Literacy Program: South Africa (Sepedi)

2016 Midline Evaluation Report

Annika Rigole

Peter Cooper

5/5/2017

In February 2016, Room to Read began a two-year internal evaluation of its newly-contextualized Sepedi language Literacy Program in South Africa. The aim of the evaluation was to determine the impact of the intervention on children's reading skills. Data collected one year into the evaluation in October and November 2016 revealed that the program is having a positive impact on the development of Grade 1 learners' reading skills. Learners at schools benefiting from the program made reading skill gains from the beginning to end of Grade 1 that were nearly two times larger than learners at comparison schools not receiving the intervention. The evaluation also highlights that there are areas for improvement in 2017.

Table of Contents

1 Executive Summary

Room to Read's Literacy Program is a multi-faceted intervention that seeks to develop children's reading skills and reading habit in the primary grades. The program strengthens reading and writing instruction in Grades 1 and 2 and provides regular access to reading materials through the establishment of school libraries and the provision of quality local language books. The instruction component of the program works in conjunction with a country's existing language curriculum and includes detailed lesson plans, classroom materials, and comprehensive teacher training and coaching support. Overall, the program has had a substantial impact on reading skills and reading habits in nearly all of the countries in which it operates. In South Africa, a newly-contextualized Literacy Program was launched in 70 schools in Limpopo Province in January 2016.

In February 2016, a two-year impact evaluation was launched to accompany the roll-out of the new program and measure its impact on learners' reading skills development. The evaluation included 26 schools benefiting from the program and 24 comparable schools in the same region not benefiting from the program. Room to Read assessed learners' reading skills using a version of the the Early Grade Reading Assessment (EGRA)[1] that was adapted to the Sepedi language by local experts. The baseline assessment found that children from Literacy Program schools and comparison schools entered Grade 1 with similarly low reading skills.

In October-November 2016, we conducted a midline follow-up assessment with the same cohort of pupils to measure the impact of the program after one academic year**.** The results showed that the Literacy Program had a positive impact on improving Grade 1 learners' reading skills. Learners in project schools made one-year gains in reading skills that were approximately two times larger than those of comparison school learners. Thirty-eight percent of project school learners met Room to Read's Grade 1 letter sounding benchmark (45 letters per minute), 30 percent reached our "emergent reader" benchmark (20 correct words per minute), and 6 percent reached our "fluent reader" benchmark (45 correct words per minute). The larger gains seen in the letter sounding task suggest that the Grade 1 instructional intervention was most effective at building letter-sound recognition skills.

As expected in the first year of a new program, the evaluation results also highlighted that room for improvement exists. A large percentage of project school learners remained unable to read or achieve the letter sounding benchmark by the end of Grade 1. Although the newly contextualized program is having a bigger impact (relative to comparison school children) than the earlier version of the program, the average reading levels achieved by learners benefiting from the new program were lower. The staffing challenges that limited literacy coaches' support to teachers and the learning curve associated with the rollout of the new program may have reduced the potential impact of the program in its first year.

In the coming years, Room to Read aims to improve Literacy Program implementation and thereby increase the number of learners reading fluently with comprehension. To build upon the positive results from this Grade 1 program evaluation, the Room to Read South Africa team proposes the following actions and strategies for 2017:

1. **Better utilize formative assessment data** to target struggling students and ensure that resulting improvement strategies are implemented;
2. **Investigate why letter sounding performance is such a success** and the other tasks are still too difficult for the learners;

[1] The EGRA was developed by RTI International in 2006. For more information, please see: https://globalreadingnetwork.net/resources/early-grade-reading-assessment-egra-toolkit-second-edition.

3. **Improve operational practices to ensure there are sufficient field staff** to support teachers and that the program is implemented with fidelity to its design;
4. **Utilize the school-by-school EGRA assessment results to target teachers** at low performing schools with additional support, and develop a strategy for sharing results with district and provincial education officials;
5. **Analyze data from classroom observation forms and program implementation monitoring** to better understand the quality of implementation and how this correlates to reading skills outcomes;
6. **Incorporate a discussion of assessment into teacher trainings** to help teachers utilize the assessment components to effectively revise and check learning.

Through these and other routine strategies to review and improve the Literacy Program, Room to Read hopes to achieve its goal of having all children in the program reading fluently with comprehension by the end of Grade 2.

2 Introduction

Room to Read's Literacy Program is a school-based intervention that seeks to develop children's reading skills and reading habit in the primary grades. The program includes two main components: reading and writing instruction for children in Grades 1 and 2, and access to reading materials through the establishment of school libraries and provision of quality local language reading materials. The instruction component of the program works in conjunction with the country's existing language curriculum and includes the provision of detailed lesson plans, classroom materials, and comprehensive teacher training and coaching support (see *Appendix A* for more detail); its primary goal is for children to become fluent readers by the end of Grade 2.

Room to Read Literacy Program teams in all countries participate in an impact evaluation that includes reading assessments at the beginning of Grade 1 (baseline) and at the end of Grade 1 (midline) and Grade 2 (endline). The data from these evaluations enable Room to Read to:

1. Determine whether the Literacy Program is having an impact on learners' reading skills;
2. 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
3. Identify reading skills that could be better supported by the program and determine how to improve these reading skills quickly and effectively.

To date, the Literacy Program has had a substantial impact on reading skills in nearly all of the countries in which it operates. For example, the program has improved reading fluency scores by a magnitude of 1.5 to 3.5 across eight of the nine countries in which it has been implemented.

In South Africa, a newly-contextualized Literacy Program was launched in Sekhukhune District in Limpopo Province in January 2016. The program targeted Grade 1 teachers in 70 project schools and expanded to supporting Grade 2 teachers in these same schools in 2017. A two-year impact evaluation was launched to accompany the roll-out of the new program and measure its impact on children's reading skills development. The evaluation included 26 schools benefiting from the program, randomly selected from the 70 new project schools; it also included 24 non-project "comparison" schools, which were randomly selected from a set of comparable schools in nearby wards in Sekhukhune District.[2] Baseline reading skills assessments were conducted in February 2016 with 30 Grade 1 learners per school to measure the initial reading skills of learners who would and would not be benefiting from the Literacy Program. The baseline results indicated that learners from the project schools entered Grade 1 with the same level of reading skills as learners from the comparison schools. In late-October and early-November 2016, a midline follow-up assessment was conducted with the same cohort of learners to measure the program impact after one school year of project support (see *Section 3: Results* for details on the results).

Room to Read assessed learners' reading skills using a version of the Early Grade Reading Assessment (EGRA)[3] that was adapted to the Sepedi language by local experts. The EGRA featured the following four tasks: letter sounding, non-word reading, oral passage reading, and reading comprehension. The assessments were administered

[2] Initially 25 project and 25 comparison schools had been identified for the evaluation sample, but one comparison school was changed to a project school by program staff after baseline data collection completed.

[3] The EGRA was developed by RTI International in 2006. For more information, please see: https://globalreadingnetwork.net/resources/early-grade-reading-assessment-egra-toolkit-second-edition.

individually with learners by external data collectors trained by Room to Read. Please see *Appendix B* for a full description of the research design and methodology.

3 Results

3.1 School and Learner Background Characteristics

Since the literacy intervention was not allocated randomly to project and comparison schools, it is important to assess whether the two groups were comparable. At the school level, on average, project schools had significantly larger enrollment, Grade 1 class sizes, and pupil-teacher ratios than comparison schools. At the learner level, the project school learners were significantly more likely to have a television at home and to have a literate mother than comparison school learners at baseline. In the midline evaluation sample, meanwhile, those differences did not exist, but project school learners were significantly less likely to have a collection of books at home. Statistical comparisons of assessment results took these differences into consideration. See Tables C.1 and C.2 in *Appendix C* for more details on the school and learner background characteristics.

3.2 End-of-Grade-1 Reading Results

3.2.1 Changes in Average Reading Scores

Across each of the four reading skills assessed, project school learners made gains from the beginning to end of Grade 1 that were approximately two times larger than those of comparison school learners. On average, learners benefiting from the Sepedi Literacy Program could sound 37 letters correctly per minute and read nearly 14 words correctly per minute at the end of Grade 1. Comparison school learners, meanwhile, sounded 21 letters correctly per minute and read seven words correctly per minute at the end of Grade 1. Figure 1 depicts these key fluency results, and Table 1, below, shows the baseline and midline mean scores by project and comparison school group for each assessment tasks.

FIGURE 1: Key Grade 1 Fluency Results

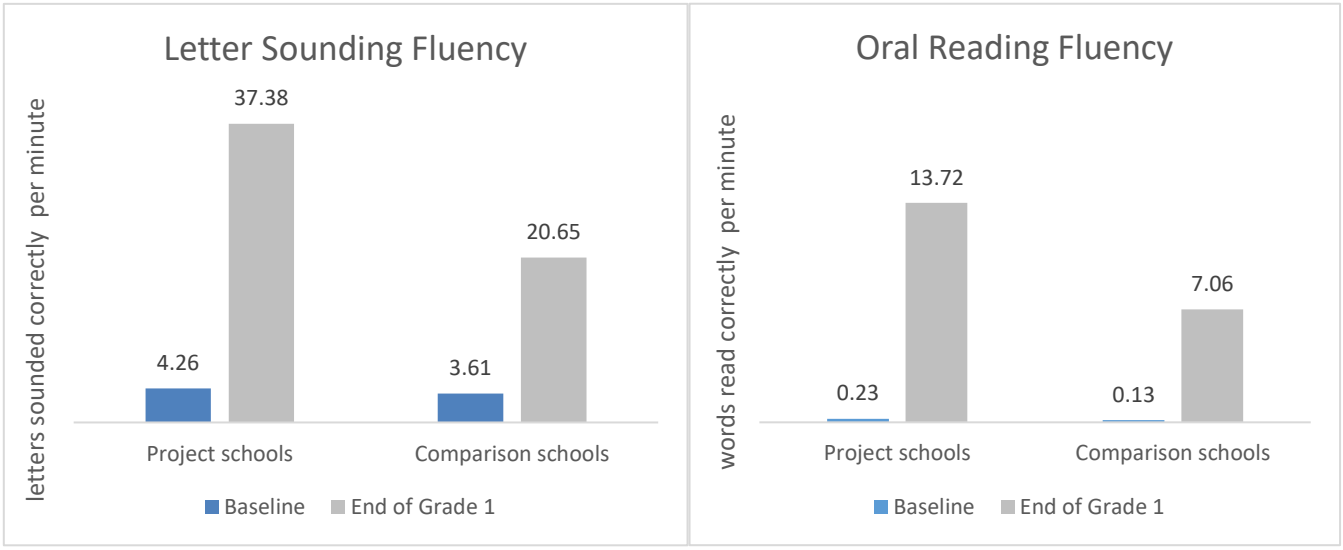


TABLE 3.1: End of Grade 1 Reading Assessment Results

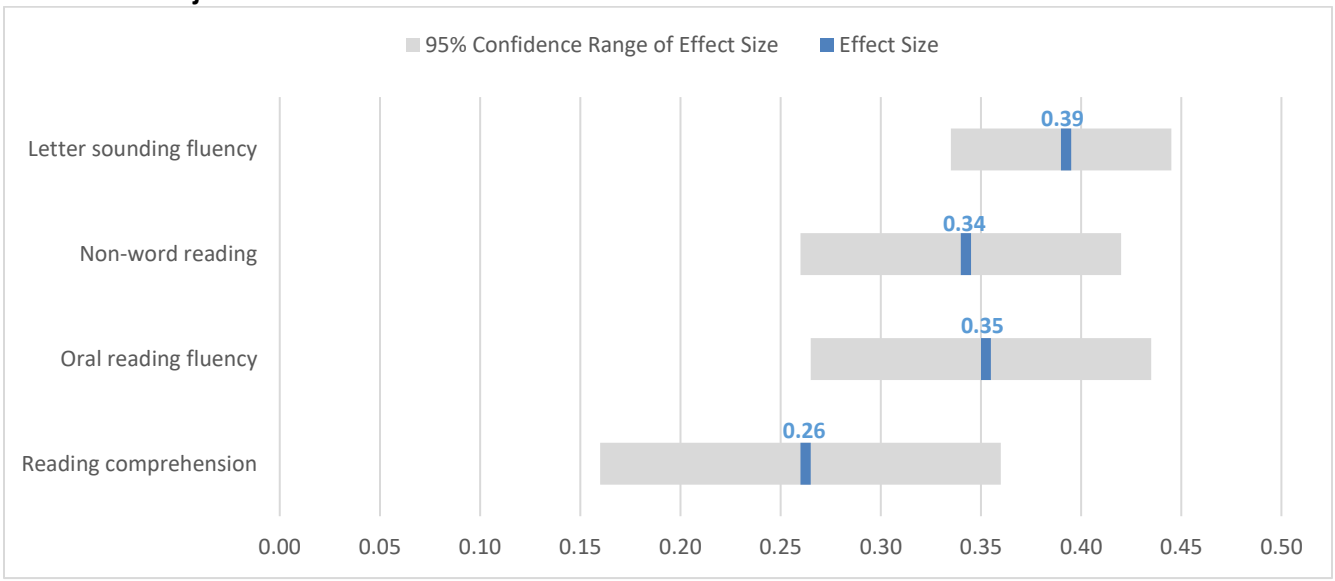
Assessment Task	Group	Baseline Mean	End of Grade 1	Gains over 1 year	Adjusted Difference in Gains†
Letter sounding fluency***	Project	**4.26**	**37.38**	+33.13	**+16.02**
(letters per minute)	Comparison	**3.61**	**20.65**	+17.04	
Non-word reading fluency *	Project	**0.21**	**8.87**	+8.66	**+4.41**
(correct words per minute)	Comparison	**0.14**	**4.34**	+4.20	
Oral reading fluency***	Project	**0.23**	**13.72**	+13.50	**+6.47**
(correct words per minute)	Comparison	**0.13**	**7.06**	+6.93	
Reading comprehension***	Project	**0.19**	**0.75**	+0.56	**+0.33**
(questions answered correctly)	Comparison	**0.16**	**0.39**	+0.23	

Legend of statistical significance of differences between project and comparison schools: *** < .001, ** < .01, * < .05

†Adjusted difference in gains reports the difference in gains after controlling for differences in school and learner background between project and comparison schools.

Differences in student gains are further examined by looking at the adjusted effect sizes for the Literacy Program across the assessment tasks (see Figure 3.2 below). The effect size statistic is used to make comparisons across measures that use different scales or units. For the purposes of this analysis, we used the standardized mean effect size statistic, through which an effect size of .80 or higher is considered large. We calculated effect sizes by determining the adjusted difference in scores between project school students and comparison school students through linear regression analysis (see *Appendix B: Research Design*) and then dividing this difference by the pooled standard deviation of project and comparison learners' scores. **The effect size for the Literacy Program intervention was small to moderate for the four reading tasks, with the program having its largest effect (0.39) on improving the foundational skill of letter sounding and its smallest effect (0.26) on improving reading comprehension.**

FIGURE 3.2: Adjusted Effect Sizes across Assessment Tasks

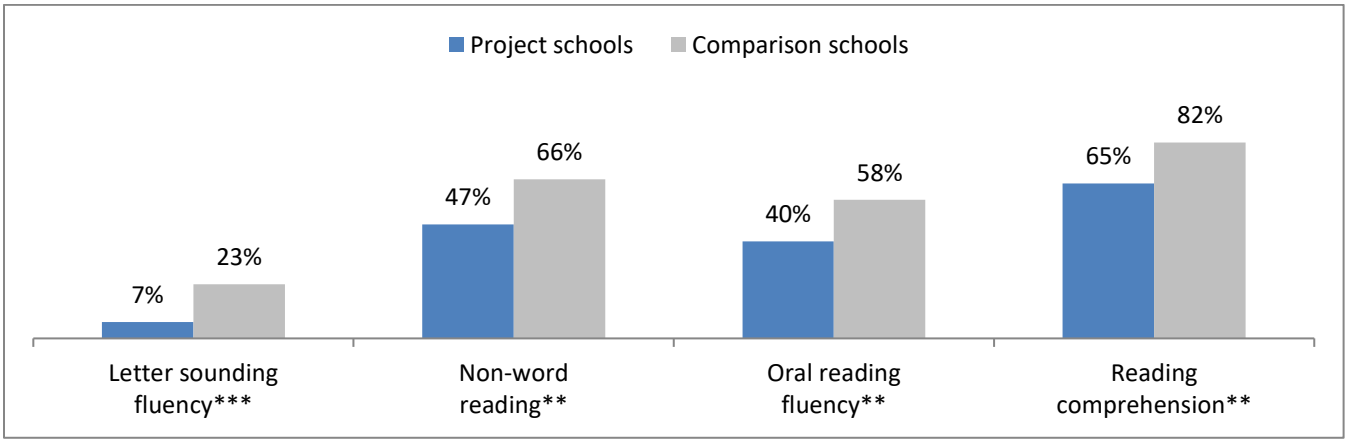


3.2.2 Zero Score Prevalence

Analysis of zero scores provides another view of how learners performed, with a particular focus on children with the lowest achievement. In the EGRA, zero scores include those instances in which a learner does not respond correctly to any item in the first line of a reading assessment task (also known as a discontinued task), or does not provide correct responses to a question. Zero scores on tasks show the subset of learners who can be characterized as nonreaders.

Figure 3.3 depicts the number of children scoring zero on each assessment task for project and comparison school learners. The percentage of project school learners scoring zero was significantly lower than the percentage of comparison learners for all four reading tasks ($p<.01$). In fact, **the Literacy Program reduced the prevalence of zero scores by 15 to 20 percentage points on each of these assessment tasks.** Nevertheless, 40 percent of project school learners remained unable to read a single word correctly at the end of Grade 1.

Figure 3.3: Zero Score Prevalence by Assessment Task



Legend of statistically significant differences between project and comparison groups: *** < .001, ** < .01, * < .05

3.2.3 Fluency Benchmarks

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 correct words read per minute as a prerequisite to read with comprehension.[4] Room to Read aims for learners to reach this fluency level by the end of Grade 2, but we also track a letter sounding benchmark (45 letters sounded per minute) and an emergent reader benchmark (20 words read per minute) to assess learner progression towards this goal at the end of Grade 1.

The evaluation revealed that **thirty-eight percent of project school learners were able to sound at least 45 letters correctly per minute, in contrast to 12 percent of comparison school learners.** Seven percent of project school learners, meanwhile, remained unable to sound a single letter correctly. Figure 3.4 depicts the distribution of learners' letter sounding scores in greater detail.

Additionally, **30 percent of project school learners were able to read at least 20 words correctly per minute, thus meeting the standard to be classified as an "emergent reader."** However, 40 percent of project school learners remained "non-readers," as they could not read a single word of a grade-appropriate reading passage by the end of the school year. Figure 3.5, below, depicts the distribution of learners' oral reading fluency scores.

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

Figure 3.4: Distribution of Learners by Letter Sounding Fluency Score

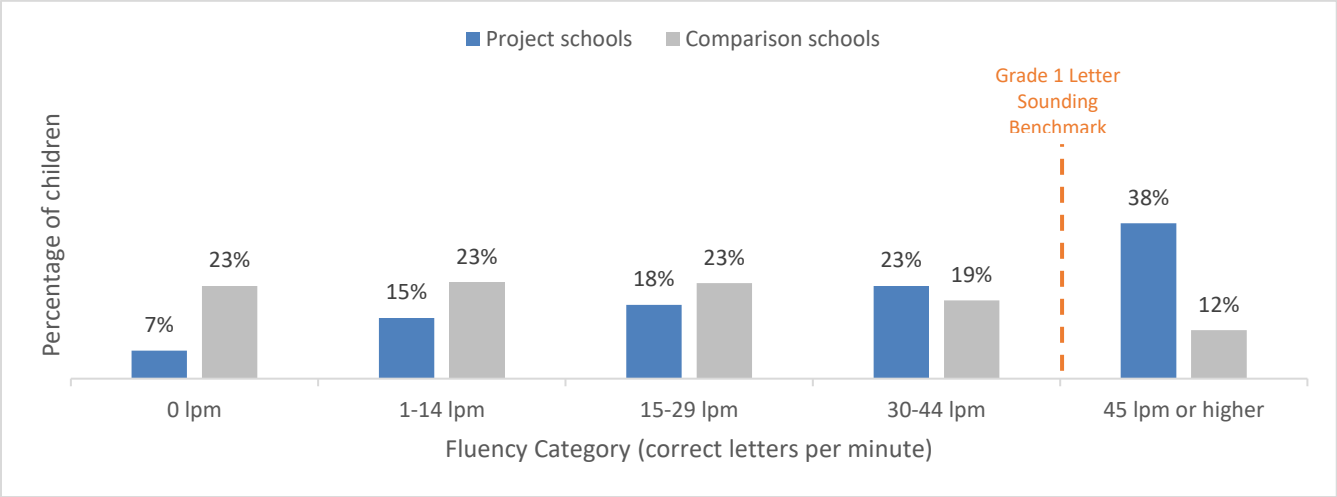
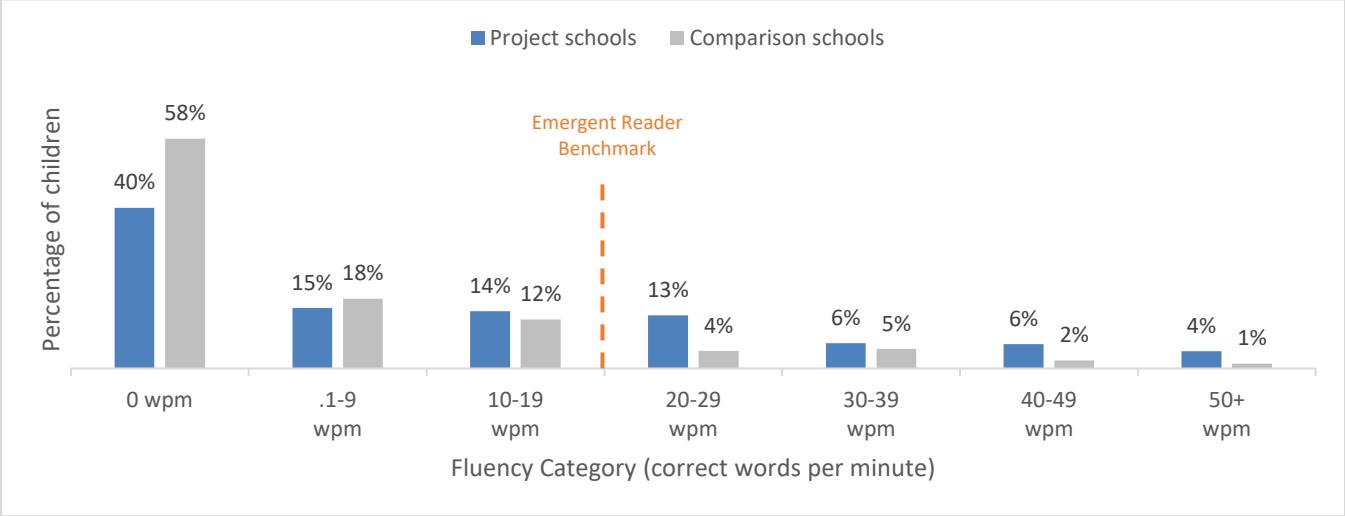


Figure 3.5: Distribution of Learners by Oral Reading Fluency Score



3.3 Gender Comparisons

Though the primary focus of the evaluation was comparing learners who benefit from the Literacy Program with those who do not, Room to Read is also interested in the interaction between the presence of the program and gender. We find that **the Literacy Program benefitted both boys and girls.** Boys in the project schools made significantly greater gains on all four assessment tasks than boys in comparison schools ($p < .01$). Similarly, girls in the project schools made significantly greater gains than comparison school girls on all four assessment tasks ($p < .001$).

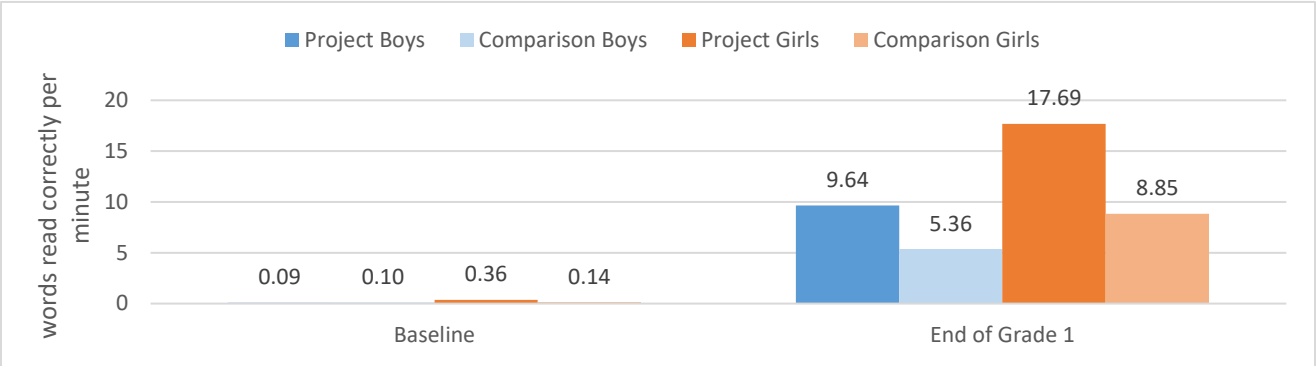
Additionally, girls scored significantly higher than boys on all assessment tasks ($p < .001$), and the Literacy Program significantly increased the gains girls made over their male counterparts on all assessment tasks ($p < .001$). For example, from baseline to midline, project school girls improved their reading fluency by 17 words per minute, while project school boys only improved by 10 words per minute. Comparison school girls also improved more than comparison school boys, but by a smaller margin ($p < .01$). Figure 3.6 depicts the gender breakdown for the oral reading fluency results, while Table 3.2, below, provides the midline assessment results disaggregated by gender for all four assessment tasks.

TABLE 3.2: Comparison of Average Scores by Gender

		BASELINE			END OF GRADE 1				
Assessment Task	**Group**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**Gains**	**Adjusted Difference in Gains†**
Letter sounding fluency [a, b, c, d]	Project boys	378	3.19	5.86	374	31.28	23.19	**+28.09**	**+13.91**
(letters per minute)	Comparison boys	307	3.03	6.44	304	17.18	16.91	**+14.15**	
	Project girls	376	5.33	8.10	384	43.33	24.68	**+38.00**	**+18.02**
	Comparison girls	292	4.21	7.01	290	24.29	19.90	**+20.08**	
Non-word reading fluency [a, b, c, d]	Project boys	378	0.09	0.98	374	6.22	9.31	**+6.13**	**+2.86**
(correct words per minute)	Comparison boys	307	0.10	0.81	304	3.34	7.20	**+3.23**	
	Project girls	376	0.33	1.81	384	11.44	12.70	**+11.12**	**+5.94**
	Comparison girls	292	0.18	1.27	290	5.39	9.02	**+5.21**	
Oral reading fluency [a, b, c, d]	Project boys	378	0.09	0.96	374	9.64	14.05	**+9.55**	**+4.26**
(correct words per minute)	Comparison boys	307	0.10	0.73	304	5.36	11.06	**+5.25**	
	Project girls	376	0.36	1.90	384	17.69	18.49	**+17.33**	**+8.67**
	Comparison girls	292	0.14	1.02	290	8.85	13.40	**+8.71**	
Reading comprehension [a, b, c, d]	Project boys	378	0.15	0.41	374	0.51	1.00	**+0.37**	**+0.27**
(questions answered correctly)	Comparison boys	307	0.17	0.45	304	0.27	0.84	**+0.10**	
	Project girls	376	0.23	0.55	384	0.98	1.40	**+0.75**	**+0.39**
	Comparison girls	292	0.14	0.38	290	0.51	1.07	**+0.37**	

[a] Differences in one-year gains between **boys in project schools and boys in comparison schools** were statistically significant ($p < .01$).
[b] Differences in one-year gains between **girls in project schools and girls in comparison schools** were statistically significant ($p < .001$).
[c] Differences in one-year gains between **boys in project schools and girls in project schools** were statistically significant ($p < .001$).
[d] Differences in two-year gains between **boys in comparison schools and girls in comparison schools** were statistically significant ($p < .01$).
†Adjusted difference in gains reports the difference in gains after controlling for differences in school and learner background between project and comparison schools.

Figure 3.6: Oral Reading Fluency by Gender Group



Figures 3.7 and 3.8 depict the program effect sizes for boys and girls, respectively. **The effect sizes were larger for girls on all four assessment tasks.** The program had a moderate effect for girls on the letter sounding, non-word reading, and oral reading tasks (0.47-0.50), and a smaller effect on improving reading comprehension (0.31). The program had a moderate effect on boy's letter sounding as well (0.49), but a smaller effect on improving boys' skills on the remaining three reading tasks (0.28-0.33).

FIGURE 3.7: Adjusted Effect Sizes – Boys in Project Schools

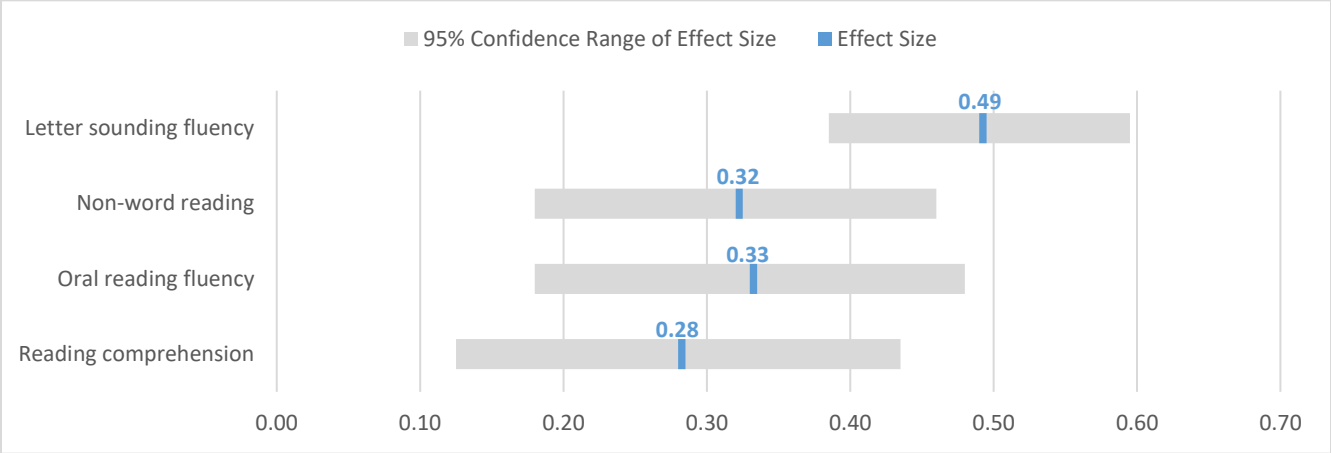
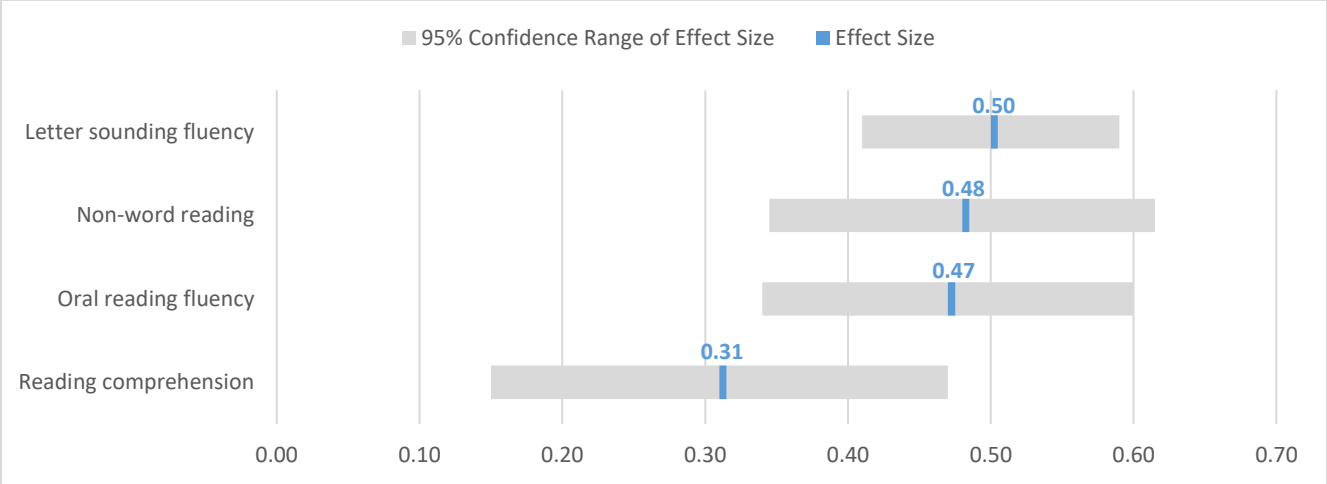


FIGURE 3.8: Adjusted Effect Sizes – Girls in Project Schools



4 Limitations and Context

The comparability of the project and comparison groups is the key technical limitation of the evaluation, while several program implementation factors provide additional context to help interpret the evaluation results. These factors include the learning curve associated with the launch of a new program and field staffing challenges that the team experienced in 2016.

4.1 Limitations

The key limitation of this evaluation was comparability. The validity of impact evaluation results rests on the strength of the assumption that the comparison schools, on average, are comparable to the group of project schools amongst all observable and unobservable characteristics that may affect the outcome being evaluated. Because project schools were not randomly assigned the intervention, it may be possible that they differed from comparison schools in ways that we did not assess. As explained in *Appendix B: Research Design*, every effort was made during sampling to ensure that a comparable set of schools was chosen, and school and learner-level characteristics were later analyzed to check whether any significant differences did in fact exist. As mentioned in Section 3.1, some observable differences did exist between the project and comparison groups, which we controlled for in the analysis of project outcomes.

4.2 Program Implementation Context

Several program implementation factors provide additional context with which to interpret the 2016 Grade 1 results. The instruction component of the Sepedi Literacy Program was revised in 2014 and 2015 as part of an effort to align Room to Read's Literacy Program globally with best practices in reading program design and delivery. Consequently, the 2016 reading skills assessment measures the first-year results of a newly implemented program. In the long-run, we expect the revised program to lead to higher learner outcomes than the previous version, but in the short-run, we recognize that the first year of implementation likely included a learning curve for field staff to fully master and effectively implement the new program.

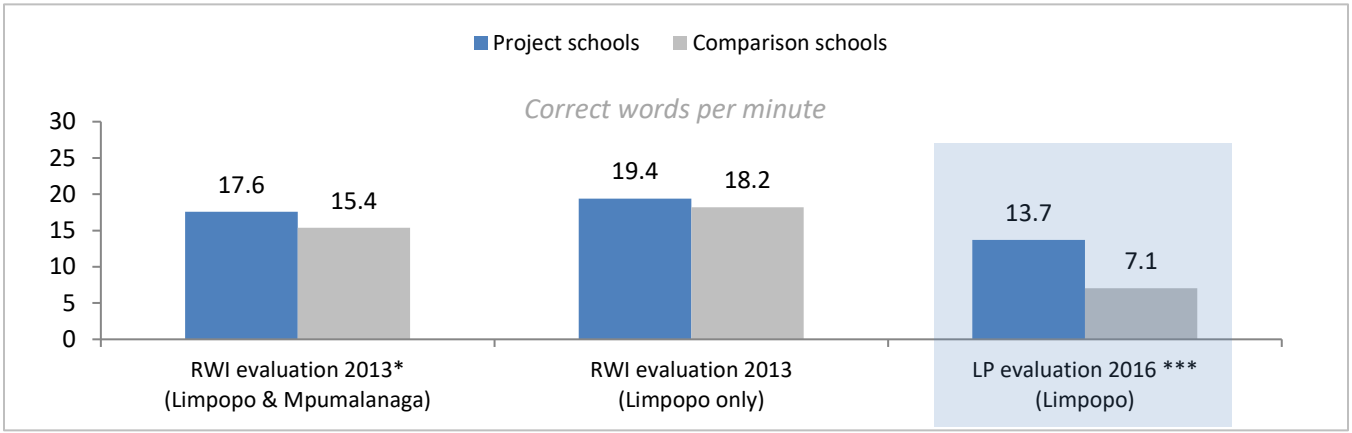
Additionally, staff transitions and hiring delays led to there being an insufficient number of Room-to-Read-hired Literacy Coaches to provide the required monitoring and coaching support to teachers in 2016. For much of the year, Sekhukhune District programs were short-staffed by almost half the required number of coaches, which reduced the regularity and quality of the guidance and support that was provided to Grade 1 teachers. As 2016 was also the first year of the program for these teachers, this more limited coaching support likely lengthened their process of adopting improved teaching methods over what may have been possible with more regular monitoring and support visits.

4.3 Results Context

Previous reading skills evaluation results in South Africa also provide context with which to interpret the 2016 results. Figure 4 below compares the 2016 Grade 1 Sepedi Literacy Program (LP) evaluation results (highlighted in the blue box) with results of a similar evaluation conducted to measure the impact of the earlier version of the

Sepedi Literacy Program (RWI) intervention in 2013.[5] The comparison shows that all learners in the LP evaluation sample achieved end-of-Grade 1 reading fluency levels lower than those achieved by learners in the RWI evaluation sample. When we restrict the 2013 evaluation sample to the learners and schools in the same province (Limpopo) as the 2016 evaluation sample, this difference is even stronger. However, while the 2013 evaluation found that the RWI program had no discernable impact on improving oral reading fluency for Limpopo province, the LP intervention had a large impact, almost doubling the reading skills of the project group over the comparison group.

FIGURE 4: End-of-Grade 1 Oral Reading Fluency Results in Sepedi, by Evaluation



Legend of statistical significance of differences between project and comparison schools: *** < .001, ** < .01, * < .05

5 Conclusion

Overall, the results showed that the Literacy Program had a positive impact on improving Grade 1 learners' reading skills. Learners in project schools made gains in reading skills from the beginning to end of Grade 1 that were approximately two times larger than those of comparison school learners. Thirty-eight percent of project school learners met the Grade 1 letter sounding benchmark, 30 percent reached the "emergent reader" benchmark, and 6 percent reached the "fluent reader" benchmark. The larger gains seen in the letter sounding task suggest that the Grade 1 instructional intervention was most effective at building letter-sound recognition skills.

As expected in the first year of a new program, the evaluation results also highlighted that room for improvement exists. A large percentage of project school learners remained unable to read or achieve the letter sounding benchmark by the end of Grade 1. Although the newly contextualized program is having a bigger impact (relative to comparison school children) than the earlier version of the program, the average reading levels achieved by learners benefiting from the new program were lower. The staffing challenges that limited literacy coaches' support to teachers and the learning curve associated with the rollout of the new program may have reduced the potential impact of the program in its first year.

[5] Interpretation of this comparison is limited, as the programs assessed were different in design and implementation, the assessments were conducted several years apart in different locations, and the assessment tasks may not have been of the same difficulty level.

In the coming years, Room to Read aims to improve Literacy Program implementation and thereby increase the number of learners reading fluently with comprehension. To build upon the positive results from this Grade 1 program evaluation, the Room to Read South Africa team proposes the following actions and strategies for 2017:

6 Next Steps

In the coming years, Room to Read aims to improve Literacy Program performance further and significantly raise the number of learners achieving reading fluency. To build upon the positive results from this Grade 1 program evaluation, the Room to Read South Africa team proposes the following actions and strategies for 2017:

1. **Better utilize formative assessment data** to target struggling students and ensure that resulting improvement strategies are implemented;
2. **Investigate why letter sounding performance is such a success** and the other tasks are still too difficult for the learners;
3. **Improve operational practices to ensure there are sufficient field staff** to support teachers and that the program is implemented with fidelity to its design;
4. **Utilize the school-by-school EGRA assessment results to target teachers** at low performing schools with additional support, and develop a strategy for sharing results with district and provincial education officials;
5. **Analyze data from classroom observation forms and program implementation monitoring** to better understand the quality of implementation and how this correlates to reading skills outcomes;
6. **Incorporate a discussion of assessment into teacher trainings** to help teachers utilize the assessment components to effectively revise and check learning.

Through these and other routine strategies to review and improve the Literacy Program, Room to Read hopes to achieve its goal of having all children in the program reading fluently with comprehension by the end of Grade 2.

7 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.

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.

Appendix A: Literacy Program Instruction Component Overview

The instruction component of the Literacy Program is a classroom intervention designed to complement and increase the effectiveness of the government language curriculum. The process of developing the intervention includes the completion of a scope and sequence of instruction, detailed lesson plans, classroom materials, and comprehensive teacher professional development. Literacy facilitators, or coaches, provide classroom support to teachers throughout the intervention.

During the research and development stage of the instruction component for any language, our country teams analyze the language curriculum and classroom instruction to determine whether all five core elements necessary in a comprehensive language curriculum are included. These elements, which are best addressed through a combination of listening, speaking, reading, and writing activities and lessons, include:

- **Phonological awareness**: Phonological awareness is knowing the sound structure of spoken language.
- **Phonics**: use of the code (sound-symbol relationships) to recognize words.
- **Vocabulary**: The knowledge of the meaning and pronunciation of words.
- **Fluency**: Fluency is determined by how quickly, accurately, and expressively someone reads, which, taken together, facilitate the reader's construction of meaning. It is demonstrated during oral reading through ease of word recognition, appropriate pacing, phrasing, and intonation. It is a factor in oral and silent reading which can limit or support comprehension (Kuhn et al., 2010).
- **Comprehension**: A definition of reading comprehension that captures the purpose of reading is "intentional thinking during which meaning is constructed through interactions between text and reader" (Harris & Hodges, 1995, p. 207). Reading comprehension consists of three elements: the reader, the text, and the activity of reading (Snow, 2002). Writing skills are incorporated into the instructional approach through all components. In addition, teachers teach children how to write and learner workbooks provide daily opportunities to practice the writing skills taught.

Appendix B: Research Design

Methodology

The impact evaluation employs a quasi-experimental design that includes learners from schools that benefit from the Literacy Program (project schools) and learners from schools that do not benefit from the Literacy Program (comparison schools). Data collection occurs at three points in time and follows the same cohort of learners over two academic years.[6] At the beginning of Year 1, a baseline assessment was conducted with project and comparison school learners who are entering Grade 1 to assess learners' reading level at the beginning of program exposure. Subsequent rounds of data collection assess progress of learners from the project school cohort after one and two years of the program in comparison with learners from the comparison school cohort.

The structure and design of the evaluation allows for an examination of the effects of the Literacy Program on learner achievement over time. The assessment results also help us understand learners' reading skill strengths and weaknesses and provide Room to Read staff, classroom teachers, and administrators with information on program efficacy. These data guide program improvement strategies to ensure that learners achieve the learning goals.

Sampling

The aim of sampling was to ensure that project and comparison schools were as similar as possible before the introduction of the Literacy Program. The project group initially consisted of 25 schools, randomly selected out of the 70 schools in which the Sepedi Literacy Program was launched in 2016. A list of all potential comparison schools in nearby wards of the same district as the project schools was compiled. As no objective indicator of recent school achievement was available, we could not stratify and match the project and comparison schools by achievement. Instead, a simple random sample of 25 comparison schools was selected from this group. Due to some challenges experienced during program site selection, one comparison school was ultimately selected to become a project school by program staff after baseline data collection completed. As a result, our evaluation sample consisted of 26 project school and 24 comparison schools.

At baseline and midline (i.e., the beginning and end of Grade 1), we randomly assessed 30 Grade 1 learners from each school. During learner-level sampling, we selected children who

- Did not have physical, sensory and significant cognitive disabilities[7]; and
- Were present on the day(s) of data collection.

In schools where grade-level enrollment was less than the desired sample size of 30 learners, all learners in the grade were assessed. This sampling procedure resulted in an achieved sample of 1,353 learners at baseline and 1,352 learners at the end-of-Grade 1 midline.

[6] During each data collection point (i.e., baseline, end of Grade 1, and end of Grade 2), a new sample of pupils was randomly selected from the same cohort of pupils in project and comparison schools. Though the same pupils may be selected at multiple data collection points, Room to Read is not intentionally following the same pupils over the two years of the study.

[7] We were not able to identify or exclude children with learning and/or reading disabilities as such disabilities are difficult to detect in Grade 1 and 2.

Reading Assessments

In this evaluation, Room to Read assessed learners' literacy skills using a version of the Early Grade Reading Assessment (EGRA)[8] that was adapted from English to Sepedi by local experts. Room to Read used a version of the EGRA that was designed according to the expected reading levels of Grade 2. The EGRA was comprised of four common tasks:

- *Letter sounding fluency:* Ability to sound letters of the alphabet without hesitation and naturally. This is a timed test that assesses automaticity and fluency of letter recognition. Children are given one minute to sound 100 letters.
- *Non-word fluency:* 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 are given one minute to read 50 words.
- *Passage reading fluency:* Ability to read an approximately 60-word passage that tells a story. Children are given up to three minutes to read the passage. The assessment rotated through four different versions of the reading passage.
- *Reading comprehension:* Ability to answer five questions based on the passage.

Assessments were administered individually with learners by external data collectors using tablets.

Enumerator Training

In total, 18 individuals were trained to administer the midline reading assessments, including 12 enumerators, 4 team leaders, and 2 supervisors. These individuals were recruited by Room to Read as native speakers of Sepedi with, at minimum, a post-secondary diploma or degree and some data collection experience; several had also assisted Room to Read with previous data collection activities.

The training lasted five days and was facilitated by Room to Read's Research Officer. The training content included background information on Room to Read and the purpose of the assessment, introduction to the EGRA tool and tablet technology, in-depth review of the different assessment tasks, enumerator skills and data collection procedures, data quality considerations, school visit and random selection procedures, and the logistical arrangements for field work. The training also included significant hands-on practice, including two days of field testing spread across eight schools, and three inter-rater reliability tests to verify the accuracy and consistency of the team's data collection. At the end of the enumerator training, the average enumerator inter-rater reliability score was 95 percent.

Data Collection

Data collection took place from October 17 to Nov 4, 2017 at a total of 50 schools – 26 from the project group and 24 from the comparison group. There were four data collection teams, each consisting of three enumerators and one team leader; each typically collected data from one school per day. Each team was accompanied to the field by one of the two supervisors or Room to Read's Research Officer or Research Manager on the first days of data collection, and alternately supported by one of these individuals on later days. Team leaders took the lead in organizing data collection at their schools, while supervisors provided oversight and assisted with data

[8] The EGRA was developed by RTI International in 2006. For more information, please see: www.ineesite.org/uploads/files/resources/EGRA_Toolkit_Mar09.pdf.

verifications and quality assurance. The Research Officer also conducted daily team debriefs and reviewed collected data on a daily basis to ensure proper assessment administration.

Data Entry

Data was entered directly into the tablet by enumerators during data collection and was uploaded onto the cloud at the end of each day. The Research Officer conducted daily review of the downloaded data to check for completeness and proper assessment administration and to make minor corrections. Once data collection was complete, the full dataset was shared with Room to Read's Global Office for further quality checks and data analysis.

Data Analysis[9]

The first aim of the data analysis was to determine if there were significant differences in school and learner background variables between the project and comparison groups. For the school background variables, we examined differences in mean school enrollment and pupil-teacher ratio by conducting *t* tests. We also examined differences in school location (rural vs. semi-urban vs. urban) using chi-square tests. For the learner background variables, we examined differences in age, gender, learner participation in pre-school, home language, whether the child had television at home, whether the child had a collection of books at home, and mother's literacy by conducting regression analysis (linear regression for the continuous variables and logistic regression for the categorical variables) with random effects at the school level. The equations included the learner background variables (age, gender, etc.) as the dependent variable and school type (project or comparison) as the independent variable. The results of these analyses appear in *Appendix D*.

The second aim of the data analysis was to determine if learners in the project group made greater gains from baseline to the end of Grade 1 than learners in the comparison group. The analysis strategy was to compare reading levels in the two assessment periods (baseline versus end of Grade 1) among the two experimental groups (project versus comparison). An impact of the Literacy Program is evident if there is a greater gain from baseline to end of Grade 1 among the project groups compared to the comparison group. This is demonstrated by a statistically significant interaction between experimental group and assessment period. To determine this, we conducted linear regression analysis 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 end-of-Grade-1 assessment scores as the dependent variable and age, gender, whether the child had a collection of books at home, and pupil-teacher ratio as covariates[10]. We followed a similar procedure to analyze differences in gains by gender. We created a dummy variable for gender and conducted linear regression analysis with random effects at the school level to examine differences in gains across each of the following: project school boys versus comparison school boys and project school girls versus comparison school girls. Each analysis included one of the assessment scores as the dependent variable and age, whether the child had a collection of books at home, and pupil-teacher ratio as the covariates.

Next, we examined the percentage of learners from each group achieving the Grade 1 fluency target of 45 letter sounds per minute. We conducted logistic regression analysis with random effects at the school level to determine if significant differences existed between the percentages of project school versus comparison school learners

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

[10] Age and gender were included in the regression model because of their known effects on pupils' reading performance. Whether a child had a collection of books at home and pupil-teacher ratio were included of a significant differences between project and comparison school groups for these variables at midline (see *Appendix D*).

achieving the fluency target. The equations included a dummy variable for whether learners achieved the target as the dependent variable, school type as the independent variable, and age, gender, and pupil-teacher ratio as the covariates. A similar analysis was done to determine the percentage of learners reaching the "emergent reader" benchmark for oral reading fluency.

Finally, we analyzed zero scores 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). We conducted logistic regression analysis with random effects at the school level to determine if significant differences existed between the percentage of project school vs. comparison school learners registering zero scores (vs. non-zero scores). The equations included the presence or absence of a zero score as the dependent variable, school type as the independent variable, and age, gender, whether the child had a collection of books at home, and pupil-teacher ratio as covariates.

Appendix C: School and Learner Background Characteristics

Table C.1: Background Characteristics of Sample Schools

	Project Schools		Comparison Schools	
	n	Mean (SD) or %	N	Mean (SD) or %
District				
Sekhukhune	26	100%	24	100%
Location	26	-	26	-
Urban	0	0%	0	0%
Semi-urban	2	8%	0	0%
Rural	24	92%	24	100%
School Enrollment – Total				
Baseline**	26	525.27 (215.93)	24	341.75 (193.98)
Midline**	26	519.38 (228.25)	24	329.42 (184.79)
School Enrollment – Grade 1				
Baseline**	26	67.77 (34.04)	24	43.54 (26.76)
Midline**	26	68.08 (31.47)	24	41.88 (23.98)
Pupil Teacher Ratio - Overall				
Baseline	26	36.25 (5.44)	24	32.57 (10.08)
Midline	26	35.78 (7.08)	24	33.61 (8.48)
Pupil Teacher Ratio – Grade 1				
Baseline**	26	44.33 (14.27)	24	32.57 (10.88)
Midline*	26	42.05 (15.66)	24	32.71 (15.01)
Percentage of Learners Present				
Baseline	26	97%	24	98%
Midline	26	94%	24	95%

Legend of statistical significance: *** p < .001, ** p < .01, * p < .05.

TABLE C.2: Background Characteristics of Sample Learners

	Project Schools		Comparison Schools	
	n	Mean (SD) or %	n	Mean (SD) or %
Baseline learners	754	-	599	-
Male	378	50%	307	51%
Female	376	50%	292	49%
End of Grade 1 learners	758	-	594	-
Male	374	49%	304	51%
Female	384	51%	290	49%
Age				
Baseline	751	5.69 (0.59)	598	5.73 (0.58)
End of Grade 1	748	6.50 (0.60)	592	6.55 (0.60)
Attended preschool				-
Baseline	752	99%	598	100%
End of Grade 1	758	98%	594	98%
Home language is Sepedi				
Baseline	754	99%	599	100%
End of Grade 1	758	95%	594	97%
Has a television at home:				
Baseline*	754	89%	599	84%
End of Grade 1	758	82%	594	84%
Has a collection of books at home:				
Baseline	751	86%	598	82%
End of Grade 1***	758	59%	594	70%
Mother can read and write:				
Baseline*	753	86%	595	82%
End of Grade 1	757	82%	593	84%

Legend of statistical significance: *** p < .001, ** p < .01, * p < .05.