



# **Two-year (2021-22) Impact Evaluation of the Literacy Program in Tanga Region in Tanzania**

## **Evaluation Report**

**April 2023**

*Prepared by*

Pinaki Joddar, Rebecca Westbrook and Japheth Ogot  
Global Research Monitoring & Evaluation Team, Room to Read

*With support from*

Room to Read Tanzania

*This page is left blank intentionally.*

## Table of Contents

<b>Executive Summary</b> .....	1
<b>1. Introduction</b> .....	3
1.1 Literacy Program in Tanzania .....	3
1.2 Evaluation of the Literacy Program .....	3
1.3 Evaluation Methodology .....	3
1.4 Sampling .....	5
1.5 Assessment Tool .....	7
<b>2. Results</b> .....	9
2.1 Assessment Scores .....	9
2.2 Fluency and Comprehension Benchmarks .....	11
2.3 Zero Scores .....	12
2.4 Comparisons by Gender .....	13
<b>3. Limitations</b> .....	16
<b>4. Conclusion</b> .....	18
<b>References</b> .....	19
<b>Annex A: Room to Read's School Selection Criteria – An Abridged Version</b> .....	20
<b>Annex B: Data Analysis</b> .....	21
<b>Annex C: School and Children Background Characteristics</b> .....	23
School background characteristics during 2022 endline .....	23
Children background characteristics during 2021 baseline and 2022 endline .....	23
<b>Annex D: Correlation between Assessment Scores during 2022 Endline</b> .....	25

## List of Figures and Tables

Figure 1.1: Visual presentation of difference-in-difference methodology .....	4
Table 2.1: Status of children's early grade literacy skills and the impact of the Program .....	9
Figure 2.1: Effect sizes across indicators .....	10
Figure 2.2: Children's distribution across oral reading fluency score categories during endline .....	11
Figure 2.3: Share of children answering reading comprehension questions correctly during endline .....	12
Figure 2.4: Zero scores during endline .....	13
Figure 2.5: Effect sizes across indicators for boys .....	14
Figure 2.6: Effect sizes across indicators for girls .....	14
Table 2.2: Status of children's early grade literacy skills and the impact of the Program by gender .....	15

## Executive Summary

Room to Read's Literacy Program (or the *Program*) in Tanzania is a school-based intervention that aims to develop children's literacy skills in early grades and the habit of reading among all primary grade children to help them become life-long, independent readers. The Program includes three key components supplementing each other, such as early grade literacy instruction in Kiswahili language, child-friendly school library for all primary grade children, and development and/or procurement of quality reading materials for children. For the instruction component, Room to Read worked with the Government of Tanzania to develop early grade teaching and learning resources that introduce phonics content at an appropriate pace and promote the development of key reading and writing skills. The intervention also includes detailed language lesson plans, materials for use during language classes and library periods, manuals for library management and reading activity during library periods, professional development training for teachers and librarians, family and community engagement activities, and school-based monitoring and evaluation in collaboration with local government and school authorities, respectively. The instruction component of Room to Read's Literacy Program was first introduced in 2016 in Bagamoyo district of Pwani region. In 2020 and 2021, the Program was initiated in a set of government primary schools in two new districts, Muheza District Council (DC) and Tanga City Council (CC), of Tanga region.

Results of the Impact Evaluation				
Task	Group	Mean Score – Baseline (2021)	Mean Score – Endline (2022)	Impact <sup>^</sup>
Letter sound	Project	2.60	21.59	<b>8.37***</b>
	Comparison	1.25	11.90	
Nonword reading	Project	1.32	14.44	<b>2.37***</b>
	Comparison	1.01	11.50	
Sentence choice	Project	0.27	5.55	<b>0.84***</b>
	Comparison	0.21	4.49	
Oral passage reading	Project	1.65	25.81	<b>5.6***</b>
	Comparison	1.12	18.99	
Reading comprehension	Project	0.35	3.08	<b>0.31**</b>
	Comparison	0.24	2.59	
Sentence dictation	Project	4.10	6.02	<b>0.28</b>
	Comparison	3.87	5.45	

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ . <sup>^</sup>Impact: Difference between project and comparison groups on the gains in literacy skills from baseline to endline.

To determine the Program's effect on children's early grade literacy skills, Room to Read initiated a two-year (2021-22) impact evaluation study in the two districts of Tanga region in 2021. The evaluation included 50 Program schools, defined as *project* group, where a new cohort of Grade 1 children started benefiting from the intervention from 2021 school year. The project group schools were selected randomly in proportion to the distribution of 120 Literacy Program schools across Muheza DC and Tanga

CC, respectively. The evaluation also included 50 non-intervention schools, defined as *comparison* group, from the same two districts; these schools were selected randomly from a list of non-Program schools that had observable pre-intervention background characteristics similar to project group schools. In this evaluation, Room to Read assessed children's early grade literacy skills using a tool that was prepared on the framework of Early Grade Reading Assessment (EGRA) by in-house local language and curriculum experts. The assessment tool included six tasks, such as letter sound, nonword reading, sentence choice, oral passage reading, reading comprehension, and sentence dictation. Using a quasi-experimental design, this study evaluated the Program's impact by comparing the gains in children's early grade literacy skills from the beginning of Grade 1 (baseline) to the end of grade 2 (endline) from project and comparison groups.

The baseline data collection was done in February-March 2021 in the 50 project and 50 comparison group schools, respectively. In each school, early grade literacy skills of a sample of start-of-Grade 1 children were assessed during the 2021 baseline. Baseline results showed that children across project and comparison groups entered Grade 1 with comparably low levels of early grade literacy skills. Endline evaluation was conducted in November 2022 in the same schools. In each school, early grade literacy skills of a freshly drawn sample of end-of-Grade 2 children were assessed during the 2022 endline.

Overall, the study showed that children from project group schools outperformed comparison group school children during 2022 endline. Moreover, across tasks, project group children made gains in their early grade literacy skills between 2021 baseline and 2022 endline that were 1.2 to 1.8 times greater than those seen in comparison group schools. Differences in the two-year gains between project and comparison group children were statistically significant on letter sound ( $p < 0.001$ ), nonword reading ( $p < 0.001$ ), sentence choice ( $p < 0.001$ ), oral reading ( $p < 0.001$ ), and reading comprehension ( $p < 0.01$ ) tasks, respectively. Also, a significantly higher proportion of Grade 2 children in project schools in contrast to comparison group achieved the benchmarks on (i) oral reading fluency of correctly reading 45 or more words per minute, and (ii) reading comprehension of scoring 80 percent or more, respectively. Across all early grade literacy skills assessed in this study, the maximum impact of the Program was seen on the letter sound task as per the effect size statistics. The evaluation also found that the Program benefitted both girls and boys; however, girls seemed to have benefitted more from the intervention than boys.

Despite the program's positive outcomes after two years of intervention, the study provided ample evidence that indicates scope for further improvements in the effectiveness of the Program. Low average scores by Grade 2 children in project schools during the 2022 endline and small effect sizes on most of the literacy skills testify to this. Moreover, a noticeable number of children remained as non-readers in project schools even after two years of intervention as shown by the statistics on zero scores across assessment tasks. Also, the Program did not make any noticeable impact on children's early grade writing skills as per the results from the sentence dictation task.

Room to Read recognizes that the disruptions to the Program implementation due to Covid-19 might have played a crucial role in affecting its effectiveness. The shift from onsite delivery to remote delivery of the integral elements of the Program due to the lack of physical access to schools during the pandemic was sudden. However, the situation has improved a lot in recent times, and Room to Read in Tanzania is committed to continuing its effort to improve children's early grade literacy skills in the coming days in collaboration with the government. This commitment and buy-in from the government about the Literacy Program are evident from the recent expansions of the intervention in new geographies in 2022 and 2023 school years, respectively.

## **1. Introduction**

### **1.1 Literacy Program in Tanzania**

Room to Read's Literacy Program<sup>1</sup> in Tanzania is a school-based intervention that aims to develop children's literacy skills in early grades and habit of reading among all primary grade children to help them become life-long, independent readers. The Program includes three main components: (i) reading and writing instruction in Kiswahili for children in grades 1 and 2; (ii) establishment of child-friendly school libraries providing access to quality materials to read inside and outside schools for all primary grade children; and (iii) development and/or procurement of locally appropriate quality children's materials. For the instruction component, Room to Read worked with the Ministry of Education, Science, and Technology, Government of Tanzania to develop early grade teaching and learning resources that introduce phonics content at an appropriate pace and promote the development of key reading and writing skills. The intervention also includes detailed language lesson plans, materials for use during language classes and library periods, manuals for library management and reading activity during library period, professional development training for teachers and librarians, and family and community engagement activities, respectively. These elements of the Program are complemented by concurrent school- and classroom-based mentoring and coaching for the early grade language teachers and school librarians by Room to Read appointed Literacy Facilitators, and monitoring and evaluation of the intervention, respectively. The Program provides a solid foundation for all children in Room to Read supported government primary schools, with the goal that children will become fluent readers by the end of Grade 2. In Tanzania, the Kiswahili instruction component of the Program was first launched during the 2016 school year in Grade 1 classrooms in Bagamoyo district of Pwani region. In 2020 and 2021, the Program was initiated in a set of primary schools in two new districts, Muheza District Council (DC) and Tanga City Council (CC), of Tanga region.

### **1.2 Evaluation of the Literacy Program**

To determine the effect of the intervention on children's early grade literacy skills in Kiswahili, Room to Read initiated a two-year (2021-22) impact evaluation of the Literacy Program in the two districts of Tanga region in 2021. Objectives of the impact evaluation were as follows:

- Evaluate the impact of the Program on children's early grade literacy skills.
- Determine whether the implementation of the Program facilitated the acquisition of early grade literacy skills at a rate that ensures children would become fluent readers by end of Grade 2.
- Identify reading and writing skills that could be better supported by the Program and determine how to improve these skills quickly and effectively.

Room to Read, with support from an independent data collection agency, conducted this impact evaluation following the methodology described below.

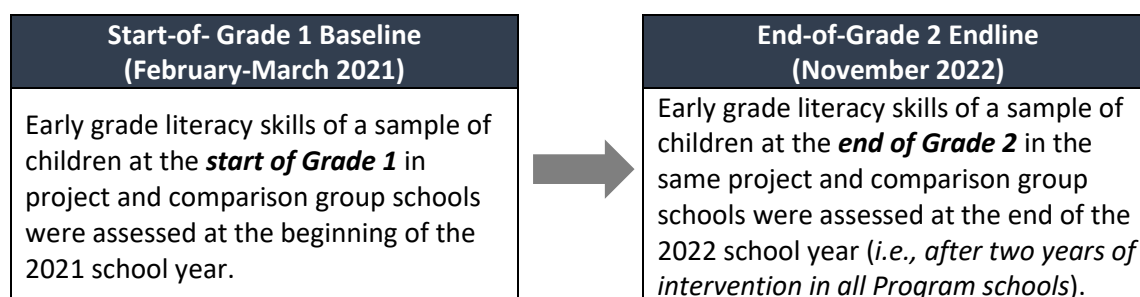
### **1.3 Evaluation Methodology**

Room to Read was implementing the Program in 120 government primary schools in Muheza DC and Tanga CC at the time when this evaluation was planned. These schools were selected for participation in

---

<sup>1</sup> Also referred as the Program or intervention alternatively in this document.

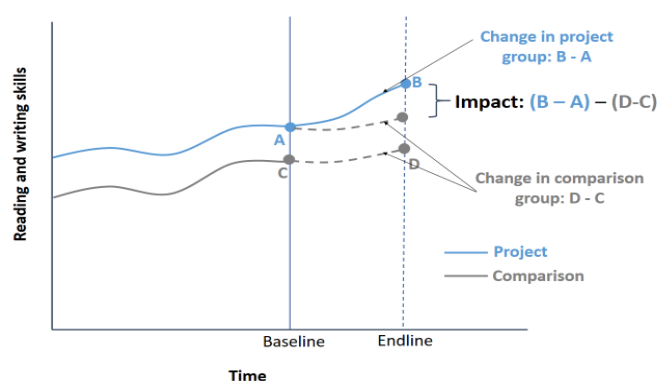
the intervention non-randomly, using pre-decided school selection criteria<sup>2</sup> in consultation with the government. As a result, traditionally used experimental methods for impact evaluation like randomized control trial was not deployed for this two-year study. Instead, Room to Read opted for a quasi-experimental impact evaluation methodology. Here, following a semi-longitudinal study design, early grade literacy skills data were collected at two different points of time over two academic years from (i) children in 50 out of 120 government primary schools benefiting from the Program (defined as *project* group) and (ii) children in another 50 government primary schools not benefiting from the Program, but having observable pre-intervention background characteristics similar to the project group schools (defined as *comparison* group). The following two boxes present the details of the two waves through which data on children's early grade literacy skills were collected from project and comparison group schools included in the evaluation sample frame for the study.<sup>3</sup>



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

- **Category 1 (first difference):** Gains in children's early grade literacy skills between start-of-Grade 1 baseline and end-of-Grade 2 endline in project group schools after two years of Program intervention; and
- **Category 2 (second difference):** Gains in children's early grade literacy skills between start-of-Grade 1 baseline and end-of-Grade 2 endline in comparison group schools without getting the benefits of the Program.

**Figure 1.1: Visual presentation of difference-in-difference methodology**



<sup>2</sup> Room to Read uses a host of criteria to decide about the intervention schools. These criteria are customized in each intervention geography and final school selection are always done in consultation with local stakeholders like schools and government. An abridged and generic version of Room to Read's global school selection criteria is included in *Annex A*.

<sup>3</sup> During each data collection wave (*i.e., baseline, and endline*), Room to Read conducted assessments with a fresh sample of randomly selected children from the appropriate grade in project and comparison schools. Though a particular child may get selected in both baseline and endline, Room to Read did not intentionally track the same children over the two years of this study.

Comparison of the difference between the above two categories enabled Room to Read to statistically estimate the extent of influence (measured through *effect size*) of the Program on children's early grade literacy skills in project group schools. Also, very importantly, this methodology helped Room to Read to establish a direct causal relationship between the extent of influence of the Program on children's early grade literacy skills and the intervention. In impact evaluation literature, this methodology is defined as "*Difference-in-difference (DID)*" or "*Double difference (DD)*" and a visual presentation of the same is provided in Figure 1.1. Additionally, the evaluation results also helped in understanding the strengths and weaknesses in children's reading and writing skills and provided Room to Read, language teachers, school administrators, and the government with information that would help in refining the Program design and implementation process in the future.

## 1.4 Sampling

Schools were the primary sampling units in this evaluation. Through power calculations, it was determined that a sample size of 100 schools (50 units each from the project and comparison groups) with 10 children per school would be sufficient to detect a minimum effect size of 0.4<sup>4</sup> from the evaluation (assuming a level of significance to be 0.05, power to be 80%, baseline-endline correlation to be 0.7, and an intra-cluster correlation of 0.2)<sup>5</sup>. Consequently, the aim of the sampling was to select the required number of project and comparison schools for the evaluation that were as similar as possible at the time when the Program was starting with a new cohort of start-of-Grade 1 children at the beginning of 2021 school year in all 120 intervention schools in Tanga region. Following steps were taken to select this sample.

First, a list of all Program schools in the two districts of Tanga region (Muheza DC and Tanga CC) was assembled. From the list, 50 Program schools were selected randomly from the two districts in proportion to the distribution of all intervention schools across districts. As a result, 30 and 20 Program schools were selected for the project group for evaluation from Muheza DC and Tanga CC, respectively. For comparison group sampling, a list of non-intervention schools in the same two districts that met the general selection criteria for the Program was prepared. Then thirty schools from Muheza DC and 20 schools from Tanga CC were randomly selected from the list as comparison group sample for the study.

During 2021 start-of-Grade 1 baseline and 2022 end-of-Grade 2 endline data collection, information on an exhaustive list of school background characteristics was collected and analysed to verify the similarity between project and comparison group schools. For 2022 endline, analyses of school background characteristics did not reveal any statistically significant difference between project and comparison groups on the following indicators: enrollment in target grade<sup>6</sup>, attendance in target grade on the day of assessment, pupil to classroom ratio in target grade, monitoring/support visits from local (government) education officials in last 3 months, school having functioning school committee, school organized outreach event(s) in last 6 months, and classroom environment score<sup>7</sup>. However, analyses also showed

---

<sup>4</sup> Based on historical minimum detectable effect sizes observed in various impact evaluations for the Literacy Program conducted by Room to Read across countries and languages.

<sup>5</sup> Assumptions for level of significance and power were based on general practices in impact evaluations, and assumptions for correlations between baseline and endline, and intra-cluster correlations were based on historical data from Room to Read Literacy Program evaluation studies, respectively.

<sup>6</sup> Target grade: Grade 1 for 2021 baseline and Grade 2 for 2022 endline.

<sup>7</sup> During each wave of data collection, information on availability and condition of the following seven items in a classroom for the target grade in each sampled school was collected: *blackboard or whiteboard, visual aids on the wall, teacher guide, student textbooks/workbooks, adequate seating, intact roof without signs of damage, and intact wall without visual structural damage.*



statistically significant differences between project and comparison schools on the following indicators in 2022 endline: location (rural or urban/semi-urban;  $p < 0.001$ ), pupil to teacher ratio in target grade ( $p < 0.001$ ), at least one teacher in target grade received training on teaching language at the beginning of school year ( $p < 0.001$ ), and school facilities score<sup>8</sup>. Statistical comparisons of early grade literacy skills assessment results between project and comparison groups presented in this report accounted for these differences in the school level background characteristics at endline.<sup>9</sup>

Children were the final sampling units in this evaluation. During baseline and endline data collection, attempts were made to randomly select 10 children from each school for the assessment who: (i) did not have physical, sensory, and significant cognitive disabilities<sup>10</sup>, and (ii) were present in the school on the day(s) of data collection. Also, it was decided that all children present in the target grade in a school would be selected for assessment if attendance in the grade on the day of visit is less than 10. Moreover, assessors were instructed not to deny a child to participate in the assessment if s/he expresses willingness, even if the child was not part of the children sample.

During the study, the data collection was faced with several challenges. One such key challenge was restricted physical access to the schools due to Covid-19 during 2021 baseline. The data collection team could not physically verify the status of enrollment and attendance in Grade 1 prior to the actual administration of 2021 baseline assessment in schools located in several clusters. It was decided that a total of 16 children (i.e., six additional children on top of 10 children) in Grade 1 would be targeted for the assessment during 2021 baseline. However, the situation was much more conducive during 2022 end-of-Grade 2 endline, and the target sample size was brought down to the original level of 10 children in Grade 2 per school. The children sampling process as described here yielded a sample size of 1,539 start-of-Grade 1 children (766 from project schools and 773 from comparison schools) in 2021 baseline and 1,000 end-of-Grade 2 children (500 in project schools and 500 in comparison schools) in 2022 endline, respectively.

Comparability of the children sampled from project and comparison group schools is another vital factor that lends credence to the findings on the impact of the Program. During each wave of data collection, information on a set of children background characteristics were collected and analysed to verify this. In both baseline and endline, no statistically significant difference was found between the children sampled from project and comparison groups on the following indicators: gender, age, household size, mother's literacy, anybody supports in studies at home, anybody read something to the child in the last week, anybody told a story to a child in the last week, collection of children's book at home, and availability of other types of printed materials at home. However, on the following indicators, a statistically significant difference was found between the children sampled from project and comparison

---

For each available item in acceptable condition, a classroom was given with a score of one. Sum of all items was then calculated as the *classroom environment* score for a school.

<sup>8</sup> For each school, data on availability of the following 10 facilities were collected during baseline and endline: *electricity, potable drinking water, separate toilets for boys and girls, clean toilets, fence or boundary around the school, playground, information technology equipment for students' use, library, school feeding program, and paved road leading to the school*. For each available facility, a school was given with a score of one; sum of all facilities in each school was then calculated as the *school facilities* score.

<sup>9</sup> During 2021 baseline, no statistically significant difference between project and comparison groups was found for majority of the school background characteristics. Few school characteristics on which there was statistically significant difference between the two groups at baseline were used as additional controls in the regression analyses conducted on the data. The full list of school-level controls included in the analyses are provided in *Section 2: Results* of the report.

<sup>10</sup> During data collection, no attempts were made to identify and/or exclude children with learning and/or reading and writing disabilities as such disabilities are difficult to detect in early grades.

groups: child speaking Kiswahili at home ( $p < 0.001$  in baseline and  $p < 0.05$  in endline), absent from the school on any day in the last week ( $p < 0.05$  only in endline), had all three main meals in last 24 hours ( $p < 0.01$  only in endline), presence of both parents at home on most days in the last week ( $p < 0.01$  only in endline), spent time with anyone else like a caretaker other than the parents at home in the last week ( $p < 0.05$  in baseline and  $p < 0.001$  in endline), read at least once without prompting or for fun in last week ( $p < 0.01$  only in baseline), and household standard of living score<sup>11</sup> ( $p < 0.001$  only in baseline). Statistical comparisons of early grade literacy skills assessment results between project and comparison groups accounted for the differences in the children background characteristics (either in baseline or endline or both) mentioned above.<sup>12</sup>

## 1.5 Assessment Tool

In this impact evaluation, Room to Read assessed children's early grade literacy skills in Kiswahili language using a tool prepared on the framework of Early Grade Reading Assessment (EGRA) by in-house local language and curriculum experts.<sup>13</sup> The tool contained six early grade literacy skills assessment tasks, with difficulty level for each task fixed at Grade 2 level, as outlined below.

- **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, a child was asked to read as many letters as possible from a menu of 100 letters in 60 seconds with the correct sound. Data were analyzed to calculate the *letter sound fluency* score, measured in the units of *correct letters per minute*, for each child.
- **Nonword reading:** Ability to read words that do not exist, but whose letter combinations follow the rules of the language. This task assesses a child's ability to "decode" words fluently as distinct from her ability to recognize words seen before. Here, a child was given 60 seconds to read up to 50 nonwords correctly. For each child, data were analyzed to calculate the *nonword reading fluency* score, measured in the units of *correct nonwords per minute*.
- **Sentence choice**<sup>14</sup>: Ability to read and understand sentences. For this task, a child was presented with 10 pairs of sentences, with each pair containing one true and one matching false sentence placed randomly in the list. The child was given a score of one if the nature of both sentences in a pair was identified correctly. Then the sum of all pairs out of 10 identified correctly was calculated as the *sentence choice* score. Total time assigned for this task was five minutes.
- **Oral passage reading:** Ability to read a story fluently with comprehension. A child was asked to read a 49-word long passage in three minutes in this task. Then, the *oral reading fluency* score

---

<sup>11</sup> For each child, (self-reported) data on availability of the following nine items/ amenities at home were collected during baseline and endline: *mobile phone, refrigerator, television, bicycle, motorcar, electricity, water supply, latrine, separate kitchen, and paved floor*. For each available item/ amenity, household of a child was given with a score of one. Sum of all items/ amenities at home for each child was then calculated as the *household standard of living* score.

<sup>12</sup> Data on school and children background characteristics are presented in summary format in *Annex C*.

<sup>13</sup> Room to Read implements the Literacy Program in several countries in various languages. For a summative evaluation of the Program (e.g., *2-year impact evaluation*) in each language Room to Read prepares an assessment tool following a rigorous process of desk-based drafting of the tool and multiple rounds of field testing. The version of the Kiswahili early grade literacy assessment tool used in this study was prepared, tested, and finalized in 2019-20. Preparation of such tool in a language is a highly resource intensive activity and Room to Read uses the same tool for four to five years for all types of summative Program evaluation for the language. For this reason, Room to Read keeps the detailed content of the assessment tool used in a summative evaluation confidential. However, Room to Read provides a high-level summary of the tool and few additional analyses (e.g., correlation between scores across assessment tasks as presented in *Annex D*) in the evaluation report.

<sup>14</sup> Sentence choice is an alternative measure of reading comprehension skills. This task is delinked from the combined tasks of oral passage reading and reading comprehension that are being used traditionally to assess both fluency and comprehension.

measured in the units of *correct words per minute*, was calculated for the child using the data on number of words read correctly and total time spent reading.

- **Reading comprehension:** Assessment of comprehension skill based on response to a set of questions from the oral reading passage. A child was asked up to five questions depending on the length of the story read during the oral passage reading task. Then *reading comprehension* score of the child was calculated as the total number of questions answered correctly. The questions were evenly placed across the length of the story. The first four questions were fact-based or literal, and the last question was simple inferential in nature.<sup>15</sup>
- **Sentence dictation:** Ability to write words and a sentence. A child was asked to write a seven-word long Kiswahili sentence that was read aloud by an assessor. The written sentence was then scored on a 13-point scale based on correct spelling of each word and syllable, space between each word, capitalization of appropriate letters, direction of writing (left to right in Kiswahili), and punctuation, respectively.

Assessments were administered individually to each child by assessors from DataVision International, an independent data collection firm hired by Room to Read.<sup>16</sup>

---

<sup>15</sup> Correct answer to a fact-based or literal question was mentioned in the story. But correct answer to the simple inferential question was not mentioned in the story; a child was required to answer it based on his/her understanding of the whole story.

<sup>16</sup> In this exercise, a Tanzania-based reputed survey agency, Datavision International (DVI), was engaged by Room to Read for baseline and endline data collection as per the methodology and timelines mentioned in the report. As part of this assignment DVI hired a pool of experienced assessors and conducted a 5-day long training workshop for them with support and supervision from Room to Read. The training included: (i) desk-based training on how to administer early grade literacy skills assessment and collect data on school and children background information, and (ii) field practice in schools that were not included in the evaluation sample frame, respectively. Multiple rounds of inter-rater reliability test were also conducted with the assessors during the training. During the evaluation, early grade literacy skill assessment was administered to the children individually by an assessor. These assessments as well as collection of school and children background characteristics data were performed through tablets with the help of SurveyCTO software. In each school the entire data collection process was overseen by a dedicated supervisor and data quality assurance processes were performed by both DVI and Room to Read separately.

## 2. Results

### 2.1 Assessment Scores

Status of children's early grade literacy skills during 2021 Grade 1 baseline and 2022 Grade 2 endline across project and comparison groups and the impact of the Literacy Program in Tanga region are presented in Table 2.1 below, respectively. **Overall, children from project group schools outperformed comparison group school children during endline.**

*Table 2.1: Status of children's early grade literacy skills and the impact of the Program*

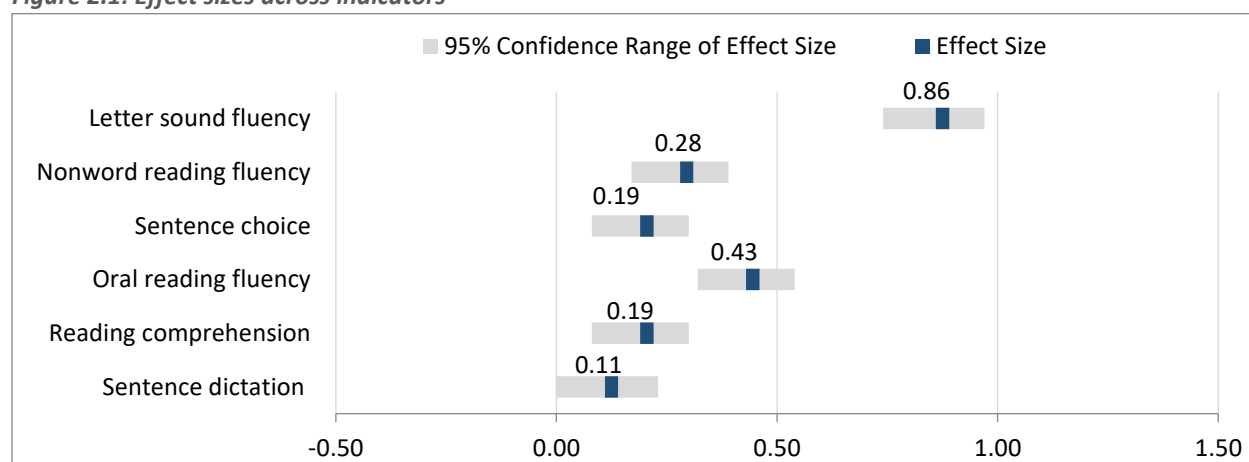
Indicator (unit of measurement)	Group	Start-of-Grade 1 Baseline (2021)			End-of-Grade 2 Endline (2022)			2-year Gains	Diff. in the 2-year Gains b/w Groups^
		N	Mean	SD	N	Mean	SD		
Letter sound fluency (correct letters per minute)	Project	766	2.60	6.40	500	21.59	17.84	18.99	8.37***
	Comparison	773	1.25	3.70	500	11.90	16.47	10.64	
Nonword reading fluency (correct nonwords per minute)	Project	766	1.32	3.47	500	14.44	8.57	13.13	2.37***
	Comparison	773	1.01	2.72	500	11.50	7.91	10.49	
Sentence choice score (score on a 10-point scale)	Project	766	0.27	0.84	500	5.55	3.66	5.29	0.84***
	Comparison	773	0.21	0.65	500	4.49	3.63	4.29	
Oral reading fluency (correct words per minute)	Project	766	1.65	4.69	500	25.81	16.66	24.16	5.6***
	Comparison	773	1.12	3.52	500	18.99	13.83	17.87	
Reading comprehension score (score on a 5-point scale)	Project	766	0.35	0.95	500	3.08	1.57	2.73	0.31**
	Comparison	773	0.24	0.72	500	2.59	1.63	2.36	
Sentence dictation score (score on a 13-point scale)	Project	766	4.10	1.90	500	6.02	3.71	1.92	0.28
	Comparison	773	3.87	1.48	500	5.45	3.81	1.58	
SD: Standard Deviation. *** $p < 0.001$ , ** $p < 0.01$ , * $p < 0.05$									
^Diff. in the 2-year Gains b/w Group i.e., the impact of the Program: Reports the estimated difference in gains from 2021 baseline to 2022 endline on each task score between project and comparison groups. An estimated difference was the coefficient of an independent variable – interaction between data collection wave (baseline or endline) and group (project or comparison) – from a regression, with random effects at school level, using a task score as dependent variable. Each regression on a task score also included the following covariates (i.e. controls): district, location, learner-teacher ratio, teacher training, school facilities score, age, gender, speaks Kiswahili at home, missed school in last week, had all 3 main meals in last 24 hours, both parents were present on most days at home in the past week, spent time with a caretaker in the past week at home other than parents, read without prompting/ for fun at least once in the past week, and household standard of living score.									

On average, Grade 2 children in project schools could read 22 letters per minute with correct sound compared with 12 correct letters per minute in comparison group schools during endline. On nonword reading task, project group children correctly read an average of 14 nonwords per minute during endline as compared to the mean score of 12 correct nonwords per minute in comparison schools. Grade 2 children in project

group could correctly identify, on average, six out of 10 pairs of sentences on sentence choice task compared with an average score of four out of 10 pairs in comparison group. Average oral reading fluency score in project group schools during endline was 26 correct words per minute in contrast with the average score of 19 correct words per minute in comparison group schools. Furthermore, project school children, on average, correctly answered more than three out of five reading comprehension questions whereas average score on the same task in comparison group was less than three. Lastly, average score on sentence dictation task in project school during endline was slightly higher than the same in comparison group school as shown in Table 2.1

**Across tasks children in Literacy Program schools made two-year gains in their early grade literacy skills that were approximately 1.2 to 1.8 times greater than those seen in comparison group schools.** Differences in the two-year gains between project and comparison group children were statistically significant on letter sound ( $p < 0.001$ ), nonword reading ( $p < 0.001$ ), sentence choice ( $p < 0.001$ ), oral reading ( $p < 0.001$ ), and reading comprehension ( $p < 0.01$ ) tasks, respectively. However, the difference in the two-year gains on sentence dictation task between the two groups was not statistically significant. It indicated that gains made by children in writing skills in their early years of schooling across project and comparison groups were similar.

*Figure 2.1: Effect sizes across indicators*



Differences in the two-year gains between the two groups were investigated further by examining the *effect size* for each assessment task (see Figure 2.1). If an intervention includes several impact indicators measured in different units<sup>17</sup>, effect sizes help in comparing impacts of the intervention across indicators. Effect sizes also help in comparing impacts of an intervention with other interventions of similar nature. Here, the standardized mean effect size statistic was used for this purpose, with an effect size of 0.8 or higher considered as large. First, the mean difference in the two-year gains on each task between project and comparison group children was estimated using linear regression after controlling for various influencing factors or covariates. Then the estimated mean difference was divided by the estimated pooled standard deviation of the distribution of the difference in the two-year gains between project and comparison groups from the same linear regression to derive the effect size on each task (See *Appendix B: Data Analysis*). **The effect sizes for the Literacy Program intervention in Tanga region were positive but mostly small.** It was within the range of 0.1 to 0.4 for nonword reading, sentence

<sup>17</sup> For example, impact indicator letter sound fluency was measured in the units of correct letters per minute whereas impact indicator oral reading fluency was measured in the units of correct words per minute.

choice, oral reading, reading comprehension, and sentence dictation tasks, respectively. Only on letter sound task, the estimated effect size was large, around 0.9.

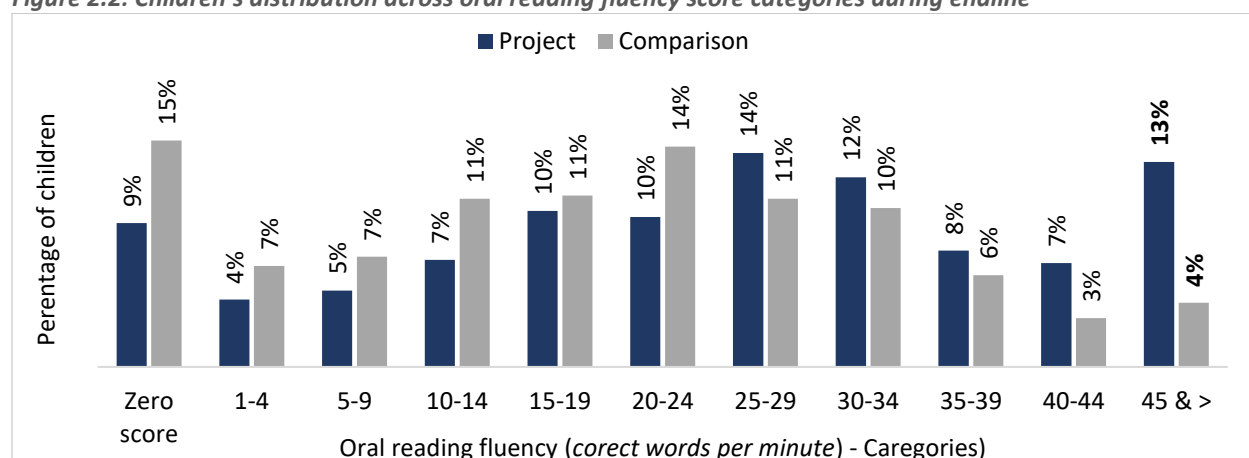
Overall, results of this study established that Room to Read’s Literacy Program in Tanga region had positive and statistically significant impacts on most of the early grade literacy skills. However, low mean scores for the children in project group during endline and low effect sizes across many tasks were noticeable. There exists room for further improvements across all early grade literacy skills.

## 2.2 Fluency and Comprehension Benchmarks

Independent studies in multiple language 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 (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 Tanzania and tracks children’s progress towards this fluency rate through a “fluent reader” benchmark that is set to at least 45 correct words per minute.<sup>18</sup>

Figure 2.2 shows how children were distributed across oral reading fluency score categories during 2022 end-of-Grade 2 endline in project and comparison groups, respectively.<sup>19</sup> During endline, only 13 percent children in project group and just 4 percent children in comparison group had met or exceeded the fluent reader benchmark of reading 45 or more words correctly per minute. The difference between project and comparison groups in children achieving the fluent reader benchmark was statistically significant ( $p < 0.001$ ).

**Figure 2.2: Children’s distribution across oral reading fluency score categories during endline**



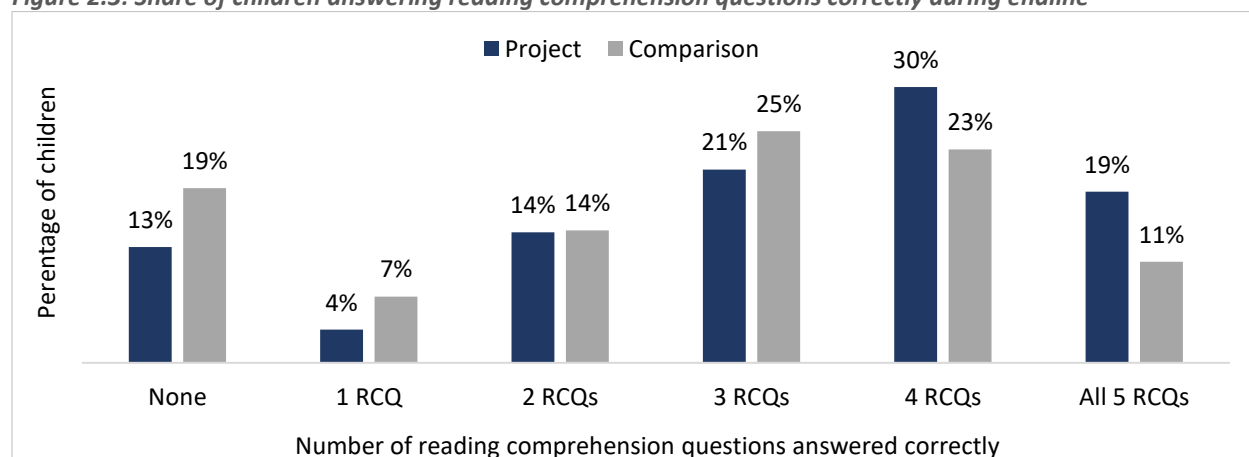
Other progress indicators include the proportion of children correctly answering at least four out of five

<sup>18</sup> Room to Read uses 45 or more correct words per minute as the oral reading fluency benchmark for all languages in which it implements the Literacy Program globally. Many countries in recent years have come up with their own language specific oral reading fluency benchmarks. Government of Tanzania uses a national benchmark of 50 correct word per minute for oral reading fluency by end of Grade 2. The recent national-level assessment done by the Ministry of Education, Science, and Technology showed only 18 percent Grade 2 children achieved the oral reading fluency benchmark of at least 50 correct words per minute. Moreover, during the national assessment 41 percent Grade 2 children could answer 80 percent or more comprehension questions correctly. Source: es Salaam, D. (2020). Standard Two National Assessment

<sup>19</sup> Oral reading fluency categories are presented in small intervals below 45 correct words per minute in Figure 2.2.

(i.e., at least 80 percent) of the questions on the reading comprehension task. Figure 2.3 shows that during 2022 end-of-Grade 2 endline, 49 percent children from project group correctly answered at least 80 percent of the reading comprehension questions. In comparison group school, a significantly ( $p < 0.05$ ) lower proportion, 34 percent, of children could do so.

**Figure 2.3: Share of children answering reading comprehension questions correctly during endline**



RCQ – Reading comprehension question

## 2.3 Zero Scores

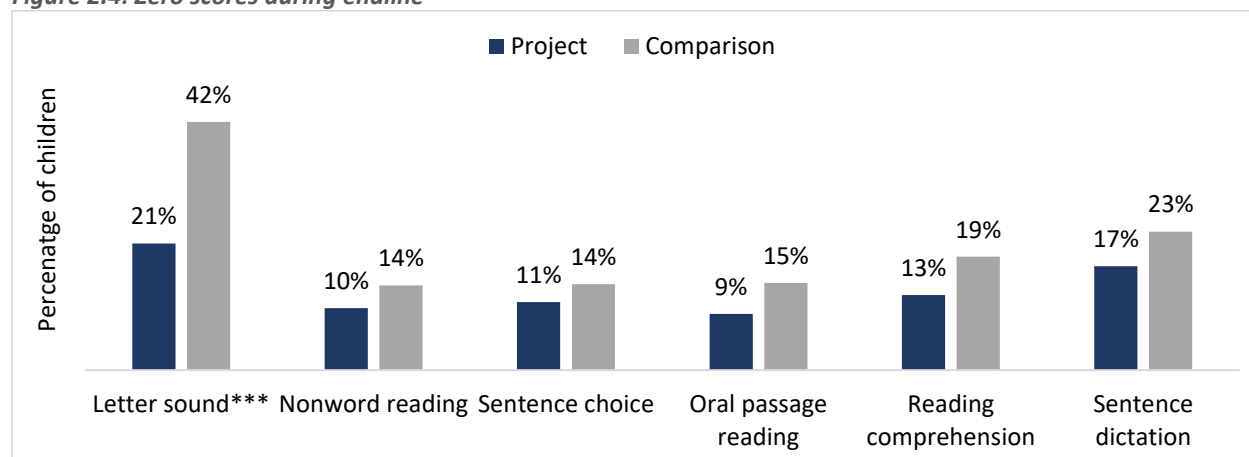
Analysis of zero scores provides a different perspective on how children performed during an assessment, with a focus on children who scored the lowest. In early grade literacy skills assessment, a zero score on a task includes instances in which a child fails to provide correct responses to any of the items attempted, as well as instances in which a child fails to answer or respond correctly to any of the first pre-determined number of items (also known as early-stop or discontinued task).<sup>20</sup> Zero scores on tasks reveal the subset of children who can be classified as non-readers.

Figure 2.4 presents percentage share of children with zero scores across project and comparison group schools on each task during 2022 end-of-Grade 2 endline. Data show that the prevalence of zero scores decreased sharply from 2021 baseline to 2022 endline across project and comparison groups. This was expected as 2021 baseline was conducted with children just starting their formal schooling in Grade 1 whereas the 2022 endline was performed with children after two years of formal schooling. What is noticeable was that after two years of intervention, substantial number of children remained as non-readers in project schools. For example, 21 percent project group children could not read a single Kiswahili letter with correct sound during 2022 endline. Also, in project schools, nine percent children could not read a single word correctly on oral passage reading task and 13 percent children could not

<sup>20</sup> For example, on letter sound task, if a child could not read any of the first 10 letters with correct sound, the task was discontinued, even if the stipulated 60 seconds total time for reading was not exhausted; in this case the child was given with zero score on letter sound task. On nonword reading task, the threshold for discontinuity was set at first five nonwords in the Kiswahili early grade literacy skills assessment tool used in this impact study. Similarly, the thresholds for discontinuity were first five sentences for sentence choice task and first five words in the story for oral passage reading task, respectively. Note that if oral passage reading task was discontinued, a child was not asked any reading comprehension question. Lastly, there was no early-stop threshold for the sentence dictation task. During administration, an assessor read the Kiswahili sentence three times with a specified interval between each reading while a child attempted to write the sentence on a paper. After an interval subsequent to the 3<sup>rd</sup> reading, the assessor ended the task and assigned a score to the written sentence based on pre-decided criteria.

answer a single question correctly on reading comprehension task, respectively, during 2022 endline. Furthermore, 17 percent children in project group scored zero on sentence dictation task during 2022 endline. Though the prevalence of zero scores in project group was lower than the same for comparison group during 2022 endline, the difference between the two groups across most of the tasks was not significantly different (See *Appendix B: Data Analysis*). Only on letter sound task, the gap between the two groups on zero scores was statistically significant ( $p < 0.001$ ); surprisingly, a substantial number (about 42 percent) of Grade 2 children in comparison group, scored zero on this task during 2022 endline.

**Figure 2.4: Zero scores during endline**



\*  $p < 0.05$     \*\*  $p < 0.01$     \*\*\*  $p < 0.001$

## 2.4 Comparisons by Gender

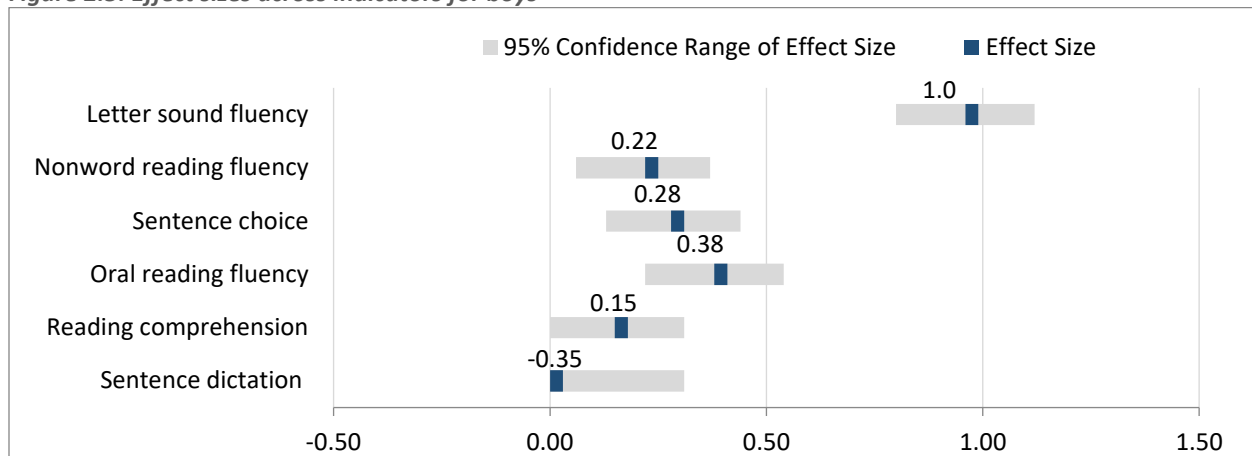
Room to Read is also interested in understanding the role that a child's gender plays alongside the intervention. Table 2.2 presents the results of this impact study broken down by gender. **Overall, Room to Read's Literacy Program in Tanga region benefited both boys and girls.** Girls in project group schools experienced two-year gains in early grade literacy skills that were 1.1 to 1.7 times higher than the gains experienced by girls in comparison group across tasks. Similarly, two-year gains experienced by project group boys were 1.1 to 1.9 times higher than the gains experienced by their counterparts in comparison group. For both girls and boys, differences in two-year gains between project and comparison groups were statistically significant on letter sound, nonword reading, sentence choice, oral passage reading, and reading comprehension tasks, respectively.<sup>21</sup> However, the same was not statistically significant on sentence dictation task for both girls and boys. Lastly, project group girls gained more in early grade literacy skills as compared to project group boys; the two-year gains for project girls on all six tasks were significantly higher than project boys.

Effect sizes were also estimated here to compare the impact of the Program on different early grade literacy skills separately for girls and boys. As shown in Figures 2.5 and 2.6, the estimated effect sizes were mostly positive but small across tasks for both girls and boys. Only on letter sound task, the estimated effect sizes were large for each gender; 1.0 for girls and 0.8 for boys.

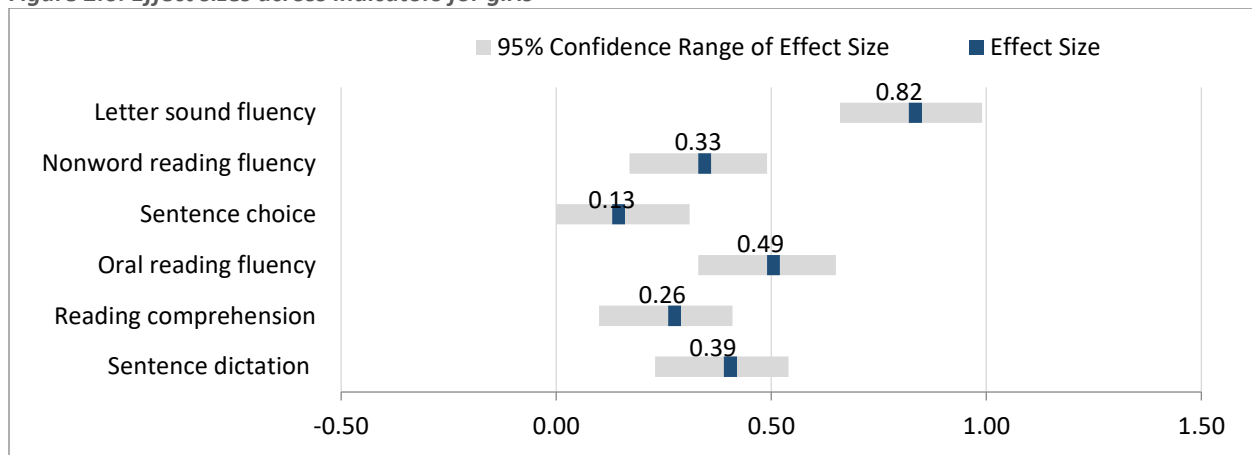
<sup>21</sup> For level of significance of the difference in two-year gains in literacy skills between, (i) project boys and comparison boys [D1], (ii) project girls and comparison girls [D2], and (iii) project girls and project boys [D3] for each task, see the statistics presented under the column heading "Diff. in the 2-year Gains b/w Groups" in Table 2.2.



**Figure 2.5: Effect sizes across indicators for boys**



**Figure 2.6: Effect sizes across indicators for girls**



**Table 2.2: Status of children's early grade literacy skills and the impact of the Program by gender**

Indicator (unit of measurement)	Group	Start-of-Grade 1 Baseline (2021)			End-of-Grade 2 Endline (2022)			2-year Gains	Diff. in the 2-year Gains b/w Groups <sup>^</sup>
		N	Mean	SD	N	Mean	SD		
Letter sound fluency (correct letters per minute)	Project boys	383	2.28	5.55	249	19.66	16.41	17.38	D1: 7.504***
	Comparison boys	383	1.07	3.14	250	10.44	17.18	9.37	D2: 9.274***
	Project girls	383	2.92	7.14	251	23.51	19.00	20.59	D3: 3.457**
	Comparison girls	390	1.44	4.17	250	13.35	15.63	11.92	
Nonword reading fluency (correct nonwords per minute)	Project boys	383	1.20	3.20	249	13.11	8.45	11.91	D1: 1.817**
	Comparison boys	383	0.73	2.43	250	10.23	7.88	9.50	D2: 2.916***
	Project girls	383	1.44	3.72	251	15.77	8.49	14.33	D3: 2.395***
	Comparison girls	390	1.28	2.96	250	12.77	7.74	11.49	
Sentence choice score (score on a 10-point scale)	Project boys	383	1.45	4.16	249	22.88	16.58	21.43	D1: 1.053***
	Comparison boys	383	0.83	3.26	250	16.45	13.76	15.62	D2: 0.633*
	Project girls	383	1.86	5.16	251	28.72	16.25	26.86	D3: 0.327
	Comparison girls	390	1.39	3.74	250	21.53	13.45	20.13	
Oral reading fluency (correct words per minute)	Project boys	383	0.23	0.72	249	5.37	3.78	5.13	D1: 4.686***
	Comparison boys	383	0.26	0.73	250	4.08	3.72	3.82	D2: 6.515***
	Project girls	383	0.30	0.94	251	5.74	3.54	5.44	D3: 5.441***
	Comparison girls	390	0.16	0.56	250	4.91	3.49	4.75	
Reading comprehension score (score on a 5-point scale)	Project boys	383	0.32	0.85	249	3.01	1.70	2.69	D1: 0.330*
	Comparison boys	383	0.17	0.63	250	2.44	1.74	2.27	D2: 0.286*
	Project girls	383	0.38	1.04	251	3.15	1.43	2.76	D3: 0.0522
	Comparison girls	390	0.31	0.80	250	2.75	1.50	2.45	
Sentence dictation score (score on a 13-point scale)	Project boys	383	4.14	1.84	249	5.24	3.81	1.10	D1: 0.0526
	Comparison boys	383	3.74	1.29	250	4.70	3.85	0.96	D2: 0.44
	Project girls	383	4.05	1.95	251	6.78	3.44	2.73	D3: 1.587***
	Comparison girls	390	3.99	1.63	250	6.20	3.62	2.21	

SD – Standard deviation \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

<sup>^</sup> Diff. in the 2-year Gains b/w Group i.e., the impact of the Program: Reports the estimated difference in gains from 2021 baseline to 2022 endline on each task score for the following three categories separately: (i) D1: between project boys and comparison boys, (ii) D2: between project girls and comparison girls, and (iii) D3: between project girls and project boys. An estimated difference was the coefficient of an independent variable – interaction between data collection wave (baseline or endline) and group (project or comparison) – from a regression, with random effects at school level, using a task score as dependent variable for each of the three categories. Each regression on a task score for a category also included the following covariates (i.e., controls): district, location, learner-teacher ratio, teacher training, school facilities score, age, speaks Kiswahili at home, missed school in last week, had all 3 mains meals in last 24 hours, both parents were present on most days at home in the past week, spent time with a caretaker in the past week at home other than parents, read without prompting/ for fun at least once in the past week, and household standard of living score.

### 3. Limitations

Alongside the results from the impact evaluation, Room to Read wants to highlight the limitations of the study. One such limitation was comparability across project and comparison groups, which was linked with the quasi-experimental design adopted for this impact evaluation. Validity of the impact 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 influence children's early grade literacy skills. As explained in *Section 1.4: Sampling*, every effort was made during sampling to ensure that a comparable set of schools was chosen. Moreover, data on observable school and children background characteristics were collected and analysed during both baseline and endline evaluation to check whether any statistically significant difference did exist. In *Section 1.4: Sampling*, it was discussed that the study did not find a statistically significant difference between project and comparison groups on many school and children background characteristics, either during baseline or endline or both. However, few school and children background indicators, as listed in *Section 1.4: Sampling*, revealed a statistically significant difference between project and comparison groups, either during baseline or endline or both. These indicators were used as controls in the regression analyses to estimate the impacts and effect sizes for the Program on each early grade literacy skill assessment task. Still, it may be possible that schools and children from project and comparison groups differed from each other in ways that the study did not explore or could not observe.

Study limitations that were linked with data collection and children sampling procedure are listed below.

- Data collection was done in the schools in the study. As a result, the study was unable to triangulate children's self-reported responses on background characteristics with those from informed adults, such as a parents or primary caregivers, who were not present in the schools during the data collection. To ensure quality, assessors were prepared through desk-based training and field practices to collect reliable data from early grade children during school-based data collection.
- During each wave of data collection, children sample was drawn from a list of children present in the target grade on the day visit to a school by assessors. Children sample in a school was not drawn from the list of all children enrolled in the target grade in a school. Authorities of each selected school were informed in advance about the time and purpose of the visit. However, possibilities for frequent and/or intentional absence on the day of visit by some children in a school could not be ruled out entirely.<sup>22</sup> Consequently, early grade literacy skills of the sampled children assessed in a school might not fully represent the current status of the same for the whole class.
- Access to the schools was restricted during the 2021 baseline evaluation. The data collection team could not physically verify the status of enrollment and attendance in Grade 1 in the selected schools prior to the visit for baseline data collection. To avoid possible shortfalls, more than required number of children was sampled in each school during the baseline. As a result, children sample sizes across 2021 baseline and 2022 endline were unbalanced.

Lastly, challenges associated with implementation of the Literacy Program in Tanga region should also be highlighted here. Onsite coaching and mentoring for the teachers teaching Kiswahili in early grades based on frequent observation of the language classes by Room to Read appointed field staff are key elements of the Literacy Program. Room to Read conducts these activities through the program

---

<sup>22</sup> For example, on average, 27 percent, and 21 percent of the enrolled children in Grade 2 were absent on the day visit in project and comparison group schools respectively during 2022 endline.

implementation monitoring processes. Multi-country analyses by Room to Read showed that these activities have positive and statistically significant influence on teachers' acquisition and mastery of classroom language instruction techniques that in turn impact children's learning.<sup>23</sup> Majority of the time during this two-year (2021-22) impact study, these activities were severely affected due to lack of regular, physical access to schools and classrooms because of Covid-19 related restrictions. Throughout this period, Room to Read had to rely mostly on remote support (e.g., regular telephonic calls) for the teachers.<sup>24</sup> Effectiveness of the remote support activities such as those implemented by Room to Read in contrast with onsite activities is not yet established. However, it does provide context for understanding and interpreting the findings of this two-year impact study.

---

<sup>23</sup> Joddar, P. (2019). 2<sup>nd</sup> Generation Multi-Country Project Implementation Monitoring System – What have we learned about implementation fidelity, program quality and reading outcomes? Presentation in the 63<sup>rd</sup> Annual Conference for Comparative and International Education Society (CIES)

<sup>24</sup> Room to Read also implemented many other program-specific activities (e.g., teacher training) remotely during this period.

## 4. Conclusion

**Results from the two-year (2021-22) study show that Room to Read's Kiswahili language Literacy Program in Tanga region of Tanzania had positive, statistically significant but modest impact on children's early grade literacy skills.** From the beginning of Grade 1 to the end of Grade 2, the gains made by project group children on most of the early grade literacy skills were significantly higher than the gains made by comparison group children. Moreover, a significantly higher proportion of Grade 2 children in project school achieved the oral reading fluency and reading comprehension benchmarks in contrast with comparison schools, respectively. Across all early grade literacy skills assessed in this study, the maximum impact of the Program was seen on the letter sound task as per the effect size statistics. The evaluation also found that the Program benefitted both girls and boys; however, girls seemed to have benefitted more from the intervention than boys.

**Despite the program's positive outcomes after two years of intervention, the study provided ample evidence that indicates scope for further improvements in the effectiveness of the Program.** Low average scores by Grade 2 children in project schools during the 2022 endline and small effect sizes on most of the literacy skills testify to this. Moreover, a noticeable number of children remained as non-readers in project schools even after two years of intervention as shown by the statistics on zero scores across assessment tasks. Also, the Program did not make any noticeable impact on children's early grade writing skills as per the results from the sentence dictation task.

Room to Read recognizes that the disruptions to the Program implementation due to Covid-19 might have played a crucial role in affecting its effectiveness. The shift from onsite delivery to remote delivery of the integral elements of the Program was sudden. However, the situation has improved a lot in recent times, and Room to Read is committed to continuing its effort to improve children's early grade literacy skills in the coming days in collaboration with the government in Tanzania. This commitment and buy-in from the government about the Literacy Program are evident from the recent expansions of the intervention in two new geographies, Ubungu Municipal Council in Dar es Salam region, and Mkuranga District Council in Pwani region, where the Program started in 2022 and 2023 school years, respectively.

## References

es Salaam, D. (2020). Standard Two National Assessment.

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.

Joddar, P. (2019). 2<sup>nd</sup> Generation Multi-Country Project Implementation Monitoring System – What have we learned about implementation fidelity, program quality and reading outcomes? Presentation in the 63<sup>rd</sup> Annual Conference for Comparative and International Education Society (CIES)

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.

Room to Read.

- Impact Evaluation of the Literacy Program in Nepal (2019). Available at: <https://www.roomtoread.org/media/42pnyd2o/2017-2019-impact-evaluation-of-the-lp-in-nepal-2019-endline-end-of-g2.pdf>
- India Literacy Program – 2017 Endline Evaluation Report. Available at: [https://www.roomtoread.org/media/gj1p5amz/in-2015-2017-reading-skills-endline-evaluation-report\\_all-states\\_final.pdf](https://www.roomtoread.org/media/gj1p5amz/in-2015-2017-reading-skills-endline-evaluation-report_all-states_final.pdf)

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](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/54857261295455628620/Impact\\_Evaluation\\_in\\_Practice.pdf](https://siteresources.worldbank.org/EXTHDOFFICE/Resources/54857261295455628620/Impact_Evaluation_in_Practice.pdf)

## **Annex A: Room to Read's School Selection Criteria – An Abridged Version**

### **General school selection criteria**

1. Each school is a government school or a government-supported/registered school
2. Schools are clustered or geographically close to other schools to enable regular and efficient support and monitoring.
3. Schools are easily and safely accessible year-round.
4. School infrastructure does not pose an imminent threat to students.
5. The school has a functioning administration and leadership committee to support project decisions in the school.
6. The principal/headmaster is supportive of core Room to Read program objectives and activities.
7. The principal/headmaster is in favour of a gender-friendly environment in the school.
8. Parents and community representatives are supportive of core Room to Read program objectives and activities.

### **Literacy Program school selection criteria (in addition to the general criteria)**

9. Schools have primary grades, at least Grades 1 to 5.
10. Enrolments are of an appropriate size to support quality, cost effective implementation.
11. Teachers are present and ready to teach class on time at least 80% of the time.
12. The school does not have an active or phased out Room to Read library.

### *Specific criteria for schools implementing the instruction component*

13. Literacy levels or existing reading assessments at the school or district are below the national/provincial average (based on available data).
14. Students entering school speak the language of instruction.
15. Classes are comprised of no more than one grade level.

### *Specific criteria for schools implementing the library component*

16. Access to appropriate books and other reading materials is limited.
17. A school library does not exist or does not function in a manner that supports a habit of reading.
18. Appropriate space is available to establish one of the Room to Read library variants (e.g., separate, classroom, shared or constructed).
19. The school has a dedicated librarian or is able to appoint a willing teacher as librarian, to coordinate library management at the school.

## Annex B: Data Analysis

Multiple types of data analyses were performed in this impact study. The first step of the analyses was to examine if a statistically significant difference between project and comparison groups on any school and children background characteristics existed. Difference between the two groups on each school background characteristics during baseline as well as endline was studied through the following statistical tests, such as (i) *t*-test for enrollment in target grade, attendance in target grade on the day of assessment, pupil to classroom ratio in target grade, pupil to teacher ratio in target grade, school facilities score, and classroom environment score, respectively, and (ii) *chi-square* test for location (rural or urban/semi-urban), monitoring/support visits from local (government) education officials in last 3 months, school having functioning school committee, school organized outreach event(s) in last 6 months, and at least one teacher in target grade received training on teaching language at the beginning of school year, respectively. For the children background characteristics, differences between the groups during baseline as well as endline were examined through (i) *linear regression* on each indicator in continuous variable format, such as, age, household size, and household standard of living score, respectively, and (ii) *logistic regression* on each indicator in categorical format, such as, attended pre-school, speaks Kiswahili at home, missed school on any day in the past week, had three main meals in last 24 hours, both parents present at home on most days in the past week, child spent time with anyone else other than parents at home in the past week, mother's literacy, anyone at home helped/prompted the child to study in the past week, anyone at home read to the child anytime in the past week, anyone at home told story to the child anytime in the past week, collection of children's books at home, availability of other types of printed items at home, and read at least once without prompting or for fun in the past week, respectively. Each regression equation included a children background characteristics as dependent variable and a dummy variable on school group (project or comparison) as the independent variable.

An impact of the Literacy Program in Tanga region would be evident if there was a greater and statistically significant gain on an early grade literacy skill between 2021 baseline and 2022 endline among project group children in contrast to children from comparison group. Here, linear regression analysis, with random effects at school level, was conducted on the score for each assessment task with dummy variables on (i) data collection wave (baseline or endline), (ii) school group (project or comparison) and (iii) interaction between data collection wave and school group as independent variables, respectively. Additionally, each linear regression included a set of controls for school and children background characteristics identified from the initial analyses on similarities or dissimilarities between the two groups; controls used in this impact study were: district, location, learner-teacher ratio, teacher training, school facilities score, age, gender, speaks Kiswahili at home, child missed school in last week, child had all 3 main meals in last 24 hours, both parents were present on most days at home in the past week, child spent time with a caretaker in the past week at home other than parents, child read without prompting/ for fun at least once in the past week, and, household standard of living score. Each estimated regression coefficient for the dummy variable on interaction between data collection wave and school group represented the impact of the Program on an early grade literacy skill. Value, sign (positive or negative), and level of statistical significance for each estimated coefficient were then reviewed to draw conclusions for study.<sup>25</sup> Lastly, a similar analytical procedure was used to assess the difference in two-year gains on each early grade literacy skill between (i) project group girls and comparison group girls, (ii) project group boys and comparison group boys, and (iii) project group girls and project group boys, respectively.

---

<sup>25</sup> See the statistics presented under column heading "Diff. in the 2-year Gains b/w Groups" in Tables 2.1 and 2.2.



Analysis was also done to check the statistical significance of the difference between project and comparison groups on zero score for each assessment task during 2022 end-of-Grade 2 endline. Logistic regression analysis, with random effects at school level, was conducted for this purpose with (i) child scoring zero (yes or no) on task as the dependent variable, and (ii) a dummy variable of school group as the independent variable, respectively. Each logistic regression on zero scores also included the set of controls described above.

Similar logistic regression analyses were conducted on the endline data to assess if significant differences existed between the two school groups on children achieving (i) the fluent reader benchmark i.e., correctly reading 45 or more words per minute on oral passage reading task, and (ii) 80% or more score on reading comprehension task, respectively.

Effect sizes were also estimated to compare the extent of influence of the Literacy Program in Tanga region across six early grade literacy skill assessment tasks. For each task, the effect size was computed as the mean two-year difference in gains between project and comparison groups divided by the pooled standard deviation of the distribution of two-year differences in scores across groups. 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 children in comparison group.

## Annex C: School and Children Background Characteristics

### School background characteristics during 2022 endline

Characteristics	Project		Comparison	
	N or mean	% or SD	N or mean	% or SD
Schools	50	-	50	-
District				
Muheza	30	60%	30	60%
Tanga	20	40%	20	40%
<b>Location***</b>				
Rural	27	54%	43	86%
Urban or semi-urban	23	46%	7	14%
Enrollment in the target grade	73.9	29.8	72.5	59.6
Attendance in the target grade on the day of assessment	72.8	17.2	79.1	17.9
Learner to classroom ratio in the target grade	73.9	29.8	63.7	45.2
<b>Learner to teacher ratio in the target grade***</b>	71.0	31.7	51.1	25.0
Monitoring / support visits from local education officials (government) in the last 3 months	42	84%	40	80%
Has a functioning school committee	49	98%	50	100%
Organized outreach event(s) in the last 6 months	12	24%	10	20%
<b>At least one teacher received training on teaching language at the beginning of the school year***</b>	38	76%	11	22%
<b>School facilities score (on a 10-point scale) ***</b>	6.0	1.5	5.0	1.5
Classroom environment score (on a 7-point scale)	2.0	0.4	2.2	1.0

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

### Children background characteristics during 2021 baseline and 2022 endline

Characteristics	Project		Comparison	
	N	% or Mean (SD)	N	% or Mean (SD)
Learner sample size				
Start-of-Grade 1 Baseline (2021)	766	-	773	-
Boys	383	50%	383	50%
Girls	383	50%	390	50%
End-of-Grade 2 Endline (2022)	500	-	500	-
Boys	249	50%	250	50%
Girls	251	50%	250	50%
Age (in years)				
Baseline	766	6.8 (1.0)	773	6.9 (0.9)
Endline	500	8.3 (1.0)	500	8.5 (1.1)
Attended preschool				
Baseline	723	94%	713	92%
Endline	457	91%	459	92%
<b>Speaks Kiswahili at home</b>				
Baseline***	693	90%	665	86%
Endline*	450	90%	428	86%
<b>Missed school on any day in the past week</b>				
Baseline	174	23%	186	24%
Endline*	165	33%	195	39%
<b>Had main three meals in last 24 hrs.</b>				

Characteristics	Project		Comparison	
	N	% or Mean (SD)	N	% or Mean (SD)
Baseline	546	71%	568	73%
Endline**	398	80%	353	71%
Household size				
Baseline	764	5.5 (1.9)	771	5.6 (1.8)
Endline	500	4.7 (1.8)	500	4.8 (1.8)
<b>Both parents present in the household on most days in the past week</b>				
Baseline	341	45%	335	43%
Endline**	134	27%	175	35%
<b>Spent time with anyone else (e.g., caretaker) other than parent(s) at home in the past week</b>				
Baseline*	523	68%	569	74%
Endline***	407	81%	362	72%
Mother literate				
Baseline	209	27%	226	29%
Endline	145	29%	154	31%
Anyone at home helped/prompted the child to study in the past week				
Baseline	460	60%	470	61%
Endline	358	72%	348	70%
Anyone at home read to the child anytime in the past week				
Baseline	408	53%	420	54%
Endline	301	60%	276	55%
Anyone at home told story to the child anytime in the past week				
Baseline	353	46%	389	50%
Endline	268	54%	237	47%
Children's books at home				
Baseline	375	49%	410	53%
Endline	238	48%	267	53%
Any other type of printed items at home				
Baseline	295	39%	305	39%
Endline	282	56%	285	57%
<b>Read at least once without prompting or for fun in the past week</b>				
Baseline**	508	66%	571	74%
Endline	412	82%	392	78%
<b>Household standard of living score (on a 9-point scale)</b>				
Baseline***	766	4.7 (2.2)	773	4.3 (2.3)
Endline	500	4.9 (2.0)	500	4.7 (2.0)
*** $p < 0.001$ , ** $p < 0.01$ , * $p < 0.05$				

## Annex D: Correlation between Assessment Scores during 2022 Endline

Task	Letter sound	Nonword reading	Sentence choice	Oral passage reading	Reading comprehension	Sentence dictation
<b>Project schools</b>						
Letter sound	1					
Nonword reading	0.6057***	1				
Sentence choice	0.4467***	0.6587***	1			
Oral passage reading	0.5730***	0.8581***	0.6932***	1		
Reading comprehension	0.4229***	0.6370***	0.6567***	0.6240***	1	
Sentence dictation	0.4698***	0.5755***	0.5448***	0.5899***	0.5089***	1
<b>Comparison schools</b>						
Letter sound	1					
Nonword reading	0.4058***	1				
Sentence choice	0.3591***	0.6972***	1			
Oral passage reading	0.4040***	0.8844***	0.7475***	1		
Reading comprehension	0.3032***	0.6624***	0.6305***	0.7124***	1	
Sentence dictation	0.4027***	0.6155***	0.5461***	0.6507***	0.5982***	1
<b>All schools</b>						
Letter sound	1					
Nonword reading	0.5345***	1				
Sentence choice	0.4245***	0.6845***	1			
Oral passage reading	0.5278***	0.8725***	0.7219***	1		
Reading comprehension	0.3869***	0.6574***	0.6510***	0.6704***	1	
Sentence dictation	0.4393***	0.5964***	0.5490***	0.6144***	0.5584***	1
*** $p < 0.001$ , ** $p < 0.01$ , * $p < 0.05$						