  **Overall: 29.7%**  (Numerator: 99,208, Denominator: 333,969)  **Male: 28.2%**  (Numerator: 44,270, Denominator: 156,849)  **Female: 31.0%**  (Numerator: 54,938, Denominator: 177,120) | **Grade 2**  **Overall: 27.1%**  (Numerator: 89,251, Denominator: 307,874)  **Male: 27.7%**  (Numerator: 43,383, Denominator: 156,732)  **Female: 26.7%**  (Numerator: 45,867, Denominator: 172,085)  **Grade 3**  **Overall: 27.9%**  (Numerator: 85,918, Denominator: 307,874)  **Male: 29.1%**  (Numerator: 39,972, Denominator: 137,165)  **Female: 26.9%**  (Numerator: 45,946, Denominator: 170,709) |
| **IND 18\_IR 4\_Custom Percent of learners who attain a minimum grade-level proficiency in reading at the end of grade 3 in targeted 32 local governments of Madhesh Province** | |
| **Baseline** | **Endline** |
| **Grade 3**  **Overall: 13.4%**  (Numerator: 1,682, Denominator: 12,519)  **Male: 11.9%**  (Numerator: 657, Denominator: 5,501)  **Female: 14.6%**  (Numerator: 1,025, Denominator: 7,018) | **Grade 3**  **Overall: 29.0%**  (Numerator: 3,874, Denominator: 13,363)  **Male: 33.7%**  (Numerator: 1,868, Denominator: 5,537)  **Female: 25.6%**  (Numerator: 2,006, Denominator: 7,826) |

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. EGRP assisted the GON in developing the minimum package. [↑](#footnote-ref-2)
2. Previous progress reports referred to this province as Province 2. However, in January 2022, Province 2 was officially renamed to Madhesh Province. [↑](#footnote-ref-3)
3. 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. [↑](#footnote-ref-4)
4. Difference in differences is a quasi-experimental analysis that enables measurement of changes in outcomes for two different groups over time. The approach entails subtracting pre-test scores from post-test scores for two groups to obtain a difference score. Then, the difference score for one group is subtracted from the difference score for the other group to arrive at the difference in differences. [↑](#footnote-ref-5)
5. The ERO subject committee, in consultation with subject experts from Nepali universities, allocated different weights to the subtasks as presented in Table 7 and Table 8 above, based upon the difficulty level. The main purpose of the weighting was to calculate overall reading achievement by using weights for all of the subtasks. [↑](#footnote-ref-6)