World Change Starts with Educated Children.®

Impact Evaluation of the Literacy Program in Nepal

2019 Endline (End-of-Grade 2) Evaluation Report

Pinaki Joddar

May 2019

In 2017, Room to Read began a two-year (2017-19) impact evaluation study of its Literacy Program in three districts (Nuwakot, Palpa and Tanahun) of Nepal where the Program was launched at the beginning of the 2017-18 academic year. The evaluation included 30 schools benefiting from the Literacy Program (defined as project group) and 30 matching non-intervention schools not benefiting from the Literacy Program (defined as comparison group). The aims of this evaluation were (i) estimate the extent of impact (i.e. effect size) of the Program on reading and writing skills of children in the project schools in the three districts after two years of intervention, and (ii) establish a direct causal relationship between the estimated impact on reading and writing skills and the Program. Room to Read collected data on reading and writing skills in project schools at the time of the launch of the Program from start-of-Grade 1 children in May-June 2017 (baseline) and after two years of implementation from end-of-Grade 2 children in March 2019 (endline); data from comparison school children was also collected following the same timeline. Results from this study show that the Literacy Program in Nepal is having a large positive impact on children's reading and writing skills. Children from project schools in the three districts included in the evaluation performed significantly better than the children from comparison schools at endline. Moreover, children in Literacy Program schools in Nepal experienced two-year gains in reading and writing skills that were two to three times larger than those experienced by comparison school children across all five assessment tasks. Also, it is evident from the results that the Literacy Program in Nepal benefitted both boys and girls. A key concern for program will be to reach those children who are still performing at low level.

This page is left blank intentionally

Table of Contents

Executive Summary

Room to Read's Literacy Program in Nepal is a school-based intervention that seeks to develop children's reading skills and habit or reading in the early primary grades to help them become life-long, independent readers. The Program includes three main components: literacy instruction, school libraries, and the development of quality reading materials. For the instruction component, the Literacy Program team in Nepal worked with the Ministry of Education to develop Grade 1 and Grade 2 teaching and learning materials that introduce phonics content at an appropriate pace and promote the development of key reading and writing skills. Room to Read's curriculum provides a strong foundation for all children in Room to Read schools, with the goal that children will become fluent readers by the end of Grade 2. The three components of the Literacy Program strongly complement each other and together promote a holistic approach to literacy development, thus ensuring that children develop reading skills and habit of reading. In Nepal, the instruction component of the Program was launched during the 2009 school year in Grade 1 classrooms. As of March 2019, the comprehensive Program was operating in 265 government schools in five districts of Nepal.

In 2017, Room to Read Nepal began a two-year (2017-19) internal study of its Literacy Program operating in the three districts (Nuwakot, Palpa and Tanahun) to determine the impact of the intervention on children's reading and writing skills. The evaluation included 30 schools benefiting from the Literacy Program (defined as project group), selected randomly from all project schools in proportion to their distribution across the districts. The evaluation also included 30 matching non-intervention schools in the same three districts (defined as comparison group) not benefiting from the intervention; the comparison group schools were selected randomly from a list of potential comparison schools in the same three districts based on pre-decided school matching criteria (with the project group). Under this evaluation, Room to Read assessed children's reading and writing skills using a version of the the Early Grade Reading Assessment (EGRA) that was adapted to Nepali by local experts. Using a quasi-experimental impact evaluation methodology and following a semi-longitudinal design, this study intended to assess the impact of the Program by comparing the two-year gains in reading and writing skills of the project and comparison school children from the start of Grade 1 (baseline) to the end of grade 2 (endline).

In May and June 2017, Room to Read Nepal conducted the baseline evaluation and found that children in both program and comparison schools entered Grade 1 with low reading levels, though the comparison school children demonstrated better skills than their counterparts in project schools. In March 2019, Room to Read conducted end-of-Grade 2 endline evaluation in the same schools with a freshly drawn randomly selected sample of children to understand the impact of the Program on children's reading and writing skills after two years of intervention. **Results from this 2019 endline evaluation show that the Literacy Program in Nepal is having a large positive impact on children's reading and writing skills. Children from project schools in the three districts included in the evaluation performed significantly better than the children from comparison schools at endline. Moreover, children in Literacy Program schools in Nepal experienced two-year gains in reading and writing skills that were two to three times larger than those experienced by comparison school children across all EGRA subtasks. Also, it is evident from the results that the Literacy Program in Nepal benefitted both boys and girls.** However, results of this two-year impact study also show that there exists scope for improvements as few children were still struggling with reading and writing after two years of intervention in the project schools. And Room to Read is committed to do so by continuing its support to the schools in Nepal in collaboration with the government and refining the content and delivery of the Program to improve reading and writing skills of all early grade children in Nepal.

1. Introduction

1.1 Literacy Program in Nepal

Room to Read's Literacy Program (also referred to as the Program in this document) in Nepal is a school-based intervention that seeks to develop children's reading skills and habit or reading in the early primary grades to help them become life-long, independent readers. The Program includes three main components: (i) *Instruction*: reading and writing instruction for children in Grades 1 and 2; (ii) *Library*: establishment of child-friendly school libraries to provide children with access to quality materials to read inside and outside schools; and (iii) *Quality Reading Materials* (QRM): development of locally appropriate quality reading materials for children. For the instruction component, the Literacy Program team in Nepal worked with the Ministry of Education to develop Grade 1 and Grade 2 teaching and learning materials that introduce phonics content at an appropriate pace and promote the development of key reading and writing skills. The Program also includes detailed lesson plans, classroom materials, and comprehensive teacher professional development which is complemented by concurrent classroom-based mentoring and coaching for the teachers by Room to Read appointed Literacy Coaches/ Facilitators and monitoring and evaluation of the intervention. Room to Read's curriculum provides a strong foundation for all children in Room to Read schools, with the goal that children will become fluent readers by the end of Grade 2. In Nepal, the instruction component of the Program was launched during the 2009 school year in Grade 1 classrooms. As of March 2019, the comprehensive Literacy Program was operating in 265 government schools in five districts (Nuwakot, Palpa, Salyan, Syangja and Tanahun) of Nepal.

1.2 Evaluation of the Literacy Program

To determine the effect of the intervention on children's reading and writing proficiencies, Room to Read initiated a two-year (2017-19) impact evaluation study of its Literacy Program that was launched in 125 government school in 2017-18 academic year in three districts (Nuwakot, Palpa, and Tanahun) of Nepal. The objectives of the impact evaluation are:

- Determine whether the Program is having an impact on children's reading and writing skills;
- Determine whether the implementation of the Program facilitates the acquisition of early reading skills in children at a rate that ensures that they will reach the goal of becoming fluent readers by the end of Grade 2; and
- Identify reading and writing skills that could be better supported by the Program and determine how to improve these reading skills quickly and effectively.

Room to Read with support from independent data collectors conducted this impact evaluation following the methodology as described below.

1.3 Evaluation Methodology

Room to Read is implementing the Program in the three districts in a set of government schools that were selected non-randomly using pre-decided school selection criteria in consultation with the government. As a result, traditionally used experimental methods for impact evaluation like randomized control trial was not deployed here. This two-year study employed a quasi-experimental impact evaluation methodology. It included collection of reading and writing skills data at different points in time from (i) children in 30 government schools benefiting from the Program (defined as *project* group); and (ii) children in equal number of matching schools not benefiting from the Program but have observable pre-intervention background characteristics similar to the schools in the project group (defined as *comparison* group). Following a semi-longitudinal study design, it was planned that data

would be collected from different cross-sections of children studying in the project and comparison schools included in the evaluation over two academic years[1] at three points in time as shown below.

Baseline (May-June 2017)	Midline (March 2018)	Endline (March 2019)
Assessed reading and writing skills of a sample of children at the start of Grade 1 in both project and comparison schools at the beginning of the school year 2017-18 at the time of the launch of the Program	Assessed reading and writing skills of a sample of children at the end of Grade 1 in the same project and comparison schools at the end of the school year 2017-18 after one year of implementation of the Program in the project schools	Assessed reading skills of a sample of children at the end of Grade 2 in the same project and comparison schools at the end of the school year 2018-19 after two years of implementation of the Program in the project schools

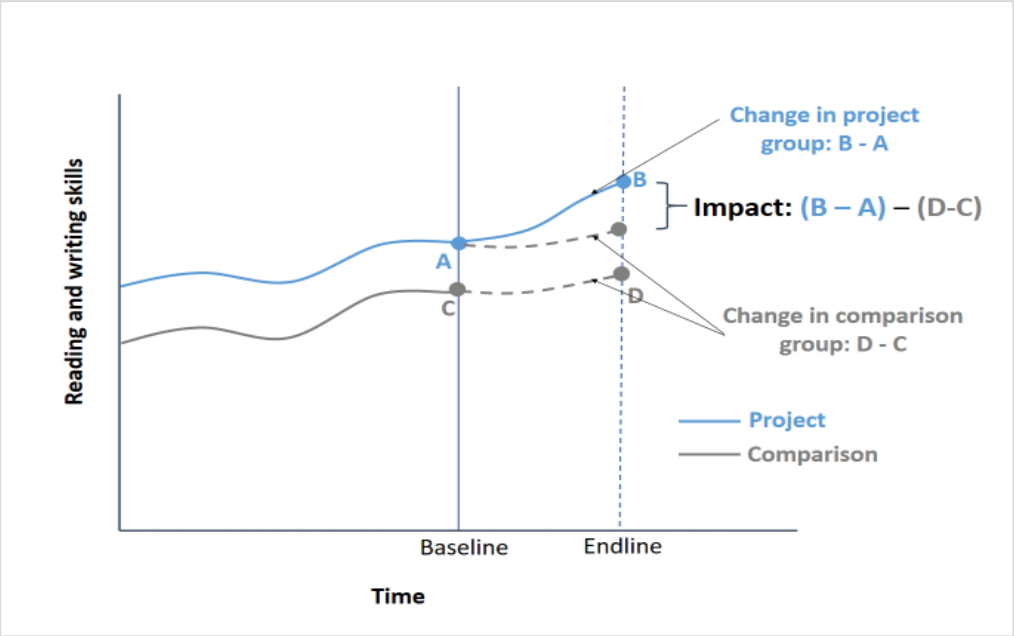
Using the these three waves of data, the quasi-experimental methodology allowed Room to Read to estimate the impact of the Program through regression analysis by comparing the difference between the following two categories after controlling for various school- and child-specific background characteristics that may influence their reading and writing skills:

- ***Category 1 (first difference)***: Changes in reading and writing skills of children from baseline to endline (or midline) in the sample project schools included in the study after two years (or one year in case of midline) of program intervention; and
- ***Category 2 (second difference)***: Changes in reading and writing skills of children from baseline to endline (or midline) in the sample non-intervention comparison schools included in the study following same timeline as above.

Comparison of the difference between these two categories enabled Room to Read to statistically estimate the extent of influence (measured through effect size) of the Program on reading and writing skills of children in project schools. Also, very importantly, this helped Room to Read to establish a direct causal relationship between the extent of influence and the Literacy Program intervention in the three districts. In impact evaluation literature, this methodology is defined as "Difference-in-difference (DID)" or "Double difference (DD)" and a simple graphical presentation of the same is provided below. Additionally, the evaluation results would help in understanding children's reading and writing skill strengths and weaknesses and provide Room to Read, classroom teachers, school administrators, and the government with information on program effect.

[1] During each data collection point (i.e. baseline, midline and endline), Room to Read planned to conduct assessment with a fresh sample of randomly selected children from appropriate grade in project and comparison schools. Though a particular child may get selected at multiple data collection points, Room to Read did not intentionally track the same children over the two years of this study.

Figure 1.1: Difference-in-Difference Methodology used in this Impact Evaluation



Following the methodology described above, Room to Read conducted the baseline evaluation in May-June 2017 with a sample of start-of-Grade 1 children in 30 project schools (seven from Nuwakot, six from Palpa and 17 from Tanahun). Simultaneously, for contrast, baseline data was also collected from a sample of start-of-Grade 1 children in 30 comparison schools; comparison schools were distributed similarly across districts as the project schools. Baseline results showed that children in both program and comparison schools entered Grade 1 with low reading levels, though children from comparison schools performed better than children from project schools.

In March 2018, Room to Read conducted end-of-Grade 1 midline evaluation in the same schools with a freshly drawn randomly selected sample of children to understand the impact of the Program on children's reading and writing skills after one year of intervention. The impact of the Program was assessed by comparing the changes in reading and writing skills of the children in project schools between baseline and midline with the changes in reading and writing skills of the children in comparison schools between baseline and midline. The midline evaluation results indicated that children in Literacy Program schools experienced higher and statistically significant gains in skills than children in comparison schools from the beginning of Grade 1 to the end of Grade 1.

Finally, in March 2019, Room to Read conducted end-of-Grade 2 the endline evaluation in the same schools with a freshly drawn randomly selected sample of children to understand the impact of the Program on children's reading and writing skills after two years of intervention. The results of this final round of evaluation is presented in *Section 2: Results.* (See *Appendix A* for a full description of the data analysis process used to derive the evaluation results.)

1.4 Sampling

Schools were the primary sampling units (PSU) in this two-year (2017-19) impact evaluation study. Through power calculations, it was determined that a sample size of 60 schools in Nepal (30 units each from the project and comparison groups) with 15 children per school would be sufficient to detect a minimum effect size of 0.25 from the evaluation (assuming α to be 0.05, power to be 0.9, baseline-endline correlation to be 0.50, and an intra-class correlation of 0.05). Consequently, the aim of the sampling was to select required number of project and

comparison schools for the evaluation that were as similar as possible at the time of the introduction of the Program in the three districts. Following steps were taken to select this sample.

First, a list of all project schools in the three districts (Nuwakot, Palpa and Tanahun) where the Program was launched in 2017-18 school year was assembled and 30 project schools were selected randomly in proportion to the number of intervention schools across districts. For comparison group sampling, a list of potential comparison schools across the same three districts that met the general selection criteria for identifying the Program intervention unit was prepared. Among these criteria were school size (enrollment in Grades 1 and 2 and all primary grades) and percentage of children in the schools who speak Nepali as their home language. Followed by this, a total 30 comparison schools, distributed as per the same proportion of the 30 project schools across three districts, were selected randomly.

During 2019 end-of-Grade 2 endline evaluation, data on observable school background characteristics was collected and analyzed to assess the comparability of the schools from the project and comparison groups.[2] Analyses of these school background characteristics like attendance rate in Grade 2 on the day of assessment, regular access to drinking water in the schools, availability of separate toilets for boys and girls, and whether schools running any student feeding program (as observed during the visits) did not reveal any statistically significant differences between project and comparison groups. However, analyses also showed statistically significant differences between project and comparison schools on the following: enrollment in Grade 2 ($p < 0.05$), enrollment in all primary grades ($p < 0.01$), total number of teachers in primary grades ($p < 0.05$), pupil teacher ratio in primary grades ($p < 0.05$), number of parent teacher meetings (PTM) conducted in current school year ($p < 0.05$), and availability of a functional library ($p < 0.01$). Statistical comparisons of reading and writing skills assessment results between project and comparison groups presented in this report took into account these differences in the school level characteristics.

Children are the final sampling units (FSU) in this two-year (2017-19) impact evaluation study. It was planned that during data collection, attempts would be made to select 15 children randomly from each school for assessment who:

- Do not have physical, sensory and significant cognitive disabilities[3]; and
- Present on the day(s) of data collection.

Also, it was decided that all children present in the appropriate grade in a school would be selected for assessment, if attendance in the grade on the day of the visit is less than 15. Moreover, assessors were instructed to not to deny a chance to any child to participate in the assessment if s/he expresses willingness. However, due to various challenges faced during data collection, the child sample size was lower than what was planned originally. The selection procedure as described here yielded a sample of 536 children (343 from projects schools and 193 from comparison schools) during the start-of-Grade 1 baseline in May-June 2017 and 537 children (343 in project schools and 194 in comparison schools) during the end-of-Grade 2 endline in March 2019.

As comparability of the children from project and comparison schools is also vital for this evaluation, during the 2019 endline evaluation data on the following child-background characteristics was collected and analyzed: age,

[2] Data on observable school background characteristics was also collected and analyzed during 2017 start-of-Grade 1 baseline and 2018 end-of-Grade 1 midline evaluations respectively.

[3] During data collection, it was not attempted to identify and/or exclude children with learning and/or reading and writing disabilities as such disabilities are difficult to detect in Grades 1 and 2.

gender, whether child speaks Nepali at home, whether child attended Early Child Development (ECD) classes, whether child's mother and father can read and write, whether child has a collection of books at home for reading[4], and whether child's family has radio, television, motorcycle and mobile phone at home. Analyses revealed that difference between the children in project and comparison schools included in the evaluation across any of these background characteristics was not statistically significant.

(See Tables B.1 and B.2 in *Appendix B* for more details about school and child background characteristics in project and comparison schools.)

1.5 Tool used for Assessment

In this impact evaluation, Room to Read assessed children's reading and writing skills using a version of the Early Grade Reading Assessment (EGRA) that was adapted from English to Nepali by local experts. Room to Read used a version of the EGRA that was designed according to the expected reading levels at Grade 2. The EGRA was comprised of five common subtasks:

- *Letter sound identification*: Ability to recognize the graphemic features of each letter and accurately map it to its corresponding name and sound to read letters of the alphabet without hesitation and naturally. Here, children read up to 100 letters in Nepali alphabet. This was a timed test and children read as many letters as possible in 60 seconds and data was analyzed to calculate *letter sounding fluency* scores.
- *Non-word reading*: Ability to read words that do not exist, but whose letter combinations follow the rules of the language. This task assesses the child's ability to "decode" words fluently as distinct from their ability to recognize words they have seen before. Children were given 60 seconds to read 50 non-words and data was analyzed to calculate *non-word reading fluency* scores.
- *Oral reading fluency*: Ability to read a passage that tells a story. Children were given 180 seconds to read a 60 word long passage of Grade 2 level of difficulty in reading and data was analyzed to calculate *oral reading fluency* scores. For oral reading fluency test, three different version of Nepali reading passages, each having 60 words with similar level of difficulty in reading, were used randomly across children.
- *Reading comprehension:* Ability to answer up to five questions (depending on the length of the passage read in the oral passage reading subtask) based on the oral reading passage.
- *Sentence dictation:* Ability to write words and sentences. Children wrote an 8-word long Nepali sentence that was read orally by an assessor. The sentence was scored on a 39-point scale based on correct spelling, space between each words, correct matra, correct dika, and full stop at the end.

Assessments were administered individually by external data collectors who were hired and trained by Room to Read.[5]

[4] Collection of books is defined as five or more books that are not school textbooks or any book read by senior siblings or adult members in the family.

[5] Room to Read Nepal's Research Monitoring and Evaluation (RM&E) team conducted the selection and a 5-day training workshop of the assessors for data collection. The training involved desk-based training on how to conduct EGRA and collect other relevant data from children and schools authorities and field practice in schools that were not included in the evaluation. During the evaluation, EGRA data was collected using tablets with the help of Tangerine software. During actual data collection, EGRA was conducted with each child individually by an assessor in the schools and in each school the entire process was overseen by dedicated supervisors and data quality assurance process was done by the Room to Read Nepal's RM&E team separately.

2. Results

2.1 Assessment Scores

Table 2.1 below provides an overview of the start-of-Grade 1 baseline (May-June 2017) and end-of-Grade 2 endline (March 2019) evaluation results by project and comparison groups across each assessment subtask. **Overall, children from project schools performed significantly better than children from comparison schools at endline.** By the end of Grade 2 children from project schools could read 54 letters per minute with correct sound, while children from comparison group could only read 29 letters per minute with correct sound. Also, project school children in endline assessment could correctly read an average of 29 words per minute, while children from comparison group could correctly read only 12 words per minute. On reading comprehension, children from project schools, on average, could answer three questions correctly (out of five), compared only two questions answered correctly by comparison school children. On sentence dictation subtask, average score of the Grade 2 children in project schools was 22 (out of 39), whereas children in comparison school, on average, scored only 15. **Moreover, children in Literacy Program schools in Nepal experienced two-year gains in reading and writing skills that were *two to three times* larger than those experienced by comparison school children across all five assessment subtasks.** Maximum two-year gains for the project school children was observed on letter sounding fluency, followed by oral reading fluency, non-word reading fluency, sentence dictation and reading comprehension respectively.

TABLE 2.1: Average Assessment Scores across Subtasks in Baseline and Endline and Impact of the Program

Assessment Subtask (*unit of measurement*)	Group	Baseline 2017 (Start-of-Grade 1)			Endline 2019 (End-of-Grade 2)			2-Year Gains in Skills	Impact (Diff. in 2-year Gains in Skills b/w Project & Comparison Groups)†
		n	Mean	SD	n	Mean	SD		
Letter sounding fluency (*correct letters per minute*)	Project	343	8.2	8.5	343	54.0	18.7	+45.8	+28.5***
	Comparison	193	12.7	14.6	193	29.4	17.4	+16.7	
Non-word reading fluency (*correct non-words per minute*)	Project	343	1.3	3.0	343	19.7	9.7	+18.5	+12.0***
	Comparison	193	3.1	5.2	193	9.4	9.1	+6.4	
Oral reading fluency (*correct words per minute*)	Project	343	0.3	2.0	343	29.1	18.5	+28.8	+19.1***
	Comparison	193	2.6	6.2	193	12.0	15.7	+9.4	
Reading comprehension (*Qs answered correctly out of five*)	Project	343	0.0	0.3	343	2.9	1.5	+2.9	+1.7***
	Comparison	193	0.4	1.0	193	1.5	1.7	+1.2	
Sentence Dictation (*score on a 39 point scale*)	Project	343	0.8	2.8	343	22.4	11.4	+21.6	+10.4***
	Comparison	193	3.7	7.8	193	14.5	13.0	+10.8	

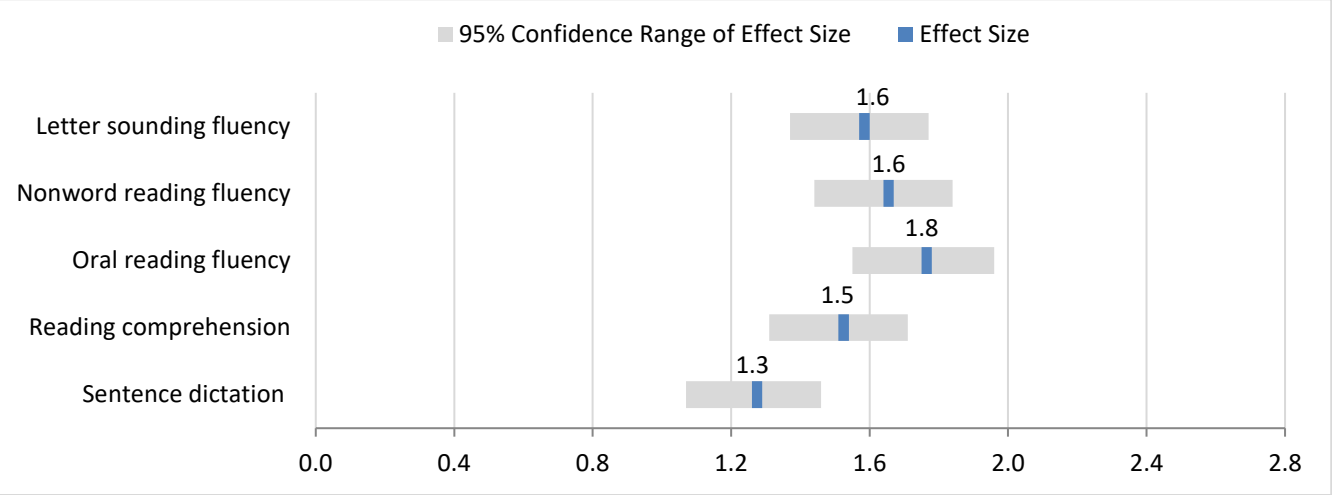
Legend of statistical significance of differences between project and comparison schools: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

†*Impact (Diff. in 2-year Gains in Skills b/w Project & Comparison Groups)*: Reports the estimated difference for each subtask between the (i) gains in reading or writing skills of children in project schools from baseline to endline after two years of intervention and (ii) gains in reading or writing skills of children in comparison schools during the same timeframe without the intervention. Each value presented in this column is the coefficient of the regression analysis conducted with random effects at the school level and age, gender, enrollment in target grades (Grade 1 at baseline and Grade 2 at endline), number of teachers in primary grade, pupil teacher ratio in primary grade, number of PTMs conducted in current school year, whether school has a functional library, and district (with Nuwakot as base) as covariates. Additionally, version of passage used during the tests were included in the regression analysis for reporting results for oral reading fluency and reading comprehension, respectively.

Differences in child gains are further examined by looking at the effect sizes for the Literacy Program across the assessment subtasks (see Figure 2.1). The effect size statistic is used to make comparisons across measures that use different scales or units. For the purposes of this analysis, the standardized mean effect size statistic was used, through which an effect size of 0.8 or higher is considered large. Effect size for a subtask was calculated by

determining the adjusted (i.e. after controlling for various influencing factors or covariates) difference in two-year gains in reading or writing skills between project and comparison school children through linear regression analysis and then dividing the difference by the adjusted pooled standard deviation of the project and comparison groups at the end of Grade 2 (see *Appendix A: Data Analysis*). **The effect sizes for the Literacy Program intervention in Nepal were large across all five assessment subtasks.** The largest effect size was found in case of oral reading fluency (*1.8*), followed by non-word reading fluency (*1.6*), letter sounding fluency (*1.6*), reading comprehension (*1.5*) and sentence dictation (*1.3*) respectively. These results suggest that the Literacy Program in Nepal was particularly effective at building children's reading fluency and comprehension skills.

FIGURE 2.1: Effect Sizes across Assessment Subtasks



2.2 Fluency and Comprehension Benchmarks

Although there has been limited research into fluency in Nepali language, independent studies in multiple countries have shown that children at the end of Grade 2 need to reach a fluency rate of roughly 45-60 words read correctly per minute (or local-language equivalent) as a prerequisite for reading with comprehension (Abadzi, 2011). Room to Read aims for children to reach this fluency level by the end of Grade 2 in Nepal and tracks children's progress towards this fluency rate through a "fluent reader" benchmark (correctly reading at least 45 words per minute).

The distribution of oral reading fluency by groups at 2019 end-of-Grade 2 endline is presented in Figure 2.2 below. This is indicative of how the Program has progressed to achieve the fluent reader goal after two years of intervention. **About 20 percent of the children from project schools included in the evaluation met or exceeded the fluent reader benchmark of correctly reading at least 45 words per minute by the end of Grade 2 (compared to just seven percent for comparison school children)**. The difference between the children from project and comparison schools in achieving the fluent reader benchmark was statistically significant ($p < 0.05$).

Other benchmark indicators that reflect progress include the proportion of children answering at least four questions out of five (i.e. at least 80 percent of the questions) correctly on the reading comprehension subtask. As shown in Figure 2.3 below, **42 percent of the Grade 2 children from project schools included in the evaluation could correctly answer at least 80 percent of the comprehension questions at endline.** This is significantly higher than the proportion of children in comparison school (17 percent) who could do so ($p < 0.01$).

FIGURE 2.2: Distribution of Children by Oral Reading Fluency Groups at Endline 2019

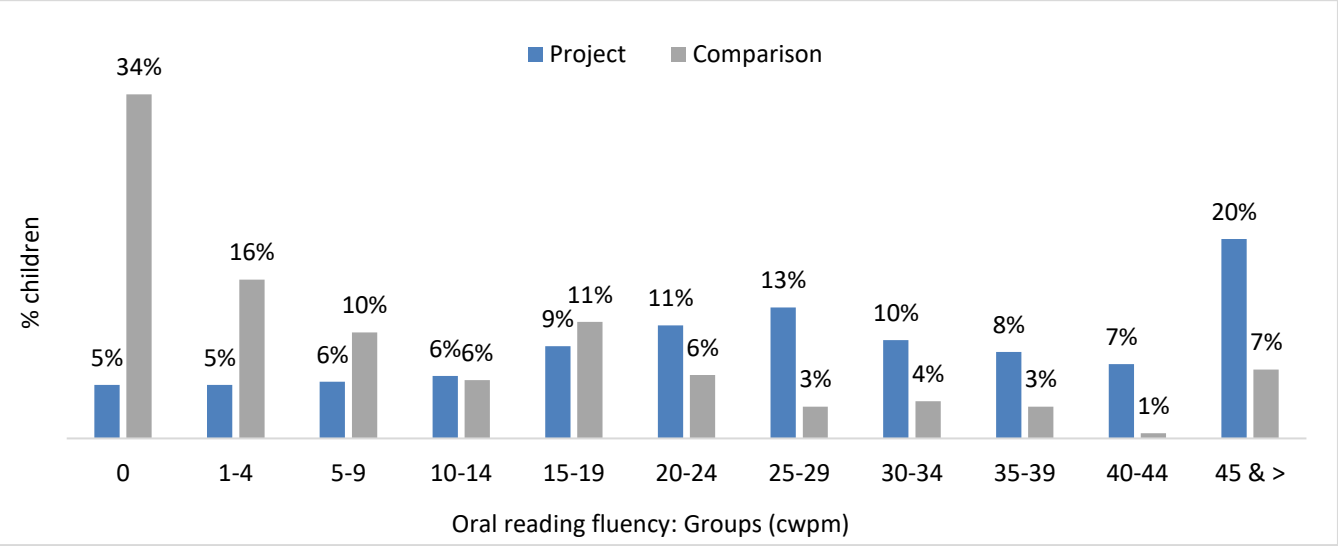
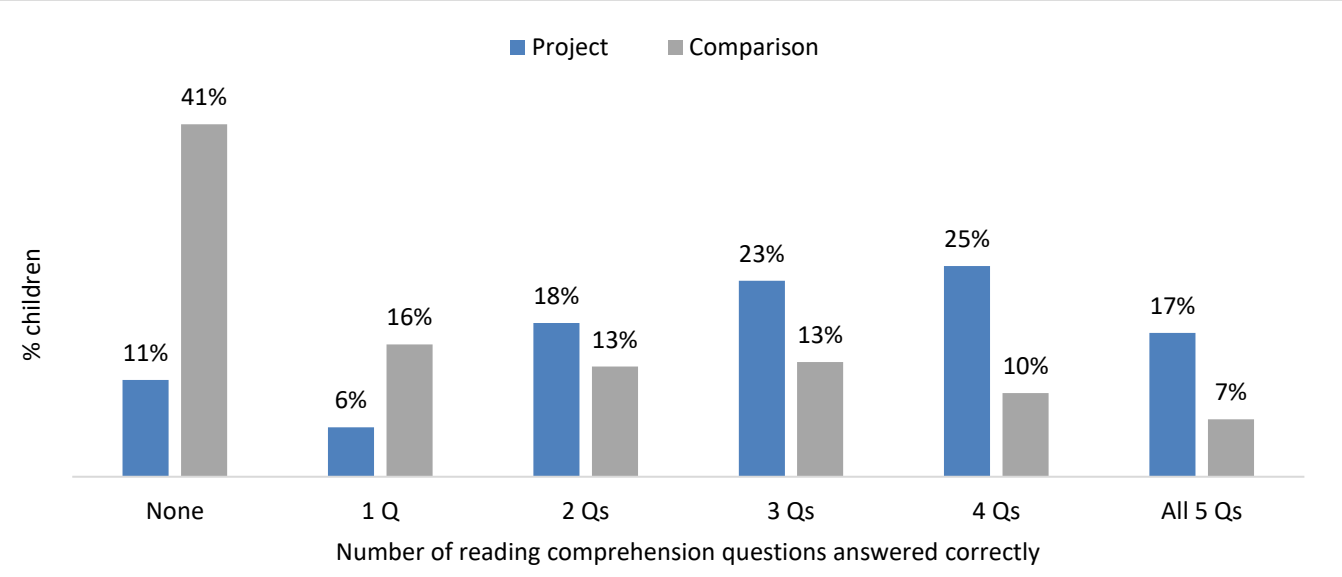


FIGURE 2.3: Distribution of Children Responding Correctly to Comprehension Questions at Endline 2019



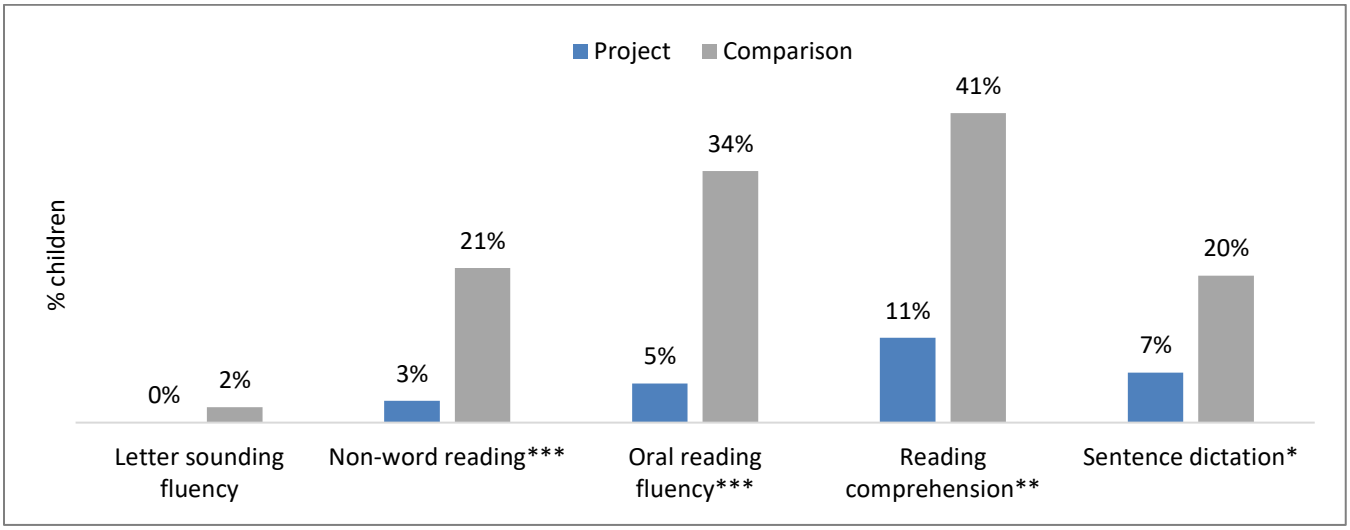
2.3 Zero Scores

Analysis of zero scores provides another view of how children performed, with a focus on children with the lowest achievement. In the EGRA, zero scores include those instances in which a child does not provide correct responses to any of the items attempted in a particular assessment task, as well as those instances when a child does not answer or respond correctly to any of the first pre-decided number of items for a particular assessment task (also known as a discontinued task)[6]. Zero scores on tasks show the subset of children who can be characterized as nonreaders.

[6] For example, on letter sounding fluency test, if a child could not give correct answer on any of the first 10 letters, the test was discontinued and the child was given 0 score on the test. Similarly, on non-word reading test, the threshold for discontinuity was set at first 5 non-words. For oral reading fluency test, we have used three different version of Nepali passages, each having 60 words, with similar level of difficulty in reading. The threshold for discontinuation for each these passages was slightly different from each other, ranging between first 6 to 14 words across passages.

Figure 2.4 below compares zero scores between project schools and comparison schools at endline. The data shows that prevalence of zero scores in both project and comparison schools reduced substantially at endline from the levels at baseline. **On letter sounding fluency none of the Grade 2 children scored zero in project schools at endline; and on non-word reading fluency ($p < 0.001$), oral reading fluency ($p < 0.001$), reading comprehension ($p < 0.01$), and sentence dictation ($p < 0.05$) subtasks prevalence of zero scores among Grade 2 children were significantly lower in project group than comparison group.** However, in project schools, highest proportion of zero scores at endline were noted for reading comprehension (11 percent), followed by sentence dictation (seven percent), oral reading fluency (five percent), and non-word reading fluency (three percent) subtasks respectively. Overall, these data suggest that project school children made substantial improvements after two year of intervention in contrast to comparison schools; however, few of them were struggling with higher-order reading skills like blending and reading for understanding.

FIGURE 2.4: Percentage of Children Scored Zero on Various Assessment Subtasks at Endline 2019



Legend of statistical significance of differences between project and comparison schools: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

2.4 Comparisons by Gender

Room to Read is also interested in understanding the role gender of a child plays alongside the program intervention. Table 2.2 below provides the assessment results disaggregated by gender for project and comparison school children.

The Literacy Program in Nepal benefited both boys and girls. Boys in Literacy Program schools in the three districts included in this study experienced two-year gains in reading and writing skills that were two to four times larger than those experienced by comparison school boys across all five assessment subtasks. On other hand, girls in Literacy Program schools experienced two-year gains in skills that were two to three times higher than the gains experienced by girls in comparison schools. Also, as shown in Figures 2.5 and 2.6, the effect sizes estimated separately for boys and girls were large (i.e. greater than 0.8) across all subtasks; however, these effect sizes were slightly larger for boys than girls. **Lastly, two-year gains in reading and writing skills made by the girls in project schools significantly higher than the gains made by the boys in project schools across all five assessment subtasks.**

TABLE 2.2: Average Assessment Scores across Subtasks in Baseline and Endline and Impact of the Program by Gender

Assessment Subtask (*unit of measurement*)	Group	Baseline 2017 (Start of Grade 1)			Endline 2019 (End of Grade 2)			2-Year Gains in Skills	Impact (Diff. in 2-year Gains in Skills b/w Project & Comparison Groups)[†]
		n	Mean	SD	n	Mean	SD		
Letter sounding fluency (correct letters per minute)	Project boys	146	7.5	7.8	156	49.7	18.3	+42.2	D1: +6.4*** D2: +9.2*** D3: +7.1***
	Comparison boys	78	9.5	12.0	98	24.6	15.4	+15.1	
	Project girls	197	8.8	8.9	187	57.6	18.3	+48.8	
	Comparison girls	115	14.9	15.8	95	34.4	18.0	+19.4	
Non-word reading fluency (correct non-words per minute)	Project boys	146	1.0	2.5	156	17.5	9.4	+16.5	D1: +1.1*** D2: +1.9*** D3: +3.7***
	Comparison boys	78	1.9	3.8	98	6.7	7.5	+4.8	
	Project girls	197	1.5	3.3	187	21.6	9.6	+20.2	
	Comparison girls	115	3.9	5.9	95	12.2	9.8	+8.3	
Oral reading fluency (correct words per minute)	Project boys	146	0.3	1.7	156	24.7	17.7	+24.4	D1: +7.8*** D2: +9.6*** D3: +8.1***
	Comparison boys	78	1.5	4.7	98	7.9	12.9	+6.4	
	Project girls	197	0.3	2.2	187	32.8	18.3	+32.4	
	Comparison girls	115	3.2	7.0	95	16.2	17.3	+12.9	
Reading comprehension (Qs answered correctly out of five)	Project boys	146	0.1	0.5	156	2.7	1.7	+2.7	D1: +1.6*** D2: +1.7*** D3: +0.5**
	Comparison boys	78	0.3	1.0	98	1.2	1.5	+0.9	
	Project girls	197	0.0	0.1	187	3.1	1.4	+3.1	
	Comparison girls	115	0.4	1.0	95	1.9	1.8	+1.5	
Sentence Dictation (score on a 39 point scale)	Project boys	146	0.7	2.4	156	20.2	11.8	+19.5	D1: +9.9*** D2:+10.2*** D3: +4.0***
	Comparison boys	78	2.2	5.6	98	11.4	12.1	+9.2	
	Project girls	197	0.9	3.0	187	24.2	10.8	+23.3	
	Comparison girls	115	4.8	8.8	95	17.8	13.2	+13.0	

D1: Differences in gains from baseline to endline between boys in project schools and boys in comparison schools.
D2: Differences in gains from baseline to endline between girls in project schools and girls in comparison schools.
D3: Differences in gains from baseline to endline between girls in project schools and boys in project schools.
Legend of statistical significance of differences between project and comparison schools: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.
[†]*Impact (Diff. in 2-year Gains in Skills b/w Project & Comparison Groups)*: Reports the estimated difference for each subtask between the (i) gains in reading or writing skills of boys and girls in project schools from baseline to endline after two years of intervention and (ii) gains in reading or writing skills of boys and girls in comparison schools during the same timeframe without the intervention. Each value presented in this column is the coefficient of the regression analysis conducted with random effects at the school level and age, enrollment in target grades (Grade 1 at baseline and Grade 2 at endline), number of teachers in primary grade, pupil teacher ratio in primary grade, number of PTMs conducted in current school year, whether school has a functional library, and district (with Nuwakot as base) as covariates. Additionally, version of passage used during the tests were included in the regression analysis for reporting results for oral reading fluency and reading comprehension, respectively.

FIGURE 2.5: Effect Sizes across Assessment Subtasks – Boys

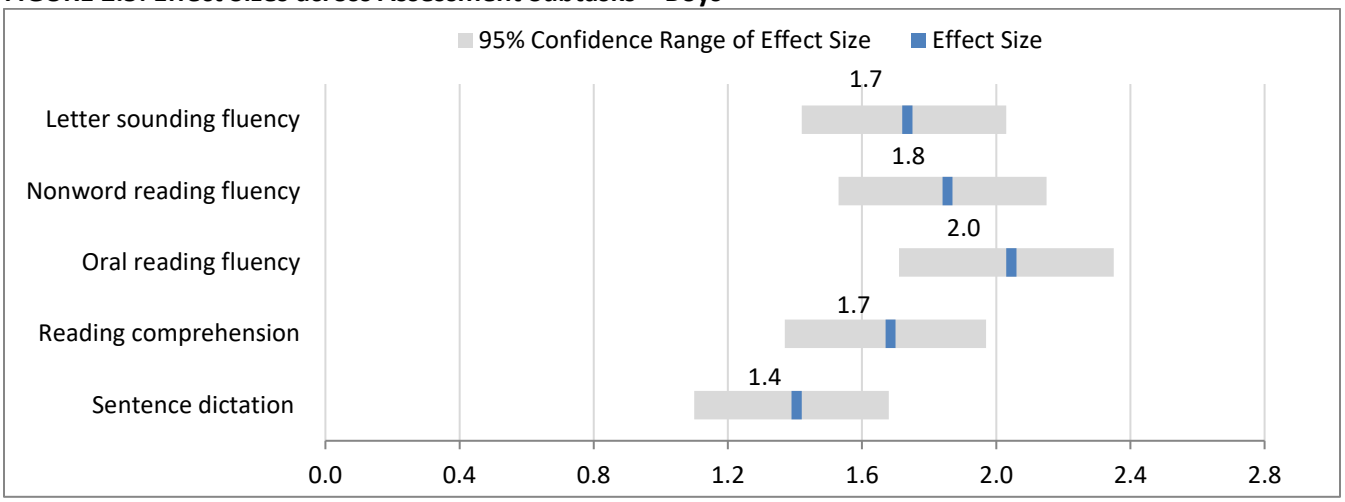
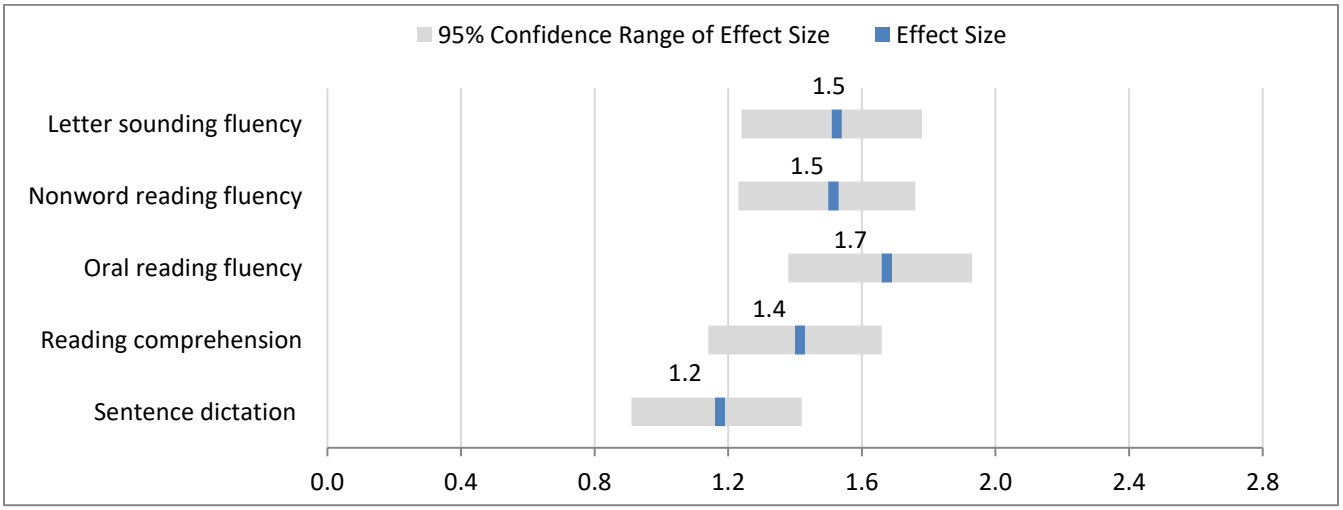


FIGURE 2.6: Effect Sizes across Assessment Subtasks – Girls



3. Limitations

Interpretation of the above results should include the limitations of the impact evaluation design. One such limitation of this evaluation was comparability. The validity of impact evaluation results rests on the strength of the assumption that the comparison group, on average, was similar to the project group across all observable and unobservable characteristics that may affect the outcomes (reading and writing skills) being evaluated. As explained in *Section 1.4: Sampling*, every effort was made during sampling to ensure that a comparable set of schools was chosen, and school and child-level characteristics were later analyzed to check whether any significant differences did, in fact, exist. As also discussed in this section, project and comparison group schools differed significantly in enrollments (in Grade 2 and all primary grades), total number of teachers in primary grades, pupil teacher ratio in primary grades, number of parent teacher meetings conducted in current school year and availability of a functional library. Statistical comparisons of reading and writing assessment results between project and comparison groups done through the regression analyses controlled for these differences across various school level characteristics. On the other hand, analyses of data on child-level background characteristics collected during the endline evaluation survey showed that the differences between the children from project and comparison groups were not statistically significant. However, because project group schools were not randomly selected for the Literacy Program in the three districts, it may be possible that they differed from comparison schools in ways that the evaluation did not assess or could not observe.

Another limitation of this evaluation was that the child sample size for assessments was lower than what was planned originally. It was planned that EGRA data would be collected from 15 children in each school. Information about the date of visit was shared with each school well before the data collection to ensure that the desired sample size could be achieved. However, children sample size achieved during the endline data collection was lower than the planned sample size due to various reasons, such as,

- *Tendency to inflate the enrolment numbers by schools*: While selecting schools for the survey, enrolment figures collected from the school registers were used. In areas with low population and/ or low enrollment, some schools showed inflated enrolment number to get grant from the government[7] and also to avoid merging of schools. Also, it was found that schools do not update the enrolment figures despite few children being absent for long[8]. This had caused lower than expected student sample in some schools.
- *Accessibility*: Some schools (especially in Nuwakot district) that were located in remote areas, attendance was in general low on the days of the assessment, despite repeated visits.
- *Social or community events/ school holidays*: In few schools, it was found that a large number of children or even the whole class was absent at the time of visit to attend some social or community-level event (e.g. marriage or religious festival) in the locality. In such situation, enumerators with the help of school authority tried to call back as many Grade 2 children as possible to the schools to participate in the assessment. Low attendance were also seen in some schools which according to the school authorities was due to the holiday at the middle of the week. Additionally, data collection was also affected by some unexpected events like teacher's strike and nationwide one day school closure due to unfortunate death of a government minister.

[7] Government of Nepal provides grants to schools on the basis of enrollment.
[8] It was informally communicated to the evaluation team that long absence by children from schools was mostly due to seasonal or (in some cases cases) long-term migration by the families.

As the plan was to assess children in the selected schools, this evaluation could not triangulate student self-reported responses on child-background characteristics with those of an informed adult, like a parent or a teacher (e.g. parents' education, availability of reading materials at home). Anticipating this challenge, the assessor were provided with training (desk- and field-based) in collecting reliable self-reported data from children in early primary grades to the extent possible.

Lastly, in this evaluation EGRA data collected from a sample of children who were present in the schools on the days of assessment rather than drawing a sample from full classroom list. The possibility of systematic student absences[9] might have induced a risk of sampling bias by doing assessment only with the subset of children present in the schools. However, the overall sampling approach adopted here tried to ensure a large enough sample (for both school and children) for the evaluation and sampling consistency across schools. Also, the school sample sizes for both project and comparison groups, unlike the children sample, was not affected.

[9] In the context of Nepal, systematics student absences can be caused by factors like seasonal/ cyclical migration of the families, health of children specifically those from vulnerable socio-economic backgrounds, remoteness of the schools, etc.

4. Conclusion

Results from the 2019 end-of-Grade 2 endline evaluation show that the Literacy Program in Nepal is having large and positive impact on reading skills of children. **Children from project schools performed significantly better than children from comparison schools across all assessment subtasks at endline. Moreover, children in Literacy Program schools in Nepal experienced two-year gains in reading and writing skills that were two to three times larger than those experienced by comparison school children across all five assessment subtasks.** Significantly, higher number of children in program schools, in contrast to comparison schools, reached the fluent reader benchmark of reading 45 or more words correctly per minute and the level of 80 percent or more comprehension. Also, the Program was effective for both boy and girls.

Despite the favorable results for the program after two years of intervention, room for improvement exists. For example, five percent of the Grade 2 children on the program could not read a single word of a reading passage, 11 percent could not correctly answer a single reading comprehension question, and seven percent could not write a single word correctly. Moreover, shares of children reaching the fluent reader benchmark and 80 percent comprehension in program schools were lower than expected. These results suggest that more can be done to improve fluency and comprehension of some of the children who were still struggling with reading and writing after two years of intervention. And Room to Read is committed to do so by continuing its support to the schools in Nepal in collaboration with the government to improve reading and writing skills of early grade children with specific focus on the following

- Continue to refine the framework for and use of program monitoring data related to inputs, activities and outputs (e.g. monthly classroom observation of teacher's adoption of instruction techniques, bi-annual classroom-based student tracking) to track quality of implementation, provide focused support to the teachers and schools and to better understand how this correlates with reading skills outcomes.
- Focus more on supporting teachers and schools to encourage children for library use and book check-out to improve the habit of reading and reading comprehension.
- Continue to focus on parental engagements in schools and communities to improve children's attendance in school and encourage parents to provide more support to children in reading at home.
- Keep improving vital elements of the instruction component of the Literacy Program (e.g. blending to read unfamiliar words) to further enhance children's reading skills in early primary grades.
- Continue to advocate Room to Read's approach to literacy development with government stakeholders and other organizations within the education sector.

References

Abadzi, H. (2011). Reading Fluency Measurements in EFA FTI Partner Countries: Outcomes and Improvement Prospects. Working Paper.

Harris, T.L. & Hodges, R.E., (Eds.). (1995): The literacy dictionary: The vocabulary of reading and writing. Newark, DE: International Reading Association.

Kuhn, M., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and the definitions of fluency. *Reading Research Quarterly*, 45, 230-251.

RTI International. (2015). Early Grade Reading Assessment (EGRA) Toolkit, Second Edition. Washington, DC: United States Agency for International Development.

Snow, C.E. (2002). Reading for understanding: Toward and R&D program for reading comprehension. RAND Corp. Available at: http://www.rand.org/pubs/monograph_reports/MR1465.html

StataCorp. (2013). Stata Statistical Software: Release12. College Station, TX:: StataCorp LP.

The World Bank. (2011). Impact Evaluation in Practice. Available at: https://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726129545562862O/Impact_Evaluation_in_Practice.pdf

Appendix A: Data Analysis[10]

The first aim of the data analysis was to determine if there were significant differences in school and child background characteristics between the project and comparison groups. For the school background variables, (i) differences between the groups in enrollments (in Grade 2 and in all primary grades), attendance rate in Grade 2 on the day of endline assessment, number of primary grade teachers, pupil-teacher ratio in primary grade, and number of parent teacher meetings conducted in current school year were examined by conducting *t* tests, and (ii) differences between groups on whether schools have regular access to drinking water, separate toilets for boys and girls, functional libraries, and are operating any student feeding program (as observed during the visits) were examined by conducting *chi-square* tests. For the child background variables, differences between the groups in age, gender, whether child speaks Nepali at home, whether child's mother and father (separately) can read and write, whether child attended Early Child Development classes, whether child has a collection of books at home for reading, and whether child's family has radio, television, motorcycle and mobile phone at home were examined by conducting regression analyses (linear regression for the continuous variables and logistic regression for the categorical variables) with random effects at the school level. The equations included the child background variables as dependent variables and school group (project or comparison) as independent variable. The results of these analyses explained in *Section 1.4: Sampling*.

The second aim of the data analysis was to determine if children in the project group made greater gains from baseline to endline than children in the comparison group. The analysis strategy was to compare reading skills in the two assessment periods (baseline versus endline) among the two experimental groups (project versus comparison). An impact of the Literacy Program in Nepal is evident if there is a greater gain from baseline to endline among the project group compared to the comparison group on any of the reading and writing subtasks. This is demonstrated by a statistically significant interaction between experimental groups and assessment period. To determine this linear regression analyses were conducted for the assessment subtasks with random effects at the school level and dummy variables for the assessment period, experimental group, and the interaction between the two. Each analysis included one of the assessment subtask scores as the dependent variable and age, gender, enrollment in target grades (Grade 1 at baseline and Grade 2 at endline), number of teachers in primary grade, pupil teacher ratio in primary grade, number of PTMs conducted in current school year, whether school has a functional library, and district (with Nuwakot as base) as covariates. A similar procedure to analyze differences in gains by gender was used. A dummy variable for gender was created and linear regression analysis for each subtask was conducted with random effects at the school level to examine differences in gains across each of the following: (i) *D1*: project school boys versus comparison school boys, (ii) *D2*: project school girls versus comparison school girls, and (iii) *D3*: project school girls versus project school boys. Each analysis included one of the assessment SUBTASK scores as the dependent variable and age, enrollment in target grades (Grade 1 at baseline and Grade 2 at endline), number of teachers in primary grade, pupil teacher ratio in primary grade, number of PTMs conducted in current school year, whether school has a functional library, and district (with Nuwakot as base) as covariates.

Further, zero scores on the subtasks were analyzed to determine the impact of the intervention on prevalence of non-readers. The analysis of zero scores is particularly appropriate when the distribution of scores is skewed towards zero (i.e. is not in a bell-shaped curve). Logistic regression analysis with random effects at the school level was conducted for each subtask to determine if significant differences existed between the percentage of project

[10] All data analyses were conducted using Stata statistical software (Stata Corp, 2013).

school vs. comparison school children registering zero scores (vs. non-zero scores) at endline evaluation. The equations included the presence or absence of a zero score as the dependent variable, type of school (project versus comparison) as the predictor, and age, gender, enrollment in target grades (Grade 1 at baseline and Grade 2 at endline), number of teachers in primary grade, pupil teacher ratio in primary grade, number of PTMs conducted in current school year, whether school has a functional library, and district (with Nuwakot as base) as covariates. Similar logistic regression analyses with random effects at the school level was conducted to assess whether significant differences existed between the percentage of project school vs. comparison school children reaching (i) the fluent reader benchmark (i.e. reading 45 or more words correctly per minute), and (ii) achieving 80% or more on reading comprehension at endline evaluation.

Effect size for each of the five reading assessment subtasks was also computed to demonstrate the magnitude of effect of Literacy Program in Nepal. Effect size was computed as the mean difference in gains between the project and comparison schools divided by the pooled standard deviation of scores in project and comparison schools. An effect size is exactly equivalent to a "Z-score" of a standard normal distribution. For example, an effect size of "+1" means that, on average, the score of a child in project school is "+1" standard deviation above the score of a child in the comparison school, and hence exceeds the scores of 84 percent of the children in comparison group.

Appendix B: School and Child Background Characteristics

TABLE B.1: Background Characteristics of Sample Schools at Endline

	Project		Comparison	
	n	Mean (SD) or %	n	Mean (SD) or %
School	30	-	30	-
Districts				
Nuwakot	7	23%	7	23%
Palpa	6	20%	6	20%
Tanahun	17	57%	17	57%
Location				
Rural	27	90%	28	93%
Semi-urban	3	10%	2	7%
Enrollment				
Primary**	30	85 (34)	30	56 (37)
Grade 2*	30	15 (5)	30	10 (8)
Percentage of Grade 2 children present on day of endline assessment	30	82 (16)	30	75 (23)
Total teachers in primary grades*	30	9 (2)	30	8 (2)
Pupil-teacher ratio (primary grades only)*	30	10 (4)	30	7 (4)
Number of PTMs conducted for primary grade in current school year*	30	4 (2)	30	3 (2)
Has regular access to drinking water	25	83%	27	90%
Has separate toilet for boys and girls	28	93%	26	87%
Has a function library***	29	97%	8	36%
Observed any school feeding program during visit	7	23%	9	30%

Legend of statistical significance of the difference between project and comparison school: ** $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$*

TABLE B.2: Background Characteristics of Sample Children*

	Project		Comparison	
	n	Mean (SD) or %	n	Mean (SD) or %
Children				
Baseline (2017)	343	-	193	-
Midline (2017)	364	-	226	-
Endline (2018)	343	-	194	-
Gender - Baseline (2017)				
Boys	146	43%	78	**40%**
Girls	197	57%	115	**60%**
Gender - Midline (2018)				
Boys	170	47%	98	**43%**
Girls	194	53%	128	**57%**
Gender - Endline (2018)				
Boys	156	45%	99	**51%**
Girls	187	55%	95	**49%**
Age (Endline only)	343	7.6 (1.2)	194	**7.7 (1.3)**
Speaks Nepali at home (Endline only)	204	59%	142	**73%**
Attend Early Child Development classes (Endline only)	323	94%	167	**86%**
Father can read and write (Endline only)	306	89%	171	**88%**
Mother can read and write (Endline only)	268	78%	153	**79%**
Has radio at home (Endline only)	114	33%	59	**30%**
Has television at home (Endline only)	173	50%	95	**49%**
Has motorcycle at home (Endline only)	41	12%	26	**13%**

	Project		Comparison	
	n	Mean (SD) or %	n	Mean (SD) or %
Has mobile phone at home (Endline only)	339	99%	190	**98%**
Has collection of books at home (Endline only)	20	6%	8	**4%**

Differences between project and comparison schools on any of the children background characteristics at endline were not statistically significant.

TABLE B.3: Correlations between Scores on Different Reading Assessments at Endline

	Letter sounding fluency	Non-word reading	Oral reading fluency	Reading comprehension	Sentence dictation
Project					
Letter sounding fluency	1.000				
Non-word reading	0.7868***	1.000			
Oral reading fluency	0.7507***	0.7629***	1.000		
Reading comprehension	0.5345***	0.5100***	0.6217***	1.000	
Sentence dictation	0.6254***	0.6200***	0.7224***	0.6446***	1.000
Comparison					
Letter sounding fluency	1.000				
Non-word reading	0.0055	1.000			
Oral reading fluency	0.8242***	0.0572	1.000		
Reading comprehension	0.7380***	0.0276	0.8151***	1.000	
Sentence dictation	0.8127***	0.0242	0.8291***	0.7924***	1.000
Overall					
Letter sounding fluency	1.000				
Non-word reading	0.1295**	1.000			
Oral reading fluency	0.8184***	0.1550***	1.000		
Reading comprehension	0.6821***	0.1028*	0.7360***	1.000	
Sentence dictation	0.7175***	0.1116**	0.7776***	0.7364***	1.000

Legend of statistical significance: ** $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$*