Improving reading proficiency in early childhood education classrooms: Evidence from Liberia

Michael Sebele[a], Jeffery McManus[a], Mtise Mwanza[a], Leah Mwai[b] and Mico Rudasingwa[b]*

[a]*IDinsight, Lusaka, Zambia;* [b]*IDinsight, Dakar, Senegal*

Corresponding author: Mico Rudasingwa, mico.rudasingwa@idinsight.org, IDinsight, Dakar, Senegal

Improving reading proficiency in early childhood education classrooms: Evidence from Liberia

Fewer than half of Liberian adults are literate. One policy approach to improving literacy rates focuses on early childhood education (ECE), since ECE gives children an opportunity to gain foundational literacy skills and prepare for enrolment in primary school. However, under the current curriculum in Liberia, few children are exiting ECE with these skills. The FasterReading (FR) program, designed and implemented by Rising Academy Network (RAN), is a phonics-based approach to reading instruction. We conducted a randomized controlled trial of this program in 74 government schools across 10 Liberian counties from January to July 2022. We find that ECE students exposed to the full FR program improved 0.28 SD in reading proficiency compared to students who received none of the program, although our impact estimates are imprecise due to treatment non-compliance. Given the standardized implementation protocol and low variable costs of the program, we project that the program may be at least as cost-effective and scalable as other primary and pre-primary literacy programs.

Keywords: Early childhood education (ECE); learning outcomes; FasterReading (FR); reading proficiency, randomised controlled trial (RCT)

Funding: Funding for this research was provided by King Philanthropies and the Fund for International Development.

Disclosure statement: The authors report there are no competing interests to declare.

1. Introduction

The education sector in Liberia has been characterized by the lingering effects of the Liberian civil war. The civil war had a detrimental effect on school infrastructure, equipment, management systems, and the teaching force. While estimates of damage and long-term effects are imprecise, UNESCO reported that 30 percent of public schools and 24 percent of community schools were completely destroyed, and 16 percent suffered serious damage or looting of their furniture and teaching supplies. A World Bank report places the toll higher, with nearly 75% of schools damaged or destroyed. Many teachers fled because of the violence leading to a dearth of qualified and trained teachers. School participation also fell with "a generation of children" missing the opportunity to go to school (UNESCO, 2011).

More recently, the education sector was also affected by the 2014 Ebola virus disease (EVD) outbreak and the COVID-19 pandemic. Liberia was one of the three epicentre countries of the 2014 EVD outbreak and to prevent transmissions schools were closed for seven months. Similar school closures were put in place to prevent the spread of COVID-19. Ongoing research is estimating the learning loss and effects on child well-being that these school closures had. However, the Ministry of Education in Liberia hypothesizes, given learnings from the EVD outbreak and its associated school closures, that the school closures due to COVID-19 could have far reaching effects including the interruption of learning, decrease in the number of available teachers, reduced access to clean water, and impacts on nutritional well-being due to reduced access to school feeding programs (MoE-Liberia, 2020).

In the wake of these shocks, learning outcomes in Liberia remain low. Liberia has one of the lowest adult literacy rates in the world, with half of adult men and a third of adult women able to read (World Bank, 2017). The adult literacy rate in Liberia across both sexes is less than half of all adults. This is below the regional average of two thirds and the world average of four out of five adults (USAID, 2017). While Liberia's literacy rates are trending upwards, children are still falling far short of expectations for their cohorts. A literacy assessment conducted in 2014 found that Grade 3 students in Liberia could read 20 correct words per minute, much lower than the Ministry of Education (MoE) benchmark of 45 correct words per minute (MoE-Liberia, 2016).

One of the policies that has been implemented in Liberia to address low learning outcomes in public schools is the Liberia Education Advancement Program (LEAP). LEAP is multi-partner public-private partnership that was designed to build capacity within Liberia's education system and improve student learning outcomes (MoE-Liberia, 2022). Rising Academy Network (RAN) is one of the private organizations that supports the management of LEAP schools in Liberia. RAN supports 95 government schools in Liberia as part of LEAP. A recent randomized controlled trial of LEAP found that after three years students in LEAP schools performed 0.18 and 0.21 standard deviations better in English and Math tests, respectively. When exploring private partner heterogeneity, RAN-supported schools performed 0.42 and 0.55 standard deviations better on English and Math tests than students in the control group (Romero & Sandefur, 2022). However, provider-level subgroup analysis was likely imprecise because the sample size of schools for any given provider was small.

Following the LEAP RCT, RAN has expanded in Liberia and continues to refine their curriculum. In 2021, RAN developed an accelerated, phonics-based reading program, FasterReading (FR), to support ECE and primary-grade students in foundational reading. Although students in RAN-supported schools have better reading outcomes than peers in other government schools, students in RAN schools still lag behind national and international benchmarks for reading proficiency. An EGRA assessment conducted in 2018 at RAN schools found that Grade 3 students could read 27 correct words per minute on average, which is lower than the Ministry of Education benchmark of 45 correct words per minute. RAN first piloted the FR program in the 2021-22 school year.

To evaluate the impact of the FR pilot on reading proficiency, we conducted a randomized controlled trial (RCT) in 74 RAN-supported government schools across 10 Liberian counties. RAN rolled out FR in ECE classrooms for 37 treatment schools from January to July 2022 while ECE classrooms in 37 other schools received regular ECE programming. Baseline assessments were administered to 2,712 students in December 2021, and follow-up assessments were administered to 2,307 of these students (85.1%) in July 2022.

An important feature of our evaluation is that we focus specifically on children in ECE classrooms who were 7 years or older (consisting of about 40% of all children in RAN's ECE classrooms), who were considered to be 'over-age' relative to the MoE's age expectations for children in ECE classes. We focus on this population for two reasons. First, only ECE classrooms were included in the random assignment; the program was

offered to students in Grades 3 to 6 in all RAN schools.[1] Thus we cannot causally identify the impact of the FR program outside of ECE classrooms. Second, the evaluation focuses on overage ECE students rather than all ECE students due to resource constraints (we could not collect data on all ECE students) and because overage ECE students are of particular policy importance to RAN and to the MoE. According to the Ministry of Education in Liberia, in 2020 roughly two thirds of students in ECE classrooms in government schools are overage. Overage ECE students are susceptible to early dropout as they pursue income-generating activities or become discouraged being in classrooms with much younger students (World Bank, 2016). Large age ranges in ECE classrooms also present pedagogical challenges for teachers.

One consequence of this design is that there was some confusion by teachers and principals about the treatment status of schools, leading to non-compliance: some control schools implemented the FR program in ECE classrooms, while some treatment schools did not implement the full FR program in ECE classrooms. In our analysis we present both intent-to-treat (ITT) and treatment-on-the-treated (TOT) estimates to address non-compliance. To estimate TOT, we define a treatment intensity (TI) estimator based on the percent of weeks that a school received of the program, which we instrument using assigned treatment. We also present results based on alternative TOT definitions as robustness checks.

Despite evaluation and implementation challenges, our results suggest that the FR program may be a cost-effective approach to improve reading proficiency. While our impact estimates are imprecise due to non-compliance, we find that an over-age ECE student exposed to the full FR program improved 0.28 SD in reading levels relative to a student who received none of the FR program ($p = 0.10$), or a 36% improvement over the status quo. We also find that the FR program increased school attendance by 11.4 percentage points ($p = 0.03$), but, curiously, reduced the likelihood that a student reported reading at home by 12.1 percentage points ($p = 0.05$). Given the standardized implementation protocol and low variable costs of the program, we project that the program may be at least as cost-effective and scalable as other TaRL-based programs.

This paper adds to literature on targeting instruction to the child's learning level in Africa. In Ghana, an evaluation of several targeted instruction programs on learning outcomes showed that teacher-led targeted instruction improved numeracy and literary outcomes for primary school students in Ghana's public schools (Duflo et al, 2021). A study in Kenya also found positive effects of grouping students together by ability (Banerjee et al., 2016).

Evidence on providing ECE programming has shown consistently positive and long-term impacts on education and economic outcomes. Along with canonical studies of programs such as the Perry Preschool Program (Heckman et al, 2010) and the

[1] The FasterReading program was designed as a catch-up program for students who had not yet developed foundational reading skills. The program was offered to ECE classes to ensure that overage ECE students developed these foundational skills and could progress to the grades that were expected for their age. Grades 1 and 2 classes were excluded because the FR curriculum overlapped with their core curriculum. The program was offered to Grades 3 to 6 classrooms so that students who had not mastered these foundational reading skills could master them so as to further progress in their understanding of English and their reading skills.

Abecedarian Project (Campbell et al, 2002), more recent evidence from sub-Saharan Africa has found similarly large effects from ECE programs on short-term and medium-term learning outcomes (Martinez et al., 2017; Bietenbeck et al., 2019; Krafft, 2015; Woldehanna and Araya, 2017). However, there has been relatively little evidence on the impact of ECE reforms in West Africa, or of TaRL methodologies applied to ECE classrooms. This paper aims to contribute to this evidence base, as well as inform the design and scale-up of new ECE curriculum in RAN's schools and other government and private schools in the region.

The rest of the paper proceeds as follows: we discuss the evaluation design in more detail and outline the sampling strategy, data sources, and analytical framework. We then present our results followed by a discussion of the results and our recommendations.

2. Program Design

The FR reading program is a phonics-based accelerated reading program that uses a level-based approach, instead of a grade-based approach as commonly seen in schools. The program has multiple components, including teacher training sessions, teacher guides and student workbooks, regular targeted assessments, and a mobile app to support teachers in their understanding of the program and how to teach phonics. The FR curriculum consists of foundational literacy competencies, from letter recognition to the ability to read short stories.

FR was intended as a supplement to regular classroom instruction. The program ran alongside regular ECE programming and during the school day. Implementing the program during the school day was done so that teachers could use FasterReading as part of their normal teaching routine and duties. During the 2021-22 school year, RAN trained all teachers in its 95 LEAP schools in the new FR curriculum.[2] Teacher training consisted of a one-time three-day training that involved discussing the foundations of early literacy, reviewing the student diagnostic assessment, and reviewing the content of the program. In addition to this training, RAN's school performance managers (SPMs) held regular visits to schools, monitored FR instruction and provided feedback to teachers to reinforce the FR curriculum.

After teacher training, schools were instructed to implement the program in all Grade 3, 4, 5, and 6 classes; treatment schools were instructed to include ECE classes. Students were assessed and assigned to groups with other children in the same reading level. The program was then rolled out for 20 weeks consisting of five four-week cycles, where after each cycle students were re-assessed and re-grouped to the appropriate level.

The intervention's theory of change hypothesizes that the various components of the program such as the initial teacher training, SPM coaching and monitoring of teacher performance, continuous assessments of students, and teaching to student ability would lead to teachers providing high quality instruction targeted to each student's level. Subsequently, these outcomes would lead to improvements in student reading and reading comprehension at the end of the program. This theory of change relied on several underlying assumptions such as materials being easily accessible, teacher's capacity

[2] One school dropped out of RAN's LEAP network in 2022, so that RAN was managing 94 schools by the end of the school year.

improving after receiving training from SPMs, and that SPMs and teachers understood how to deliver a phonics-based reading curriculum.

3. Evaluation Design

a. Sampling and random assignment

The population of interest for this evaluation was all over-age ECE students in RAN-supported schools under LEAP. Out of 95 schools supported by RAN in Liberia, RAN identified 75 that were eligible for the FR program. For this RCT, to meet evaluation requirements, eligible study schools had to have both an ECE section and more than 3 teachers in the whole school to support program implementation.

Students in these 75 schools were initially assessed in May 2021, and we randomly assigned schools to either administer the FR curriculum to ECE classrooms (38 schools) or to provide the standard reading curriculum to ECE classrooms (37 schools) Randomization was stratified on three observables – (i) school-average baseline FR level of overage ECE students; (ii) overage ECE student-teacher ratio; and (iii) year that RAN took over the management of the school. These stratification variables were combined into an index using principal components analysis. Eighteen strata were created (15 strata of 4 schools and 3 strata of 5 schools). Within each stratum, two schools were randomly assigned to treatment and two to control; for the leftover 3 schools in the strata with 5 schools, 2 were randomly assigned to treatment and the remainder to control.

RAN started implementing the FR program in ECE classrooms in treatment schools in June 2021. However, after three weeks of FR implementation, ECE classes were abruptly closed by the Ministry of Education due to fears about rising COVID-19 cases, and the FR program was suspended.

Schools reopened in November 2021, and a new cohort of over-age ECE students was assessed in November/December 2021. For this new baseline, enumerators assessed 2,712 over-age ECE students across 74 schools who had not participated in the previous baseline assessment or suspended implementation of FR. One treatment school dropped from the LEAP and the RAN network and was thus dropped from the study, leaving 37 treatment and 37 control schools. After confirming that baseline characteristics were still balanced across treatment and control schools, we retained the original random assignment rather than re-randomize schools. This reduced confusion among schools about their treatment status, though as described below, did not eliminate confusion about which students to include in the FR program.

According to our pre-analysis plan (PAP), registered on the American Economic Association registry for RCTs,[3] we planned on randomly sampling approximately 1,370 students for endline assessments. With 74 schools and approximately 1,370 students in our sample, our study was powered to detect moderate effect sizes of approximately 0.23 standard deviations (SD). However, during endline piloting, RAN staff found that it was easier to instruct enumerators to follow up with all students who were assessed at baseline, rather than to sample a subset of those. Thus, our endline sample was larger than anticipated though, as noted below, suffered from some attrition.

[3] RCT ID: AEARCTR-0007954, https://www.socialscienceregistry.org/trials/7954

A sampling and randomization flowchart is depicted in Figure 1.

[Figure 1 near here]

a. Data

Data was collected through enumerator-administered in-person assessments using tablets and phones.[4] The reading assessments used at baseline and endline bear similarity to other dipstick assessments of foundational literacy, such as the Annual Status of Education Report (ASER) assessment tool. Students were tested on letter recognition, word identification, and passage reading. An example of the assessment tool is included in the appendix. Similar assessments were used to group students in treatment schools into FR reading levels, though the questions varied in the versions used for grouping versus those used for the evaluation.

At endline we added a short module on basic numeracy to the student assessment. The numeracy assessment included four questions. Three questions were number identification questions and one question was a word problem that asked the child to perform basic addition. We also asked students to self-report their reading behaviour and perceptions of school and reading. We use this data to analyse if there was any crowding out of numeracy skills, reading practice, or changes in attitudes towards school and reading because of the program.

We also leverage attendance data collected by RAN staff. Over the course of the program, School Performance Managers (SPM) and School Leaders (SL) collected attendance data, but there was no consistency in how many attendance checks were conducted per school. SPMs conducted up to five attendance checks per school, and School Leaders conducted up to an additional five checks. The median student in our dataset has 4 attendance data points; 21% of students have no data points, and 3% of students have 10 attendance data points. In our analysis, we present results that average across all attendance checks for each student (dropping students who have no attendance data points). We also present results that only average across attendance data points collected by SPMs, due to RAN's concerns about attendance data reported by School Leaders.

Finally, we were provided with the results of a process evaluation of the FR program conducted by RAN. This data provided qualitative insights on the experiences of teachers and challenges encountered during implementation, which we use in the interpretation of our results in this evaluation.

b. Analytical model

In line with our pre-registered PAP, the effect of FR on student performance was estimated using the following Ordinary Least Squares model:

[4] Ethics approval was granted by the University of Liberia Pacific Institute for Research and Evaluation Institutional Review Board (protocol #21-11-291). We acquired school-level consent from deputy/head teachers at all participating schools. Children provided assent before being assessed.

$$Y_{ij}^{*} = \beta_1 T_j + X_{ij}'\gamma + \alpha_s' + \epsilon_{ij}$$

Y_{ij} denotes the outcome variable of interest (reading level, other outcomes) for student i in school j. T_j represents the treatment status of school j. $X_{ij}$is a vector of student-level covariates, including ECE baseline reading score, a binary indicator for student gender, and student age at baseline.[5] We use an analysis of covariance (ANCOVA) estimator, rather than a fixed-effects or first-differences model, to maximize statistical power (McKenzie, 2012). α_s is a vector of categorical factors corresponding to the stratum that the student is found in. ϵ_{ij} denotes the student error term i clustered at the school-level j.

c. Non-compliance

In the Results section, we report intent-to-treat (ITT) estimates, comparing students who were assigned to treatment schools to those assigned to control schools. However, we also report treatment-on-the-treated (TOT) estimates. Over the course of program implementation, it was discovered that schools deviated from their treatment assignment. Some schools did not implement FR in ECE classrooms or only implemented a subset of the five 4-week FR cycles. Some control schools implemented some FR cycles in ECE classrooms as well. Table 1 summarizes the extent of non-compliance across treatment and control.

[Table 1 near here]

Our preferred approach to address non-compliance is to measure treatment as treatment intensity. We define treatment intensity as the percent of the program that was received by a school. For instance, if a school implemented 3 FR cycles, then we code it as receiving 60% of the program. We then conduct an instrumental variables regression, where treatment intensity is instrumented with treatment assignment.

Our treatment intensity estimator relies on two assumptions: (i) We assume that treatment effects scale linearly with the number of weeks of implementation; (ii) the number of weeks of implementation in control schools is uncorrelated with the size of the potential treatment effect. For this evaluation, we believe these to be reasonable assumptions. Students likely benefited from more weeks of instruction in the FR program. Qualitative evidence from RAN's process evaluation suggests that most control schools implemented cycles of the FR program in ECE classrooms because of confusion surrounding treatment and control group treatment assignment (namely, that both treatment and control schools were instructed to implement FR in non-ECE classrooms, but that only treatment schools were instructed to implement FR in ECE classrooms).

In addition to our treatment intensity estimator, we report alternative TOT estimates in the appendix as robustness checks. Since some schools implemented only part of the FR program, it is unclear whether to treat those schools as 'treated' or 'not treated'. To address this, we estimated bounds on the TOT effects. The upper bound considers a student treated if they received all 5 cycles of the FR program, otherwise, they

[5] We also include adjustments for missing baseline data. When a student-level covariate was missing we replaced the missing value with 0 and included a binary indicator for whether the student covariate was missing at baseline

are not treated. The lower bound considers a student treated if they received any cycle of the FR program.

The TOT bounds rely on fewer assumptions than the treatment intensity estimator, but they provide much less precise impact estimates. For this reason, and since we think that the assumptions underlying the treatment intensity estimator are reasonable, we consider the TOT bounds a robustness check and report them in the appendix.

In summary, we report ITT estimates, which align with our pre-specified approach in our pre-analysis plan, alongside treatment intensity estimates, which rely on stronger assumptions but we believe are most indicative of the effectiveness of the full five-cycle FR program.

d. Attrition

We observed some attrition at endline with our assessors able to re-evaluate 85% (2,307 students) of our baseline sample at endline. Attrition was slightly higher in the control group (18%) than in the treatment group (12%) ($p = 0.06$). However, the endline sample remains balanced even with differential attrition. We therefore report estimates for the endline sample without adjusting for attrition in the Results section. As robustness checks, we present the results for the main outcomes with inverse-probability weights (IPWs), and Lee Bounds on treatment effects in the appendix. We find that these do not meaningfully change our findings.

4. Results

a. Reading proficiency at baseline

Average reading proficiency at baseline was low, but with some variance within classrooms. About half of students in the sample were able to recognize letters, half of students were not able to recognize letters, and virtually no students were able to decode simple CVC words. 92% of variance in baseline reading levels occurred within classrooms, with 8% of variance explained by differences in average reading levels across classrooms. This variance in baseline ability within classrooms was important for the FR theory of change, since treatment students were placed into groups with other students at a similar reading level, and a customized curriculum was delivered to each level, similar to the TaRL approach. In all treatment schools, two groups were initially created: A Letter-level group and a CVC-level group.

b. Balance

Student-level baseline characteristics and school-level baseline characteristics were well-balanced across treatment and control groups in our initial sample, as well as in our endline sample post-attrition, as shown in Table 2.

[Table 2 near here]

c. Main results

Table 3 presents results for all of our pre-specified outcomes. For reading levels, our ITT estimates show a modest and statistically insignificant improvement in reading proficiency of 0.07 reading levels or 0.17 standard deviations ($p = 0.14$). Taking into

account non-compliance using our treatment-intensity (TI) estimator, we find that a student exposed to the full FR program would gain 0.12 more reading levels or +0.28 SD compared to a student who received none of the program (p = 0.10). Since the average student in the control group gained only 0.33 reading levels, this modest effect represents a 36% increase over the status quo.

[Table 3 near here]

In addition to growth *across* reading levels, the FR program may have led to modest improvements in reading proficiency *within* levels. Enumerators recorded the number of mistakes made by students in each activity to determine whether a student passed or failed a reading level. Using this mistakes data, we estimate that a student exposed to the full FR program would correctly recognize 0.49 more letters than a student who received none of the FR program, though the estimate is not statistically significant (p = 0.19). Similarly, we estimate that a student exposed to the full FR program would correctly pronounce 0.16 more beginning sounds than a student who receives none of the FR program, though the estimate is not statistically significant (p = 0.33).[6] We report this in **Table A5** in the appendix.

The FR program did not have any statistically significant impacts on basic math proficiency, indicating that the FR program did not crowd out numeracy instruction. On the other hand, students exposed to the full FR program were 8.1 percentage points or 10% more likely to be present during SPM or SL attendance checks than students who received none of the FR program (p = 0.07). Focusing only on attendance checks conducted by SPMs, due to concerns about data quality from SL attendance checks, we find that students exposed to the full FR program were 11.4 percentage points or 15% likely to be present during SPM or SL attendance checks than students who received none of the FR program (p = 0.02).

There were no differences between treatment and control students in terms of perceptions about schooling or reading. Curiously, students assigned to treatment schools were 7.3 percentage points less likely to report reading at home than students assigned to the control group (p = 0.05). This difference grows to 12.1 percentage points when we compare students exposed to the full FR program versus students exposed to none of the FR program.

We find few differences in treatment effects across subgroups. The program had similar impacts on reading proficiency, attendance rates, and self-reported reading at home for boys, girls, students starting at Letter level, and students starting at CVC level

5. Discussion

The results from our RCT show that the FR program likely had a modest effect on reading proficiency. Considering non-compliance in treatment and control schools, we estimate that a student exposed to the full FR program gained 0.28 SD in reading scores compared to a student who received none of the FR program (p = 0.10), corresponding to a 36%

[6] We exclude CVC level analysis since mistakes in this section of the assessment were not recorded in a consistent format by enumerators.

increase over the status quo. The FR program also had a positive impact on school attendance, but it decreased the likelihood of students reporting that they practiced reading at home.

a. Cost Effectiveness

Cost-effectiveness analysis (CEA) is typically reserved for interventions with strong evidence of impact. In J-PAL's list of interventions that improve student learning (JPAL, 2020), for instance, the authors only report cost-effectiveness for interventions that have statistically significant treatment effects at the 10% level of significance. Our findings using our preferred treatment intensity (TI) estimator meet this burden of proof. We use these TI estimates – our best guess at the impact of the full program – to give suggestive evidence on the potential cost-effectiveness of FR, if the program were implemented according to plan.

We focus our cost-effectiveness analysis on over-age ECE students. This is a conservative decision since some of the costs are incurred at the school-level, and thus would not scale proportionally if FR were implemented in upper grades (Grades 3-6, as envisioned). However, we believe that our narrower focus is appropriate for two reasons. First, from our evaluation we only have impact estimates for over-age ECE students and not for other grades – any extrapolation of impact from over-age ECE to others would require strong assumptions. Second, not all schools implemented FR to Grades 3-6 due to capacity constraints (lack of teachers and/or lack of training), so it does not appear that RAN could implement FR to other students and grades without additional investment in teacher training or hiring.

We separate out the fixed start-up costs of the FR program (e.g., curriculum development) from the costs that would be incurred in future implementation (trainings, materials, etc.), and only include the latter in our CE estimate. We exclude any costs that would be incurred in the course of normal school programming, including teacher and staff salaries. We also exclude costs associated with COVID-induced lockdowns and restarting the program, including refresher trainings. A detailed breakdown of our cost effectiveness analysis can be found in the appendix.

We find that the FR program had a cost effectiveness of 0.91 SD learning gains per $100. This cost-effectiveness estimate is comparable to other pedagogical innovations highlighted in J-PAL's CEA of education programs. Computer-assisted learning in India, remedial education in India, and contract teachers with tracking in Kenya all have cost-effectiveness estimates around 1 SD per $100.

Our cost-effectiveness estimate of FR for over-age ECE students comes with several caveat, which may be leading us to overestimate or underestimate the true cost-effectiveness of the program. First, as noted above, extending the full program to Grades 3-6 may not require a proportional increase in costs. Second, the 2021-22 school year was the first year that RAN piloted the FR program; successive years may see program implementation improve, leading to further impact. Third, program impact was not precisely estimated due to non-compliance; the true impact may be higher or lower. Finally, the control group, against which treatment effects are standardised, had low average reading growth. Converting cost-effectiveness estimates to equivalent years of schooling in other contexts may lead to less favourable comparisons with other programs.

b. Implications for phonics programs in low-resource ECE classrooms

Aside from the typical challenges of piloting a new curriculum – including training teachers and getting materials to schools – we observed two challenges in the implementation of the FR program that provide lessons to other education providers who are considering similar reading programs in developing countries.

First, phonics-based instruction is a very different way of teaching reading in Liberia. The standard government reading curriculum uses a whole language approach coupled with memorization and repetitive call-and-response instruction. Many RAN teachers did not feel prepared to deliver a radically different approach to reading. In end-of-year surveys, 36% of teachers reported that they at least sometimes did not feel that they were sufficiently trained to deliver FR; we suspect that the true proportion may be higher, as some teachers may have felt pressure to provide an agreeable response. The fact that fewer than half of schools reported implementing all five FR cycles also likely points to a lack of teacher capacity to implement the program.

To effectively deliver phonics instruction in schools where the whole language approach is standard likely requires more initial training, more frequent retraining, and ongoing coaching of teachers. During our field visits, some SPMs also noted wide variation in teachers' ability to deliver FR. There may be opportunities to leverage teacher-to-teacher training (peer learning), or train-the-trainer models across and within schools.

Second, changing learning in school may have spill over effects to learning at home. In the case of the FR program, students exposed to the full program were 12 percentage points less likely to report spending time on reading at home compared with control students. One reason may be that FR students spent more time reading at school (at least one hour extra of school in every treatment school), and so they needed to focus on other subjects (like math) in their homework to compensate. Given that math proficiency does not suffer, while there are modest improvements in reading, this might be seen as an effective trade-off.

However, there may be other reasons for why FR students spent less time reading at home, such as students not having access to reading homework that corresponded to the new curriculum, or parents not knowing how to support children in a phonics-based approach to reading. Program designers should carefully consider the externalities of a new curriculum on learning at home.

Overall, these early results of the FR pilot suggest that there is potential for phonics-based, TaRL-inspired instruction to cost-effectively improve reading outcomes for ECE students in Liberia. We believe that these learnings are of value for any program looking to scale phonics and TaRL programs in ECE classrooms in West Africa.

References

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. and Walton, M., 2016. *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India* (No. w22746). National Bureau of Economic Research.

Bashir, S., Lockheed, M., Ninan, E. and Tan, J.P., 2018. *Facing forward: Schooling for learning in Africa*. World Bank Publications.

Campbell, F.A., Ramey, C.T., Pungello, E., Sparling, J. and Miller-Johnson, S., 2002. Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied developmental science*, *6*(1), pp.42-57.

Darvas, P. and Namit, K., 2016. How do you solve a problem like over-age enrolment. *World Bank Blogs*.

Duflo, A., Kiessel, J., Adrienne M., L., 2021 *Experimental Evidence on Alternative Policies to Increase Learning at Scale*. https://www.poverty-action.org/sites/default/files/publications/duflo%20kiessel%20lucas%20202103%20experimental.pdf

Gerber, S.A., & Green, P.D., 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.

Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A. and Yavitz, A., 2010. The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics*, *94*(1-2), pp.114-128.

Hwa, Y., Kaffenberger, M. and Silberstein, J., 2020. Aligning Levels of Instruction with Goals and the Needs of Students (ALIGNS): Varied approaches, common principles. *RISE: Research on Improving Systems of Education*, *22*, pp.1-25.

J-PAL., 2019. *Tailoring instruction to students' learning levels to increase learning*. https://doi.org/10.31485/pi.2522.2019.

JPAL. *Conducting cost-effectiveness analysis (CEA),* last modified July 2020, https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea,

Krafft, C., 2015. Increasing educational attainment in Egypt: The impact of early childhood care and education. *Economics of Education Review*, *46*, pp.127-143.

Martinez, S., Naudeau, S. and Pereira, V.A., 2017. Preschool and child development under extreme poverty: evidence from a randomized experiment in rural Mozambique. *World Bank Policy Research Working Paper*, (8290).

McKenzie, D., 2012. *Beyond baseline and follow-up: The case for more T in experiments*. Journal of Development Economics 99, no. 2, pp. 210-221.

Ministry of Education, Liberia (MoE-Liberia). 2020. *Education Sector Plan 2022/23-2026/27*.

https://planipolis.iiep.unesco.org/sites/default/files/ressources/liberia_esp_2022.pdf

Ministry of Education, Liberia (MoE-Liberia). 2016. *Getting to Best Education Sector Plan 2017-2021. https://www.globalpartnership.org/content/getting-best-education-sector-plan-2017-2021-liberia*

World Bank. 2016. *Liberia education sector analysis.* Ministry of Education. https://documents1.worldbank.org/curated/en/481011575583469840/pdf/Liberia-Education-Sector-Analysis.pdf

Romero, Mauricio, Justin Sandefur and Wayne Aaron Sandholtz. 2020. *Outsourcing Education: Experimental Evidence from Liberia*. American Economic Review, 110(2):364-400.

UNESCO., 2011, Education and Fragility in Liberia. https://unesdoc.unesco.org/ark:/48223/pf0000212197/PDF/212197eng.pdf.multi

UNICEF., 2018. *The situation for children in Liberia.* https://www.unicef.org/liberia/situation-children-liberia

Vromant, Nico, Line Kuppens, Mahuba Hazemba, Faith Kyulabantu Mwamba, and Emily Cupito. 2021. "Scaling Teaching at the Right Level." *Ensuring All Children Learn: Lessons from the South on What Works in Equity and Inclusion*: 101.

Woldehanna, T. and Araya, M.W., 2017. Early investment in preschool and completion of secondary education in Ethiopia: Lessons from Young Lives.

Tables

Table 1: Non-compliance

# FR cycles implemented	Assigned to Treatment		Assigned to Control	
	# schools	% T student sample	# schools	% C student sample
0	1	10%	29	73%
1	3	5%	0	0%
2	1	5%	1	10%
3	5	8%	4	8%
4	9	26%	1	4%
5	18	45%	2	5%

Table 2: Balance checks

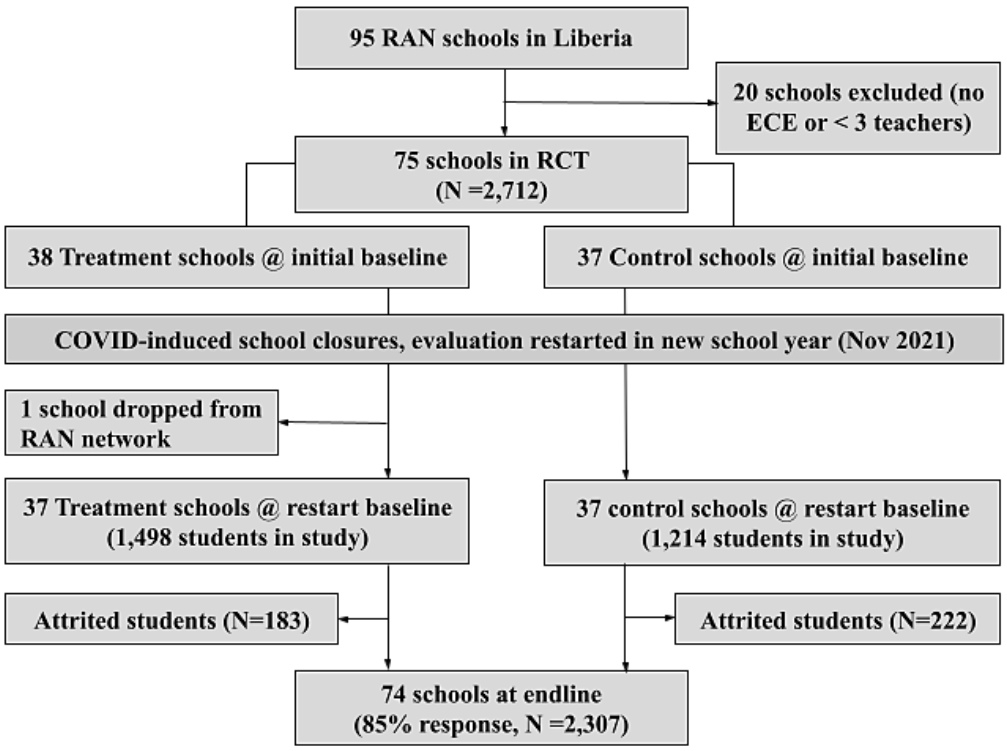
	Baseline sample (N = 2,712)			Endline sample (N = 2,307)		
	Treatment	**Control**	**p-value**	**Treatment**	**Control**	**p-value**
Student-level characteristics (at baseline)						
Reading level (out of 5)	1.49	1.51	0.69	1.50	1.51	0.86
Female	0.48	0.47	0.82	0.48	0.47	0.78
Age of student	8.97	8.89	0.60	8.94	8.88	0.65
School-level characteristics (at baseline)						
Number of over-age ECE students	40.49	32.81	0.17	*(No schools dropped out of the evaluation between baseline and endline)*		
Number of teachers in school	8.00	7.70	0.56			
Year RAN took over management of the school (Year 1 = 2016)	3.41	3.32	0.77			

Table 3: Treatment effects

	Control mean	ITT (SE)	ITT p-value	TI (SE)	TI p-value
Reading level at endline (out of 5)	1.84	0.07 (0.05)	0.14	0.12 (0.07)	0.10
Reading level at endline (standardized)	0.00	0.17 (0.11)	0.14	0.28 (0.17)	0.10
Math questions correct (out of 4)	2.80	0.13 (0.19)	0.51	0.21 (0.30)	0.48
Attendance rate (SPM & SL)	0.81	0.06 (0.04)	0.09	0.08 (0.04)	0.07
Attendance rate (SPM only)	0.76	0.08 (0.04)	0.04	0.11 (0.05)	0.03
Likes school	0.88	0.04 (0.03)	0.18	0.07 (0.06)	0.20
Likes reading in school	0.84	0.01 (0.03)	0.71	0.02 (0.05)	0.71
Has friends in school	0.97	0.00 (0.01)	0.74	0.01 (0.02)	0.73
Reads at home	0.85	-0.07 (0.04)	0.05	-0.12 (0.06)	0.05
Parents help with homework	0.82	0.01 (0.03)	0.75	0.02 (0.05)	0.74

Figures

Figure 1: Sampling and randomization



Appendices

Tables Appendix

Table A1: Descriptive statistics

	All	Treatment	Control
Reading level (baseline)			
Average level	1.50	1.49	1.51
Average Letter correct (out of 10 letters tested)	5.80	5.70	5.93
Average Beginning sound correct (out of 3 beginning sounds tested)	0.87	0.83	0.91
% Letter pass	49.9%	49.3%	50.7%
% Beginner sound 'pass'	27.7%	26.0%	29.7%
% CVC pass	0.1%	0.1%	0.2%
Reading level (endline)			
Average level	1.87	1.89	1.84
Average Letter correct (out of 10 letters tested)	8.39	8.43	8.33
Average Beginning sound correct (out of 3 beginning sounds tested)	1.62	1.69	1.53
% Letter pass	83.0%	84.1%	81.5%
% Beginner sound 'pass'	55.0%	58.7%	50.2%
% CVC pass	3.6%	4.6%	2.3%
Reading growth			
Average growth	0.36	0.39	0.33
% moved up	37.7%	39.3%	35.6%
% stayed same	59.4%	58.1%	61.1%
% went down	2.9%	2.6%	3.3%
Math problems			
Average correct (out of 4)	2.87	2.92	2.80
% recognized 12	66.9%	68.3%	65.1%
% recognized 15	63.4%	65.4%	60.7%
% recognized 18	61.1%	62.4%	59.3%

% counted beans	95.5%	95.8%	95.1%
Attendance rate			
SPM & SL	84.7%	87.3%	80.8%
SPM only	81.7%	85.5%	76.3%
SL only	88.3%	89.1%	87.0%
Perceptions of school & reading			
% like school	90.4%	92.2%	88.0%
% like reading in school	84.6%	85.2%	83.8%
% has friends in class	97.4%	97.3%	97.5%
Reading behaviour			
% reads at home	81.8%	79.6%	84.6%
% parents help with homework	82.0%	82.2%	81.7%

Table A2: Reading results

	Control mean	**ITT effect**	**p-value**	**Treat intensity effect**	**p-value**
All students	1.84	0.07	0.14	0.12	0.10
Baseline Letter	1.70	0.09	0.19	0.15	0.14
Baseline CVC	1.97	0.06	0.14	0.10	0.12
Boys	1.86	0.09	0.11	0.15	0.07
Girls	1.82	0.05	0.30	0.08	0.25
Year RAN took over management of the school					
Year 1 (2016)	1.94	0.04	0.75	0.04	0.70
Year 2 (2017)	1.84	-0.05	0.44	-0.14	0.54
Year 4 (2019)	1.82	0.17	0.01	0.25	0.01
Year 5 (2020)	1.91	-0.22	0.26	-0.23	0.19

Table A3: Reading results in SDs

	ITT effect	p-value	Treat intensity effect	p-value
All students	0.17	0.14	0.28	0.10
Baseline Letter	0.21	0.19	0.35	0.14
Baseline CVC	0.14	0.14	0.22	0.12
Boys	0.21	0.11	0.36	0.08
Girls	0.11	0.30	0.19	0.25

Table A4: Reading results TOT bounds

	Control mean	TOT (lower bound)	p-value	TOT (upper bound)	p-value
All students	1.84	0.11	0.10	0.19	0.10
Baseline Letter	1.70	0.13	0.13	0.23	0.14
Baseline CVC	1.97	0.09	0.12	0.16	0.13
Boys	1.86	0.14	0.08	0.25	0.07
Girls	1.82	0.07	0.24	0.13	0.26

Table A5: Other reading outcomes

	Control mean	ITT effect	p-value	Treat intensity effect	p-value
% moved up	35.6%	4.8%	0.20	8.0%	0.16
% Letter pass	81.5%	5.0%	0.21	8.3%	0.15
# Letter correct (out of 10)	8.33	0.30	0.23	0.49	0.19
% Beginner sound 'pass'	50.2%	6.2%	0.15	10.4%	0.13
# Beginning sound correct (out of 3)	1.53	0.10	0.34	0.16	0.33
% CVC pass	2.3%	1.8%	0.35	2.9%	0.35

Table A6: Math results

	Control mean	ITT effect	p-value	Treat intensity effect	p-value
All students	2.80	0.13	0.52	0.21	0.48
Baseline Letter	2.44	0.05	0.86	0.08	0.86
Baseline CVC	3.15	0.19	0.17	0.30	0.12
Boys	2.91	0.08	0.66	0.13	0.63
Girls	2.68	0.19	0.42	0.30	0.36

Table A7: Other math outcomes

	Control mean	ITT effect	p-value	Treat intensity effect	p-value
% recognized 12	65.1%	2.5%	0.68	4.1%	0.66
% recognized 15	60.7%	5.4%	0.44	8.9%	0.39
% recognized 18	59.3%	4.2%	0.52	6.9%	0.49
% counted beans	95.1%	0.9%	0.45	1.6%	0.43

Table A8: Attendance results (replication of Graphs 3 & 4)

	Control mean	ITT effect	p-value	Treat intensity effect	p-value
Attendance rate (SPM & SL)					
All students	80.8%	6.1%	0.09	8.1%	0.07
Baseline Letter	80.7%	5.2%	0.17	6.9%	0.14
Baseline CVC	80.8%	6.1%	0.13	8.1%	0.10

Boys	80.2%	7.3%	0.07	9.9%	0.04
Girls	81.4%	4.6%	0.23	6.0%	0.24
Attendance rate (SPM only)					
All students	76.3%	8.3%	0.04	11.4%	0.03
Baseline Letter	76.8%	6.5%	0.15	8.7%	0.12
Baseline CVC	75.9%	9.4%	0.03	13.0%	0.01
Boys	75.3%	9.6%	0.03	13.3%	0.01
Girls	77.6%	7.0%	0.12	9.7%	0.14

Table A9: Non-academic outcomes subgroups

	Control mean	**ITT effect**	**p-value**	**Treat intensity effect**	**p-value**
% like school					
All students	88.0%	4.3%	0.17	7.1%	0.20
Baseline Letter	87.1%	4.1%	0.26	7.1%	0.28
Baseline CVC	88.9%	4.2%	0.20	6.8%	0.22
Boys	89.1%	3.1%	0.29	5.2%	0.31
Girls	86.8%	5.8%	0.12	9.5%	0.15
% like reading in school					
All students	83.8%	1.2%	0.71	2.0%	0.71
Baseline Letter	82.3%	0.2%	0.95	0.4%	0.94
Baseline CVC	85.1%	1.5%	0.62	2.5%	0.62
Boys	84.3%	2.7%	0.41	4.4%	0.41
Girls	83.1%	-0.2%	0.96	-0.3%	0.96

% has friends in class					
All students	97.5%	0.4%	0.74	0.6%	0.73
Baseline Letter	97.5%	0.0%	0.99	0.0%	1.00
Baseline CVC	97.4%	0.9%	0.44	1.5%	0.43
Boys	98.1%	-0.1%	0.91	-0.2%	0.91
Girls	96.8%	1.0%	0.47	1.7%	0.45
% reads at home					
All students	84.6%	-7.3%	0.05	-12.1%	0.05
Baseline Letter	81.7%	-7.7%	0.11	-13.1%	0.13
Baseline CVC	87.3%	-5.9%	0.10	-9.5%	0.08
Boys	83.4%	-3.5%	0.44	-5.8%	0.44
Girls	85.9%	-11.4%	0.00	-18.7%	0.00
% parents help with homework					
All students	81.7%	0.9%	0.75	1.6%	0.74
Baseline Letter	80.3%	1.0%	0.79	1.7%	0.77
Baseline CVC	83.0%	0.8%	0.80	1.3%	0.81
Boys	80.0%	3.5%	0.29	5.8%	0.27
Girls	83.5%	-1.8%	0.61	-2.9%	0.61

Table A10: Reading results, IPW

	Control mean	ITT effect	p-value	IPW	p-value
All students	1.84	0.07	0.14	0.04	0.33
Baseline Letter	1.70	0.09	0.19	0.05	0.46
Baseline CVC	1.97	0.06	0.14	0.05	0.27
Boys	1.86	0.09	0.11	0.06	0.27
Girls	1.82	0.05	0.30	0.03	0.52

Table A11: Reading results, Lee Bounds

	Control mean	Upper bound	p-value	Lower bound	p-value
All students	1.84	0.10	0.00	-0.01	0.67
Baseline Letter	1.70	0.11	0.00	0.01	0.68
Baseline CVC	1.97	0.10	0.00	-0.03	0.12

Note: Lee Bounds could not be constructed for only male students and only female students.

Endline survey instrument Appendix

ECE Evaluation: Endline (STUDENT)

RISING ACADEMY NETWORK

Letter Names

E	N	A	P	J
o	i	h	v	f

Words with Blends

slap	clip	blob	limp
lamp	raft	stamp	drift

Beginning Sounds

Sentence

Janet helped her dad paint the kitchen.
She got covered in white marks.

CVC Words

tab	sip	hop	pad
dim	mat	fix	red

Developed from materials created by the ASER Center

असर ASER

ECE Evaluation: Endline (STUDENT)

RISING ACADEMY NETWORK

Story

It is John's birthday. He is 8 years old. He is having a party. John has asked two friends to come to his party. Their names are Nelson and Prince.
John is waiting for Nelson and Prince to come. They will play games and sing songs. Then they will read a story. After the story Nelson and Prince will go home.

Developed from materials created by the ASER Center

असर ASER

READING INSTRUCTIONS: Endline (ASSESSOR)

RISING ACADEMY NETWORK

Letter Names

Today we are going to do some reading activities. Just relax and feel free. Take your time and try your best. Point to the letters and tell me the name. **Record the # of letters read incorrectly and move to Beginning Sounds**

E	N	A	P	J
o	i	h	v	f

Beginning Sounds

Now we are going to play a listening game about the sounds at the beginning of a word.
The word ***cat*** *starts with the /c/ sound. Listen to the words I say and tell me which one starts with the same sound /c/ like cat.*

mouse, ***cup****, friend In this one, cat and cup both have the /c/ sound at the beginning!*

Q1*. Let's do another one. Ready?* ***name*** *starts with the* ***/n/*** *sound. Now listen to these words and tell me which one starts with the /n/ sound like name.*

baby, toes, ***neck***

Q2*. Let's do another one. Ready?* ***pen*** *starts with the* ***/p/*** *sound. Now listen to these words and tell me which one starts with the /p/ sound like pen.*

sat, ***pig,*** *man*

Q3*. Let's do another one. Ready?* ***hat*** *starts with the* ***/h/*** *sound. Listen to these words and tell me which one starts with the /h/ sound like hat.*

head*, feet, rat*

Record the # of incorrect answers and move to CVC Words

CVC Words

Next, I have some words for you to read. You can use what you know about blending sounds to read them.

tab	sip	hop	pad
dim	mat	fix	red

Record # of incorrect answers. If the students makes more than 2 mistakes, STOP the assessment.

Words with Blends

Next, I have some longer words for you to read. You can use what you know about blending sounds to read them.

slap	clip	blob	limp
lamp	raft	stamp	drift

Record # of incorrect answers. If the students makes more than 2 mistakes, STOP the assessment.

Sentence

Next, I have some sentences for you to read. You can use what you know about blending sounds to read them.

Janet helped her dad paint the kitchen.
She got covered in white marks.

Record # of mistakes. More than 3 mistakes, STOP the assessment.

READING INSTRUCTIONS: Endline (ASSESSOR)

RISING ACADEMY NETWORK

Story Reading

Next, please read this Story for me.

It is John's birthday. He is 8 years old. He is having a party. John has asked two friends to come to his party. Their names are Nelson and Prince.
John is waiting for Nelson and Prince to come. They will play games and sing songs. Then they will read a story. After the story Nelson and Prince will go home.

Listen to the child read the passage and record the number of mistakes.
If the child makes more than 3 mistakes, STOP the assessment

Story Comprehension

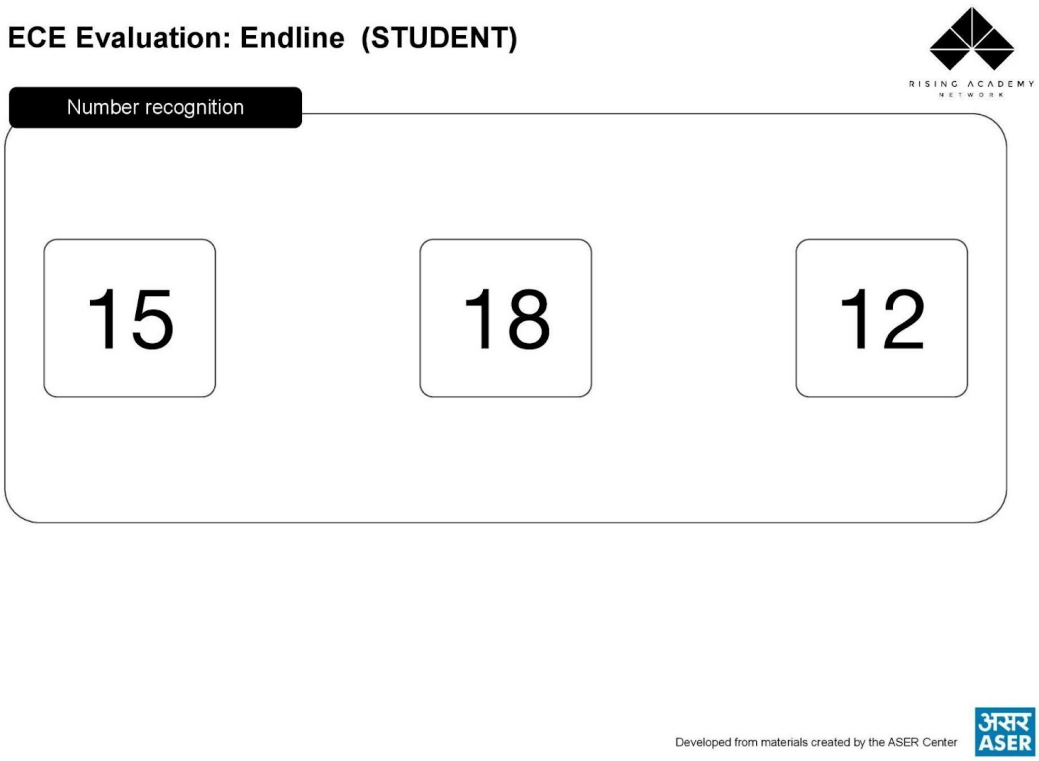
Next, please answer these questions.

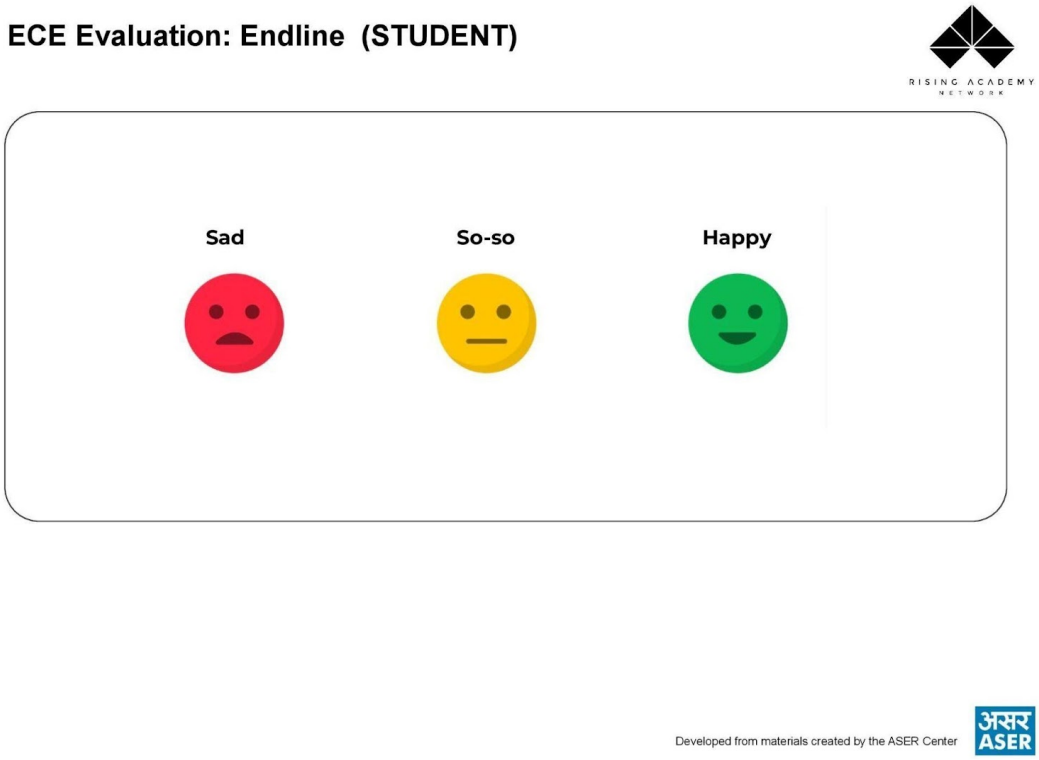
Q1*: Who came to John's party?*
Q2*: Why is John having a party?*

Example answers	
A1: Two friends Nelson and Prince	A2: It is his birthday. John is celebrating his birthday.

Record # of incorrect answers.

0 = Both answers correct.,
1 = 1 answer incorrect, 1 answer correct
2 = 2 answers incorrect





Cost-effectiveness analysis Appendix

Approach to CEA

We estimate the cost-effectiveness of the FR program for over-age ECE students. This

is a somewhat conservative estimate since some of the costs are incurred at the school-level, and thus would not scale proportionally if FR were implemented in upper grades (Grades 3-6, as envisioned). We focus on over-age ECE students for two reasons. First, from our evaluation we only have impact estimates for over-age ECE students and not for other grades - any extrapolation of impact from over-age ECE to others would require strong assumptions. Second, not all schools have been implementing FR to Grades 3-6 due to capacity constraints (lack of teachers and/or lack of training), so it does not appear that RAN could implement FR to other students and grades without significant additional investment in teacher training or hiring.

We separate out the fixed start-up costs of the FR program (e.g., curriculum development) from the costs that would be incurred in future implementation (trainings, materials, etc.), and only include the latter in our CE estimate. We exclude any costs that would be incurred in the course of normal school programming, including teacher and staff salaries. We also exclude costs associated with COVID-induced lockdowns and restarting the program, including refresher trainings.

Our calculation follows the standard approach applied by J-PAL and other researchers to estimate the cost-effectiveness of education programs:

100(# students in FR*Avg impact on learning per student in standard deviations)/(Total cost of FR in USD) = Learning gains (in SDs) per USD $100 spent*

Inputs to CEA

We solicited the cost of inputs from the RAN team and describe these below, along with any adjustments that we made:

- Training of RPMs & SPMs: $8,000
- Training of all teachers: $44,000. ECE teachers are ~⅓ of the workforce, and the treatment group is ~½ of all ECE teachers. Since training costs scale proportional to trainees, we estimate training of ECE teachers in the treatment group = ⅙ * 44,000 = $7,333
- Material costs for over-age ECE implementation: $10,000. Delivery costs: $8,000. Additional printing costs for student assessments after each cycle: $500
- No additional staffing costs to implement FR
- Total cost: $8,000 + $7,333 + $10,000 + $8,000 + $500 = $33,833

CEA calculation

- Number of over-age ECE students that received the program in the treatment group, scaled according to the amount of the program that each received: 675*1 + 396*0.8 + 119*0.6 + 73*0.4 + 80*0.2 + 155*0 = 1,108
- Impact estimate of the full 5-cycle program in SDs from Appendix **Table A3**: 0.28 SD.
- Total cost to deliver FR to over-age ECE students in the treatment group:

$33,833.

CEA = 100*1108*0.28/33,833 = **0.91 SD learning gains per $100**

Comparison to other programs

Though FR is in its early stages and is yet to be iterated for improvement, this cost-effectiveness estimate is comparable to other pedagogical innovations highlighted in J-PAL's CEA of education programs.

Our cost-effectiveness estimate of FR for over-age ECE students comes with several caveats, which may be leading us to overestimate or underestimate the true cost-effectiveness of the program:

- Extending the full program to Grades 3-6 consistently may not require a proportional increase in costs.
- Leveraging opportunities to strengthen program implementation may increase impact.
- Strengthening program implementation may incur additional costs.
- The control group, against which treatment effects are standardised, had low average reading growth. Converting cost-effectiveness estimates to equivalent years of schooling in other contexts may lead to less favourable comparisons with other programs.

+/- The full program impact (TI estimator) is not precisely estimated (not statistically significant at the 10% level, based on additional assumptions); the true impact may be higher or lower.