Design, Implementation, and Initial Impact Results from a Remedial Reading Program in Egypt

Rebecca Stone, EdD
Researcher
American Institutes for Research®

Essam Assaad
Literacy Specialist
American Institutes for Research®

Abstract

This paper reports on the design, implementation, and initial impact results of a remedial reading program in Egypt during the 2012–2013 school year. The implementation of this program and the research conducted to measure its effectiveness was done as a result of a need expressed by the Egyptian Ministry of Education and evidenced by poor Early Grade Reading Assessment (EGRA) scores in the Arabic language in 2009. To address this problem of upper primary students with inadequate literacy skills, which often led to drop out and failure, the US Agency for International Development (USAID)-funded Education Support Program (ESP) targeted this reading intervention at students in grades four and five at high risk of dropping out due to poor reading skills. The program is designed to provide these students with additional reading support during the school day and to help them catch up to their peers. To do this, we developed and applied a screening test to select students with poor literacy skills into the remedial program. During the first semester of the program, approximately 15,000 students were selected. We then administered the EGRA to a subset of these students both pre- and postintervention. We used a Regression Discontinuity Design (RDD) whereby students scoring just below and above the cut score on the screening test were selected into the treatment and comparison groups, respectively. Findings from one semester of the program show that the estimated impacts are positive for all constructs. The effect sizes that are statistically significant range from 0.3 to 0.39 standard deviations, which are substantial for a relatively short-term intervention such as this one. The most significant estimated effects were on letter-sound knowledge, phonemic-awareness ryme, and listening comprehension.

The Education Support Program (ESP) is an American Institutes for Research® (AIR)-implemented and US Agency for International Development (USAID)-funded project in Egypt. It aims to support educational service delivery for Egyptian children during a unique and very critical time in Egypt's history. While Egyptian society is going through a complex transitional process towards democratization and reform, education and the quality of educational services, in general, are at the heart of these complex changes. One critical intervention of ESP is the design and implementation of remedial reading and writing activities that aim to assist teachers in providing additional support to improve the reading and writing abilities of their students. The intervention and the research around it are the focus of this paper.

Background: Remedial Reading and Writing Activities

The ESP remedial reading and writing activities were a response to an educational need expressed by the Ministry of Education (MOE) and evidenced by poor Early Grade Reading Assessment (EGRA) scores in the Arabic language in 2009.[1] In early 2013, ESP administered a screening test for reading, which independently confirmed the 2009 results. As a result, ESP began implementing remedial reading and writing activities specifically targeting struggling readers in grades four and five, as they are at high risk for falling behind in their studies, losing motivation, and dropping out of school. This is not a one-size-fits-all intervention. These activities are designed to reach students with a range of different reading challenges to bring them to the level they need to be at *to read and to learn* in all of their subjects.

In the beginning stages of the program (October 2012), ESP staff worked with the MOE and local education stakeholders to identify schools with the lowest reading scores in the various

---

[1] The EGRA is an oral student assessment designed to measure the most basic foundation skills for literacy acquisition in the early grades: recognizing letters of the alphabet, reading simple words, understanding sentences and paragraphs, and listening with comprehension. The EGRA was conducted in Egypt in 2009, with 2,876 pupils in grades two, three, and four, by RTI International through the Girls' Improved Learning Outcomes project.

Egyptian governorates (and specifically, a cross-section of idarras [districts] within those governorates).[2] Education staff who were interested in having their schools participate in the activities and motivated to implement additional interventions were selected. In total, 75 schools were selected throughout multiple idarras in six governorates: Cairo, Qena, Ismailia, Minia, Alexandria, and Gharbeia. ESP project staff and assessment experts then developed a screening test to identify and categorize individual reading skills to determine which students needed the remedial reading and writing intervention. The screening test was given to 17,597 students in these 75 schools. By using a deliberate standard-setting effort, a cut score was established and applied to the results. Through this process, 5,515 students, whose scores on the screening test were below the cut score, were selected to participate in Level 1 of the program (i.e., the beginning level for those students with the most severe reading and writing challenges).

With AIR literacy and assessment experts based in Washington, DC, the ESP technical team was supported to develop a curriculum for Level 1 activities. The curriculum includes teacher and student guides that cover the content of the program. Teacher guides provide step-by-step instructions and activities for teachers to facilitate with their students, and the student guides provide opportunities for additional reading and writing practice that are in line with each lesson. All procedures and program materials went through an extensive process of review by AIR literacy experts and subsequent revision by teachers and MOE staff. A core group of trainers received training in how to teach their colleagues about the activities and how to use the curriculum and materials effectively. These trainers then went out to each of the targeted idarras and trained those teachers slated to administer activities. In addition to being trained in the delivery of the remedial reading and writing activities, teachers also were trained in basic

---

[2] For administrative purposes, Egypt is divided into 27 governorates (which are similar to states) and each of the governorates is divided into idarras (districts).

classroom management and teaching strategies. This additional training was added on the basis

of recommendations of ESP regional staff, supervisors, and senior teachers who observed the

initial implementation of the program.[3]

The remedial reading and writing activities were designed to be administered three times

each week for one hour each session. During this hour, teachers worked with a group of no more

than 20 identified students and focused on the individual needs of these students, as identified in

the screening test. However, some variation occurred in how individual schools implemented the

program, to accommodate the variations in resources and constraints. Most schools incorporated

activities into the school day by substituting remedial reading and writing classes for "activity

classes" offered during the same time. The number of days and hours per week that students

participated in the program varied by school, but at least one-half of the students completed 28 or

more of the 36 lessons. In total, 5,515 students began Level 1 of the reading and writing

intervention activities during the spring 2013 semester; 2,699 of those students completed at least

28 of the 36 sessions by the end of the school year or first month of summer.

Key Technical Features

The ESP remedial reading and writing activities are phonics-based and build foundational

literacy skills. The activities are provided by Arabic-language teachers who are perceived as

those with the most in-depth knowledge of the language and its instruction in Egypt. Level 1 of

the intervention focuses on building letter-sound knowledge, vocabulary, and decoding skills. All

vocabulary comes from the grade-level textbook. It follows a frequency approach by starting

with the most commonly used Arabic letters and their sounds before moving on to less

frequently used letters. Each lesson focuses entirely on one letter, its sound, and letter script,

---

[3] Because all students in a teacher's classroom benefit from these additional teacher-training activities, the impact analysis detailed below does not capture the impact of these additional services. The impact analysis is based on a comparison of students identified for remediation and students not thus identified.

including the differences in how it is written at the beginning, middle, and end of words. Every new lesson teaches one new letter and reviews only the letters previously taught. In this manner, after only five lessons, students master at least five letters and are able to write words and even simple sentences using those letters. This helps students quickly gain confidence in their new skills as they are able to read and write short words and sentences after only minimal lessons. The lessons also rely on finger coding (signs) to signify different vowels or diacritics, especially in the early lessons. Each lesson ends with an application activity to consolidate acquired skills and knowledge. The lessons focus on skill building through directed practice, games, group and partner work, and reading by both the teacher and students in multiple formats (e.g., choral reading, individual reading aloud, paired reading, silent reading).

The activities work with homogeneous groups of students at the same instructional level (grade and ability), which helps teachers combine their instruction for various ability levels at the same time. This also helps students feel that they are with their peers and not falling behind more advanced students. Level 1 of the intervention consists of 36 one-hour remedial reading and writing sessions. Sessions take place during the school day and replace one "activity class." This format was decided by the schools and the MOE. The format is preferred by teachers, as they are not required to stay after school. Since the program was not fully implemented until the middle of the school semester, many schools continued the activities during the summer break to complete all 36 lessons. Supervisors and school-based senior teachers were trained alongside the teachers in pedagogy and curriculum so as to be able to observe and provide feedback and support to the teachers.

Research Design

This section describes the empirical strategy we used to estimate the causal effect of the remedial reading and writing activities on student literacy outcomes. The goal of this strategy is to isolate changes in student achievement that are directly attributable to the program from those due to other factors. The principal challenge in the identification of the causal impact of the program is to answer the following counterfactual question: *What would have been the value of the outcome variables for students assigned to the remedial intervention had the activities not been implemented?* In practice, however, students are either subject to the intervention, or they are not. Thus, estimating this counterfactual ideally requires a comparison group that did not have access to the intervention, but is identical to the participants in terms of the distribution of characteristics that affect the outcomes of interest.

Although there are different methods to construct this counterfactual, the choice of an evaluation method depends on the mechanism by which individuals are selected to participate in the activities. In this section, we show that the way students were selected into the remedial reading and writing activities lent itself to a regression discontinuity design (RDD) to estimate its effects on student literacy outcomes. This is a highly rigorous impact evaluation design that is robust to many of the biases that are usually associated with evaluations of selective interventions such as this one.

Regression Discontinuity Design

RDD is a quasi-experimental design for impact evaluation. In this design, individuals are assigned to treatment and comparison groups using an observed continuous measure, known as the assignment (or "forcing") variable. In particular, individuals falling on one side of a

predetermined cut point of this variable receive treatment, whereas those falling on the other side of the cut point are excluded from the treatment and constitute a comparison group.

RDD identification is based on the idea that, relatively close to a pre-established cut point, the assignment rules for participation are essentially arbitrary. That is, comparing the outcomes of individuals, whose values on the assignment variable fall into a small interval around an assignment cut point will be similar to a randomized experiment at the cut point. Those just below the cut point are expected to be similar to those just above it, and statistically controlling for the assignment variable eliminates any remaining differences between the two groups. Yet, in the case of these remedial reading and writing activities, only those below the cut point are treated. Thus, a regression-adjusted comparison of the average outcomes of interest (e.g., test scores) between those just below (i.e., treated) and those just above the cut point (i.e., comparison group members) will produce good estimates of the average effects of the activities for students at or close to the cut point.

The most important limitation of impact analyses like these is that the results are not generalizable far away from the cut point. Thus, intervention impacts for students who scored much worse or much better than those near the cut point may be different from those we present in this report.

Sample Selection

Assignment to the remedial reading and writing activities was determined on the basis of a literacy screening test. The screening test was given to all fourth- and fifth-grade students within the intervention schools to determine which students were in need of the remedial reading and writing services. When designing this test, the goal was to develop a test that could: (1) accurately screen students in need of the intervention, and (2) be administered efficiently to large

7

numbers of students in a short amount of time. To reach both of these goals, we included

measures of student proficiency in the early literacy skills (listed on the following page). We

excluded oral reading fluency from this list because it was not possible to measure this skill

without doing one-on-one administration, which was not feasible for the screening test.

The screening tool was developed by assessment experts and literacy specialists at AIR

and then was piloted in 75 schools. It covers the following early literacy skills:

- Letter recognition,

- Letter/sound correspondence,

- Decoding letters and diacritics at the beginning, middle, and end of words,

- Phonological awareness,

- Spelling,

- Listening comprehension, and

- Reading comprehension.

The final version of the screening test is a 71-item measure with 10 constructs. Of the 71

items, 54 involve a simple multiple-choice task in which the student chooses the one item out of

three or four that best corresponds to a question posed by the examiner. The other 17 items

involve productive writing by the student. The tasks included in the screening test are listed in

table 1.

Table 1. Screening Test Tasks

Task #	Task instructions	Content domain	Question type
1	Teacher reads the letter name twice, and students are asked to circle the correct one from a list of four letters.	Letter name	Multiple choice
2	Teacher pronounces a letter-sound twice, and students are asked to circle the correct one from a list of four letters.	Letter-sound correspondence	Multiple choice

Task #	Task instructions	Content domain	Question type
3	Teacher pronounces a word twice, and students are asked to identify if a given letter came at the beginning, middle, or end of the word.	Decoding letters at beginning, middle, and end of words	Multiple choice
4	Teacher pronounces a word twice, and students are asked to select the word with the correct diacritics.	Letter sound with diacritics	Multiple choice
5	Teacher pronounces a pair of words twice, and students are asked to put a tick under identical or different.	Phonological awareness	Multiple choice
6	Teacher pronounces a word twice, and students are asked to circle the correct word from a list of four.	Phonological awareness	Multiple choice
7	Students are given three or four letters and are asked to form two different words from each group of letters.	Spelling/basic vocabulary	Productive writing task
8	Teacher reads aloud a 70-word text. Students then listen to six questions and select the correct answer from a list of three choices (read aloud by the teacher).	Basic listening comprehension	Multiple choice
9	Teacher reads 12 words aloud, one at a time and two times each. Students are asked to write the words.	Spelling	Productive writing task
10	Students are asked to read an 85-word text and respond to six questions by selecting the correct answer from a list of three choices.	Basic reading comprehension	Multiple choice

After piloting, the screening test was revised and administered by trained Arabic-language teachers in January and February 2013 to 17,597 students in the 75 schools participating in the remedial reading and writing activities.

To establish a content-based cut point (the score below which students would receive the intervention), we reviewed the skills being tested and determined that the weak readers (i.e., those in need of the intervention), would score a maximum of 80% on the first eight constructs and would not be able to master either the reading comprehension, or the dictation sections. Since the maximum score students could attain on the first eight constructs was 53 points, if they

achieved a mastery level of 80% that gives us 42.4 points rounded down to 42 points as the cut point. Therefore, the students scoring 42 points or below were selected into Level 1 of remedial reading and writing activities, and thus received the intervention. Of 17,597 students tested, approximately 7,500 scored 42 or below, and ultimately 5,515 received the intervention.[4] Table 2 shows the numbers of students who completed the screening test in each governorate and the percentage of those students who scored below the cut point and were selected into Level 1.

**Table 2.** Number of Students Who Completed the Screening Test and Percent Who Scored Below the Cut Point

Governorate	Total students tested	Level 1
Minia	3,315	70%
Ismailia	3,491	48%
Cairo	2,787	47%
Qena	2,857	39%
Alexandria	1,072	18%
Gharbeia	2,231	17%

It is important to note that there is quite a difference in literacy skills and remediation needs between students who scored 10 and students who scored 41, although both sets of students were selected into the same Level 1 program. Moreover, students who scored above 42 also are not necessarily proficient. Many still struggle with reading comprehension and writing skills; that is why, in subsequent semesters, Level 2 of the program will be implemented to address these learning needs.

---

[4] About 2,000 students scored below the cut-point but did not receive the intervention for the following reasons: lack of interest in participation on the part of the schools and administrators, a shortage of Arabic teachers to run the program, and political unrest and instability in certain regions.

As shown in figure 1, students who scored 42 points or below on the screening test were assigned to remedial reading and writing activities, while those who scored above this cut point were not assigned to those activities. To make efficient use of the available data, we only included in the analysis students who scored within 10 points of the cut point; this group included 1,029 students, of whom 508 were in the comparison group and 521 were assigned to the treatment.[5]

Figure 1. Screening Test Histogram for ESP Remedial Reading and Writing Assignment
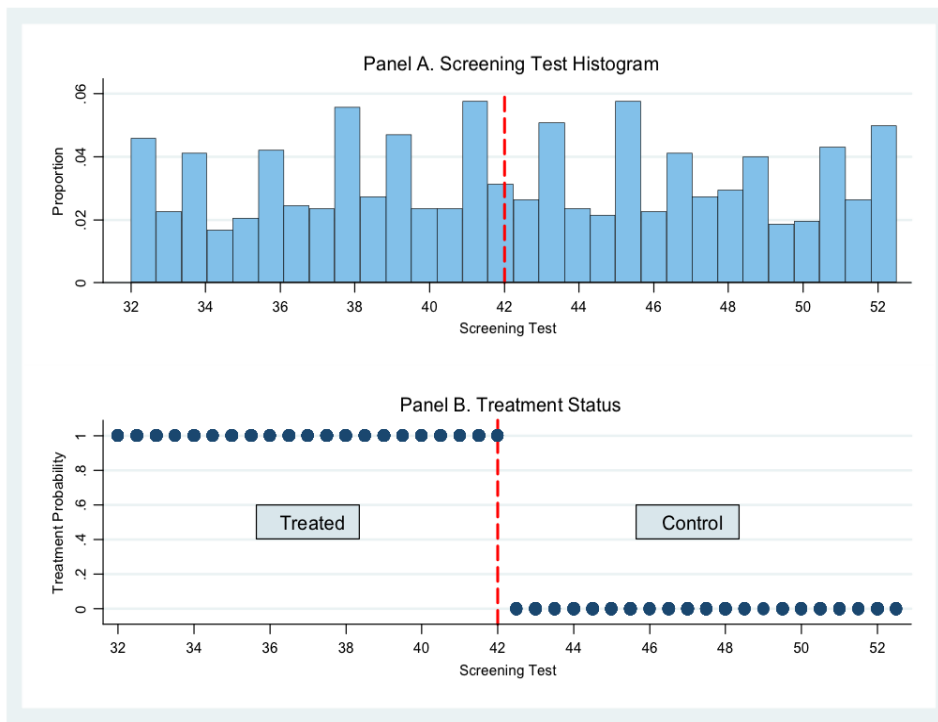
Early Grade Reading Assessment (EGRA)

To evaluate the effectiveness of the intervention, we relied on literacy outcomes from the Arabic-language EGRA. The EGRA is composed of different constructs designed to assess

---

[5] We established this 20-point range using "optimal bandwidth" procedures that are commonly used in impact studies involving an RDD design (Imbens and Kalyanaraman 2011). These procedures balance off the total available sample size for the analysis with the increased statistical noise associated with assignment variable values that are far removed from the cut point. To ensure that this choice of bandwidth did not unduly influence the impact estimates, we conducted sensitivity analyses with larger bandwidths, which produced consistent results. Thus, our impact estimates are robust to the choice of the bandwidth around the RDD cutoff point.

reading skills crucial to student success in reading and comprehension (Gove and Wetterberg 2011). These constructs are based on research that explicitly supports a comprehensive approach to reading acquisition, including alphabetic knowledge and process, phonemic awareness, reading fluency and comprehension, and listening comprehension. The EGRA is administered orally by a trained enumerator on a one-to-one basis with each student, and it takes about 15 minutes per child to administer.

To better suit the reading level of students in grades four and five, the EGRA used for our evaluation was modified to include higher level comprehension questions. In addition to explicit fact-finding questions, which are typically included in the EGRA, we added two implicit inferencing questions to the reading comprehension tasks. We also added a dictation task in which students were asked to write seven words with two to seven letters per word. The final EGRA subtasks were as follows:

- Letter-name knowledge
- Letter-sound knowledge
- Phonemic-awareness onset
- Phonemic-awareness ryme
- Familiar word reading
- Unfamiliar non-word reading
- Oral-reading fluency
- Listening comprehension
- Reading comprehension
- Dictation

Data Collection

In January 2013, we administered the screening test to 17,597 students. Results were entered at the regional level and then consolidated at the ESP project office. The cut point was applied to the results, and 5,515 students were selected to participate in Level 1 of the remedial reading and writing intervention. The RDD (described above) was then used to select treatment and comparison groups around the cut point to receive a modified version of the Arabic EGRA both before and after the intervention. The ESP technical team provided training to trainers from each of the six governorates in how to administer the EGRA. These trainers then returned to their governorates and tested students in both the treatment and comparison samples.

The EGRA was administered to approximately 1,000 students before they began the reading and writing intervention. During school hours, students attended remedial reading and writing activities for the semester. Because the start of the intervention did not coincide with the beginning of the school year, most schools decided to continue into summer vacation to complete activities. When the treatment group students in a school completed all 36 lessons, those in the treatment and comparison groups were given another EGRA to see how they progressed.

Empirical Strategy

There are different ways to implement an RDD to estimate effects. A simple way to do so is to compare average outcomes between treatment and comparison scores in a small distribution to the left and right of the cut point. Figure 2 shows the relationship between the screening test and one of the literacy constructs, namely, the number of correct answers in the dictation section. Students just below the cut point (treated) recorded about six correct answers in the dictation test; students just above the cut point (comparison) recorded just four. Thus, the RDD estimates

that students gain about two additional correct answers as a result of remedial reading and writing activities.

Although this graphic representation of an RDD estimate is convenient for expositional reasons, the effects of the intervention are in practice estimated by using multivariate regression analysis. This allows us to estimate the effect of activities more precisely after controlling for other factors that may also affect student outcomes. Thus, each one of the EGRA constructs is regressed on a dummy variable for the treatment group, a quadratic function of the screening test, idarra fixed effects to account for regional differences in student performance, and the student's baseline EGRA score on the same construct. Note that the dummy variable for the treatment group is the explanatory variable of interest, as it captures the differential effect of the program between treatment and comparison groups. A complete description of the results is presented in the Findings section.

Figure 2. Estimating Intervention Effects Using an RDD
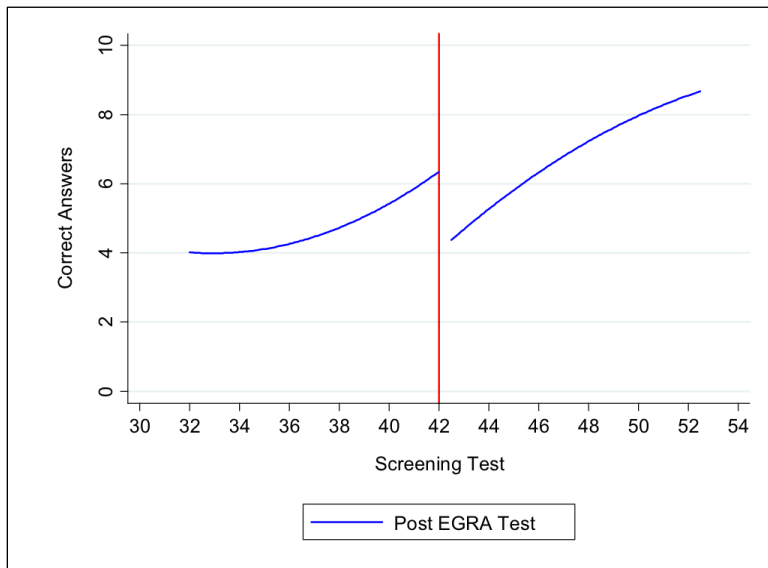
Limitations of the Study

   Attrition. As shown in table 3, all 1,029 students in the evaluation sample took the

modified baseline EGRA test in January and February 2013 before remedial reading and writing

activities were implemented. However, at the end of the intervention, only 457 of the initial

cohort took the post-EGRA test. That is, 572 students from the original sample were no longer

present at the time of the post-EGRA. The comparison group had a significantly higher attrition

rate with a 64% rate (324 students) versus a 47% rate (248 students) for the treatment group. The

high attrition rates for comparison students are not surprising, as the activities continued into the

summer months. Students in the comparison group were not receiving treatment; therefore, they

were harder to capture when the time came to administer the final EGRA. This period also was

within a time of great political instability in Egypt which may have led to disruption in students'

participation in the program. Many students also were traveling, working, or on vacation with

their families.

Table 3. Sample Size and Attrition

Gov ID		Stayers	Attrited	Total
1. Qena	N	98	87	185
	Row %	53.0	47.0	100
2. Ismailia	N	135	103	238
	Row %	56.7	43.3	100
3. Cairo	N	42	142	184
	Row %	22.8	77.2	100
4. Minia	N	161	41	202
	Row %	79.7	20.3	100
5. Alexandria	N	8	77	85
	Row %	9.4	90.6	100
6. Gharbeia	N	13	122	135
	Row %	9.6	90.4	100
Total	N	457	572	1,029
	Row %	44.4	55.6	100

A closer inspection of this issue shows that attrition was spread across all the governorates included in the study. Only 23% of Cairo students who took the baseline EGRA test stayed enrolled until the end of the program. These figures are even lower for Alexandria and Gharbeia with only 9% of the original cohort found for evaluation purposes. Although the remaining governorates exhibited lower attrition rates, these rates are still considered high.

Student attrition poses a potential bias if the decision of dropping out from remedial reading and writing activities is correlated with students' potential performance in the program. For instance, if unsuccessful treatment group students are more likely to drop out and not participate in posttesting, then the assumptions underlying the RDD estimation would be violated and there would be a potential selection bias in the impact estimates, causing them to be overestimated. On the other hand, student attrition would not impose a significant threat to the estimated results as long as students drop out of the study due to purely exogenous factors, regardless of their potential performance in the program.

One way to investigate this issue is to compare academic performance at baseline for those who stayed and those who dropped out in terms of their results in both the screening test and the baseline EGRA test. This comparison needs to be done by treatment condition—that is, for treatment and comparison students separately—given that comparison students have higher baseline achievement than treatment students and also have different attrition rates.

Table 4 allows us to investigate this point in detail. For each of the comparison and treatment groups, each entry in the table compares the outcomes between those who stayed until the end of the program (*stayers*) and those who did not leave (*attrited*). We performed a two-tailed hypothesis test of equality of means between stayers and attrited for each of these outcome variables. The results for the screening test, which is the forcing variable for the RDD strategy,

show no difference in baseline academic performance between attrited and stayers. The same results are obtained when comparing the average score for each of the EGRA constructs at baseline, and for two of the student characteristics available: gender and age. Indeed, only three of the 26 tests of mean equality conducted exhibit a statistically significant difference between stayers and attrited.

Table 4. Student Outcomes by Attrition Status

|	Control			Treatment		
	Stayers	Attrited		Stayers	Attrited	
Forcing Variable						
Screening Test Score	47.2	47.4		37.1	37.7	
Baseline EGRA Test						
Letter name knowledge	46.5	44.6	*	41.2	41.2	
Letter sound knowledge	24.6	22.9		16.7	17.1	
Phonemic awareness onset	6.5	6.3		5.7	5.6	
Phonemic awareness ryme	6.2	5.7	*	5.1	5.1	
Familiar word reading	19.5	19.6		10.0	11.7	
Unfamiliar non-word reading	18.2	16.2		10.1	11.1	
Oral reading fluency	33.2	31.9		15.6	19.8	**
Reading Comprehension	1.5	1.4		0.7	0.8	
Listening comprehension	2.2	2.2		1.8	1.7	
Dictation	6.7	6.4		3.9	4.3	
Other Covariates						
Male	0.5	0.5		0.5	0.5	
Age	10.8	10.6		10.7	10.7	
Observations	184	324		273	248	

*= statistically significant at the .05 level, ** at the .01 level (two-tailed test).

The fact that only three of the tests show a significant difference between the two groups indicates that, although attrition rates are high, students who dropped out do not seem to be statistically different in terms of their academic performance at baseline relative to program participants. Consequently, program estimates still can be estimated consistently through an RDD by using the sample of students who stayed enrolled until the end of activities.

An additional check to investigate further the implications of attrition is to compare the distribution for the screening test for all observations at baseline to the distribution that uses only those observations in the final estimation sample, that is, the sample formed by those students

who remained until the end of the program. Although a higher proportion of students from the comparison group (i.e., students with a screening test above 42) dropped out of the program, the screening test distributions—the forcing variable—around the discontinuity for both samples are not only very similar but also continuous around the cutoff point as seen in figure 3. The absence of jumps in the distributions at the cut point provides additional evidence that student attrition is not selective and therefore it does not invalidate the use of an RDD to identify the causal effects of interest.

Figure 3. Estimating Intervention Effects by Using an RDD
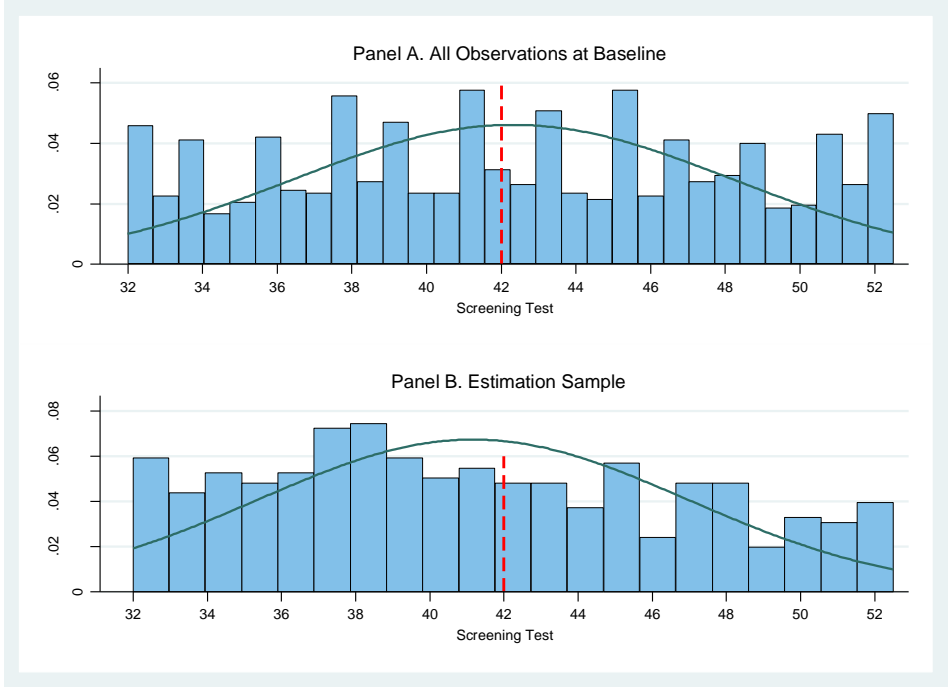
Variance in program administration. Another limitation of the study is that implementation varied from school to school. Because the MOE championed these activities and provided its staff as resources for implementation, we followed their implementation preferences and schedule. This led to differences in how the intervention was conducted from school to school. Although the activities were held during school hours in most schools, the amount of

time for implementation differed. Some schools offered the remedial reading and writing class three days per week for one hour each session (as it was originally intended); others offered it only twice per week for one hour; and still others offered it twice per week for two hours. The ideal implementation schedule was one hour three times per week so that students would not have too much time between the classes to forget what they had learned. However, each school implemented the program according to their needs and abilities as well as the availability of their teachers. Thus, although most students were able to cover the full 36 lessons, some students received more hours of instruction than others.

Findings

Using a modified EGRA reading assessment in Arabic, we estimated intervention impacts on a series of EGRA constructs.[6] Table 5 provides the results for individual EGRA constructs at baseline by treatment status. As expected, comparison group students systematically outperform treated students on all EGRA constructs before remedial reading and writing activities started. That is, while comparison and treated students close to the screening test cut point had similar baseline EGRA results (not shown), the average treated student struggles more on these literacy tasks relative to comparison students, explaining why they were selected for the intervention. The following section will discuss each EGRA outcome measure in more detail.

---

[6] The estimated results presented in this section include observations for all governorates in the study. Results without including students in high-attrition governorates are very similar and are available upon request.

Table 5. Baseline Outcomes by Treatment Condition

	Control			Treatment			
	Mean	SD	Max	Mean	SD	Max	
Letter name knowledge	45.3	8.7	50	41.2	11.6	50	**
Letter sound knowledge	23.5	13.0	50	16.9	11.7	49	**
Phonemic awareness onset	6.4	2.1	10	5.7	2.1	10	**
Phonemic awareness ryme	5.9	2.0	10	5.1	1.9	10	**
Familiar word reading	19.6	13.1	50	10.8	10.9	49	**
Unfamiliar non-word reading	16.9	12.0	50	10.6	10.2	48	**
Oral reading fluency	32.3	20.2	93	17.6	17.1	74	**
Reading Comprehension	1.4	1.4	6	0.7	1.1	5	**
Listening comprehension	2.2	1.5	7	1.8	1.4	7	**
Dictation	6.5	3.8	15	4.1	3.3	14	**
Observations		508			521		

*= statistically significant at the .05 level, ** at the .01 level (two-tailed test).

Letter-name knowledge tests students' ability to identify the letters of the alphabet separately from their sounds. For this subtask, a test assessor presents students with a list of letters in random order. Students then are asked to identify as many letter names as possible in one minute. Letter-name knowledge before formal reading instruction is a strong predictor of children's reading ability. Letter-name knowledge is thought to contribute to reading because it mediates letter-sound knowledge, and because letter-names may be a precursor of or facilitate phonological awareness. As table 5 shows, students selected for activities were able to identify 41 letter names per minute. By comparison, the maximum number of letter names identified by the highest-scoring students in the sample was 50.

The impact-analysis results presented in table 6 show that the remedial reading and writing intervention did not result in a statistically significant impact on this construct. The estimated effect size was a modest 0.11 standard deviation increase for remedial program participants. Thus, in its current form, the intervention is not particularly effective in increasing letter-name knowledge. In all likelihood, this result is partly because many students were already

close to the maximum score attainable, so we should expect smaller program impacts as students move closer to reading proficiency.

Table 6. Summary of Preliminary Level 1 Impacts

	Sample Average	Impact		Effect Size
Letter name knowledge	45.0	0.97		0.11
Letter sound knowledge	26.6	5.12	**	0.39
Phonemic awareness onset	6.6	0.39		0.19
Phonemic awareness ryme	6.0	0.85	**	0.38
Familiar word reading	15.0	1.25		0.11
Unfamiliar non-word reading	16.0	2.51		0.21
Oral reading fluency	21.3	1.6		0.09
Reading Comprehension	0.9	0.38	*	0.31
Listening comprehension	1.1	0.45	*	0.33
Dictation	5.5	1.19	*	0.3

*= statistically significant at the .10 level, ** at the .05 level (two-tailed test).

Note: The effect size is the impact divided by the sample standard deviation of the outcome

The next construct we measured is **letter-sound knowledge**. This construct tests the ability to match a written letter to its oral sound. In this subtask, a test assessor presents students with a list of letters in random order and asks them to identify as many letter sounds as possible in one minute. While the highest scoring students in the sample identified at most 50 letter sounds per minute, participants only identified 14 on average. Moreover, while 90% of treated students identified at most 34 letter sounds per minute, comparison group students identified 41. The remedial reading and writing intervention's impact on letter-sound knowledge was large (an effect size of 0.39) and statistically significant as indicated in table 6. That is, students who were provided with remedial reading and writing activities increased their letter-sound knowledge by 0.39 standard deviations, compared to how these same students would have done in the absence of the intervention.

Phonemic awareness is the ability to hear, identify, and manipulate individual sounds (phonemes) in spoken words. Whereas phonemic-awareness onset tests an ability to identify the

initial sounds of words, phonemic-awareness ryme tests the ability to identify the final sounds of words. In these subtasks, 10 simple, one-syllable words are read aloud (one at a time) and students are asked to identify either the beginning[7] or the ending[8] sound of the word, depending on the construct being evaluated. On average, as shown in table 5, treated students identify almost one fewer onset or ryme phoneme at baseline than comparison group students. The impact results indicate that the intervention had a positive effect on both of these constructs, although only the impact on the ryme subtask was statistically significant with a 0.38 standard deviation gain due to activities. **Familiar word reading** tests students' word recognition and decoding skills and **unfamiliar non-word reading also** tests students' decoding ability. For each of these subtasks, students are presented with a list of words (real or invented) and are asked to read as many words as they can in one minute, with a maximum of 50 words. By reading a list of separate words rather than a paragraph, familiar word reading isolates the skill of word recognition and decoding because students are unable to guess the next word from the context. In contrast, by using an invented word, the unfamiliar word reading construct avoids the issue of sight recognition of words that students have memorized and really tests their decoding skills.

The program effects are positive for both constructs, although the effect size for familiar word reading is rather small, with a 0.11 standard deviation increase as a result of the program. The estimated effect for unfamiliar word reading is larger, with a 0.21 standard deviation gain, although the estimated effect is imprecise. Neither of these effects is statistically significant.

---

[7] The onset is the initial consonant or consonant cluster of the word. For example, in the word bat, "b" is the onset, and in the word swim, "sw" is the onset.
[8] The ryme is the vowel and consonants that follow the initial ryme. For example, in the word bat, "at" is the ryme. In the word swim, "im" is the ryme.

Oral reading fluency is an advanced form of decoding, where the child reads with speed and accuracy but must also read with the correct stress, intonation, and emphasis.[9] The development of oral reading fluency is critical, because even students who are reading with a high degree of accuracy may have trouble understanding the meaning of what they read, if they read too slowly or with poor stress, intonation, and emphasis.[10] In this subtask, students are presented with a short story in the form of a paragraph and asked to read it aloud for one minute; the number of words read is recorded. After reading the story, or as much as they can in one minute, students are asked **reading comprehension** questions about the passage, including two explicit within-sentence comprehension questions, two explicit across-sentence comprehension questions, and two implicit critical-thinking questions.

As shown in table 6, comparison group students scored twice as high as treated students in both subtasks. Note, however, that even comparison group students only reached one-third of the maximum attainable score, which speaks to the general difficulty of these subtasks. The reading intervention did not affect oral reading fluency, for which the impact analysis found a very small and statistically insignificant effect size of 0.09 standard deviations. The remedial reading and writing activities for Level 1 did not focus explicitly on developing oral reading fluency. Instead, the activities focused on the foundational reading skills which promote fluency, and Level 2 of activities will contain more activities focused on promoting reading fluency. Oral reading fluency exhibits the lowest estimated-effect size of remedial reading and writing activities among all the different constructs. In contrast, the intervention had a sizable positive effect of 0.31 standard deviations on reading comprehension.

---

[9] National Institute of Child Health and Human Development 2000; Torgesen, Rashotte, and Alexander 2001.
[10] Daane et al. 2005; Snow et al. 1998.

Listening comprehension tests students' ability to understand spoken language. This skill requires no literacy knowledge. During this subtask, the assessor reads a short passage aloud to the student. The assessor then reads comprehension questions about that passage, and the student must orally respond to the questions. This area is another where this remedial reading and writing intervention appears to have made a meaningful difference. The estimated effect size is a statistically significant 0.33 standard-deviation gain as a result of the intervention.

Last, **dictation** tests students' spelling and oral comprehension skills. In this subtask, the assessor reads aloud a series of words, one at a time; the student must write each word down, paying attention to spelling. This task tests students' ability to hear sounds correctly and then write the letters and words corresponding to the sounds they hear. The remedial reading and writing activities also had a statistically significant and sizable impact on this outcome increasing student performance by 0.3 standard deviations relative to the comparison group.

It is notable that the estimated effects for ESP's remedial reading and writing activities are positive for all constructs. The effect sizes that are statistically significant range from 0.3 to 0.39 standard deviations, which are substantial for a relatively short-term intervention such as this one. Some impact estimates, although positive, were not statistically significant.[11]

The most significant estimated effects were on letter-sound knowledge, phonemic-awareness ryme, and listening comprehension. Since remedial reading and writing activities were aimed at the students with the weakest literacy skills, the program had a heavy focus on the foundational reading skills. In fact, every lesson was built around learning new letters and their corresponding sounds, building from those letters into syllables and words, and breaking words

---

[11] Statistical significance is a function of the sample size, the precision of the outcome measure, and the size of the actual estimated effect. When an estimated effect is not statistically significant, this could be because the EGRA did not measure the component as precisely as others. Differences across the impact estimates presented in this table may not be statistically significant even if some estimates are statistically significant and others are not.

into their component parts. There was also a significant emphasis on listening comprehension. Most lessons included stories, which were read aloud by the teacher and preceded and followed by comprehension activities. Impact estimates were less pronounced in some of the higher level skills, such as oral reading fluency and dictation, which will be more of a focus in Level 2 remedial reading and writing activities; these are expected to begin in late 2013.

Recommendations

On the basis of what we learned from the initial research on Level 1 of the reading and writing activities, we put forth the following recommendations:

- **Formalize the remedial reading and writing activities and continue holding them during activity classes.** Incorporating the remedial reading and writing activities into the regular school day held benefits for students and teachers alike. Teachers were satisfied because they were able to provide extra support to students in need of extra reading and writing help without having to work extra hours outside of their school schedule. Students were also happy not to have to stay after school for extra help, and attendance was probably much greater than it would have been if this were an after-school program. It would further encourage teachers and students if the remedial reading and writing activities were formalized as an official part of teachers' work load.

- **Establish an incentives program for teachers and students.** Teachers and students alike might perform better with incentives to encourage their attendance and adherence to the program. For example, teachers with high levels of attendance might receive some form of recognition or certificate, and the same might be true for students. Students are also motivated by small incentives, such as pencils, notebooks, or other school supplies. Classes that show the most improvement from start to finish might also receive some sort of reward or recognition.

- **Shorten the lessons, or provide more time.** During a qualitative evaluation of the program, teachers revealed that they were often unable to cover all of the lesson content during the hour-long period provided. Often, what was eliminated were the enrichment activities, which were opportunities for students to practice in fun and engaging ways the skills they were learning. These enrichment activities are extremely important—not only to concretize for students the lesson content being taught but also to motivate students to attend and participate, as the activities provide an element of fun.

- **Provide access to more and varied reading materials.** Access to print is a key element in improving reading and writing skills. Students need to be exposed to print materials of all kinds (e.g., stories, fiction and nonfiction texts, magazines, newspapers, flyers) so that they become familiar with a variety of writing types. During Level 1 of the program, we made every effort to include stories in the teachers' guides and student books, but it would be helpful to have other real materials for the students to read and practice with as well. One option might be to hold these classes in the school library (where the materials exist) so that students and teachers have access to books for reading practice.

- **Ensure that teachers receive adequate training and time to practice active-learning activities.** Although teachers were trained in active-learning techniques and were given time to practice teaching lessons during their initial training, once they got into the classroom, they often reverted to a teacher-centered style. It takes time to become comfortable with a different way of teaching and with using groups, peer work, and other classroom configurations to encourage learning. Videos of teachers using these techniques would be helpful to include during the initial training or during the midsemester refresher training.

- **Ensure that teachers are adequately supported by supervisors.** Ideally, supervisors will visit teachers during their remedial classes at least three times during the semester. Supervisors received the same training as teachers, so they know what is expected, and they also were trained in classroom observation and providing constructive feedback and support. However, the reality on the ground was that supervisors often only visited teachers one time during the whole semester and thus were not able to provide much support to the teachers they visited. Ideally, this should be a continuous relationship where they visit, observe, provide feedback and ideas for improvement, and then visit again to see how the teacher is implementing the new ideas.

- **Break Level 1 down into further levels.** Level 1 comprised students scoring from 0 to 42 points on the screening test. This range is quite broad, and students falling at the far ends of the spectrum might have very different abilities. Although both are still in need of remedial reading support, the program could serve them better by creating two separate levels which address the unique needs of students at both ends of the spectrum.

- **Reduce the number of students to a maximum of 20.** Although 20 students was the intended maximum for this program, schools often included more students in each class to provide the remedial reading and writing services to all students in need. However, the teaching and learning would be more enjoyable for all and more beneficial to students if the teacher/student ratio were lower.

Conclusion

The ESP remedial reading and writing intervention directly targets reading needs identified by the Egyptian MOE. The first semester of implementation for these activities witnessed some challenges, such as instability and civil unrest, which have made implementation

inconsistent at times and contributed to high levels of participant attrition between the pre- and posttests. However, in spite of these challenges, EGRA results indicate that the activities had a positive effect on reading skills and provided motivation and support to those students trying to achieve the reading levels of their peers.

Although the sample size was relatively small during this first semester of implementation, we can say with a strong level of confidence that this intervention helped improve students' reading outcomes. Students who completed Level 1 will transition into the second part of the program, Level 2, which provides additional practice in decoding, vocabulary, oral reading fluency, and reading comprehension skills. By the end of this two-semester reading and writing activity, students should show even greater improvement and, we hope, will be caught up to their peers.

References

Daane, M. C., J. R. Campbell, W. S. Grigg, M. J. Goodman, and A. Oranje. 2005. *The nation's report card. Fourth-grade students reading aloud: NAEP 2002 special study of oral reading.* NCES 2006-469. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Gove, A., and A. Wetterberg, eds. 2011. *The early grade reading assessment: Applications and interventions to improve basic literacy.* Research Triangle Park, NC: RTI Press.

Imbens, G., and K. Kalyanaraman. 2011. *Optimal bandwidth choice for the regression discontinuity estimator. Review of Economic Studies.* http://restud.oxfordjournals.org/content/early/2012/01/07/restud.rdr043.

National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services. 2000 April. *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* NIH Pub. No. 00-4769. Washington, DC: Government Printing Office. http://www.nichd.nih.gov/publications/nrp/smallbook.htm.

Snow, C. E., M. S. Burns, and P. Griffin. 1998. *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Torgesen, J. K., C. A. Rashotte, and A. W. Alexander. 2001. Principles of fluency instruction in reading: Relationships with established empirical outcomes. In *Dyslexia, fluency, and the brain,* ed. M. Wolf, 333–356. Parkton, MD: York Press.