This Brochure provides highlights from a synthesis report commissioned by USAID’s Office of Education related to USAID’s Education Strategy Goal 1 “improve reading skills for 100 million children in primary grades.” This review synthesizes the results for 23 evaluations covering 21 activities related to Goal 1.

**Study Objectives**

Review of findings and lessons learned of topics of interest to the Office of Education.

**Topics of Interest**

- **Learning outcomes**
  - Overall results
  - Disaggregated results
  - Subtask results
- **Support for classroom instruction**
  - Time on task
  - Instructional guides
  - Material distribution
  - Language appropriateness
- **Teacher training**
  - Pre-service
  - In-service
  - In-school mentoring and district-level coaching
  - Pedagogy
- **Community engagement**
  - PTAs/SMCs and parental engagement
  - Extracurricular activities
  - Information for parents
- **Policy and systems strengthening**
  - M&E systems
  - Policy and data
  - Capacity building

**Crosscutting Themes** – Gender, disability, information and communications technology, innovative financing, and scaling up/sustainability.

Full report available [here](#).
Inclusion Criteria

As determined by the Office of Education, evaluations to be reviewed in this study were:

- USAID-funded evaluations of education interventions;
- Published between 2013 and 2016;
- Performance and impact evaluations;
- Relevant to the Education Strategy;
- Single, latest published report (in case of reports for multiple phases of an evaluation);
- Evaluation reports from multiple countries (in case of a multicountry education intervention);
- Of acceptable quality based upon minimum evaluation quality criteria set in the synthesis phase.

Sample

<table>
<thead>
<tr>
<th>Region</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>61%</td>
</tr>
<tr>
<td>Asia</td>
<td>26%</td>
</tr>
<tr>
<td>Latin American and the Caribbean</td>
<td>13%</td>
</tr>
<tr>
<td>Middle East</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country Income</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>39%</td>
</tr>
<tr>
<td>Lower-Middle</td>
<td>52%</td>
</tr>
<tr>
<td>Upper-Middle</td>
<td>9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation Type</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact</td>
<td>26%</td>
</tr>
<tr>
<td>Performance quantitative</td>
<td>13%</td>
</tr>
<tr>
<td>Performance qualitative</td>
<td>61%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation Phase</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm</td>
<td>48%</td>
</tr>
<tr>
<td>Final</td>
<td>52%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implementation Stage</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full intervention</td>
<td>74%</td>
</tr>
<tr>
<td>Pilot intervention</td>
<td>26%</td>
</tr>
</tbody>
</table>
Intervention Components

18 Activities
Teacher Training

18 In-service
13 In-school mentoring
8 Pedagogy
6 District-level coaching
6 Pre-service

17 Activities
Support for Classroom Instruction

13 Materials distribution
11 Instructional Guides
10 Quality of Materials
6 Time on task
5 Language Appropriate

15 Activities
Community Engagement

13 PTAs/SMCs
7 Parental engagement
5 Extracurricular activities
3 Information for parents

16 Activities
Policy and Systems

15 Capacity building
9 M&E systems
9 Policy and data
1. Learning Outcomes

Reading interventions in almost all projects had significant, positive effects on student reading scores, though the magnitude of the gains was generally small. Slightly over half of the projects in the Goal 1 sample of evaluations used the Early Grade Reading Assessment (EGRA) to measure student learning, including oral reading fluency (ORF) and reading comprehension. ORF data showed that many projects had statistically significant gains but the student samples were large and the practical effects were mostly small (see below). Most evaluations did not report percentages of students meeting reading benchmarks and targets.

Students generally made more progress in ORF than in reading comprehension, though the comprehension measurements were usually lacking in reliability. As mentioned, the ORF gains were usually significant, but the comprehension gains were more mixed, sometimes with almost no practical effects. Three issues with comprehension measurements were: 1) a maximum of five or six comprehension questions (with one reading passage) on most of the EGRAs, which resulted in a lack of reliability and high measurement error; 2) near universal measurement of comprehension with oral reading but not with silent reading; and 3) focus on ORF in reporting rather than an equal balance between ORF and comprehension.

Projects with the lowest baseline scores tended to have the most difficulty showing improvements, with some difficulty also for projects with the highest scores. Projects with large percentages of low-achieving students at baseline, such as Malawi EGRA and Mozambique ApaL (Aprender a Ler), showed significant ORF gains, i.e., as a percentage of the starting point. However, even with the significant gains, fluency levels remained low, so the practical gains were small. Projects such as Indonesia PRIORITAS (Prioritizing Reform, Innovation, and Opportunities for Reaching Indonesia’s Teachers, Administrators, and Students) and Philippines Basa (Read Philippines) with more high-achieving students were able to achieve gains but the effect sizes were small. Reasons for the lack of large gains by the projects with high-achieving students were not provided in the reports.

Boys generally outperformed girls, though girls often made progress in closing achievement gaps, especially in the early primary grades. Reading outcomes data were disaggregated by gender in almost all of the relevant evaluation reports, and the scores for the boys were usually higher than those of girls, particularly in upper primary grades, e.g., grades 3 and 4. Smaller differences were often found for the scores at the lower primary grades, e.g., grades 1 and 2. In many projects, the girls in the lower grades were able make progress closing gaps in performance, such as on Kyrgyz Republic QRP (Quality Reading Project) and Mozambique ApaL. It appears that projects with interventions supporting girls’ education were generally successful.

2. Support for Classroom Instruction

New reading materials were features of most of the projects, with support for classroom use sometimes hampered by production delays and lack of full distribution to all schools. Projects generally made strong progress in producing and delivering reading textbooks – and sometimes supplementary materials – to the schools as part of their reading programs. However, there were instances of delays in producing materials, including for those projects that needed materials in multiple languages, and in promptly distributing materials to all of the project-supported schools, such as with Ethiopia READ II.
3. Teacher Training

Most projects implemented teacher in-service workshops, though the relationship between the workshops and student learning was unclear. Teacher in-service training was the most popular intervention in the sample of projects. Almost all projects organized periodic in-service workshops. Most of these workshops received positive reviews by the teachers, such as with Indonesia PRIORITAS, though there were some problems with unavailability of materials at the time of the training. Most measurements of teacher behavior change, such as through well-designed classroom observations, were either not conducted (e.g., Ethiopia READ II) or not linked to student learning outcomes (e.g., Indonesia PRIORITAS).

Some teacher coaching and mentoring models showed promise, but most had implementation difficulties due to local staffing, training, and logistical issues. There were challenges with designing and implementing the teacher coaching models to allow for sufficient support in the schools by the district specialists, such as in Dominican Republic ESP (Effective Schools Program), Guatemala REAULA (Reform in the Classroom Project), Malawi EGRA, and Zambia TTL (Time To Learn). Mentoring models were easier to implement, such as with Jamaica BEP (Basic Education Project). Unfortunately, the coaching and mentoring interventions were usually not evaluated except through anecdotal evidence, as in Liberia LTTP II (Teacher Training Program II). An exception was Kyrgyz Republic QRP, which evaluated its coaching model through ratings of teachers in relation to best teaching practices and then associated those ratings with student scores; however, it found no significant improvements in learning outcomes from teacher coaching.

4. Community Engagement

Community engagement models focused on supporting parent teacher associations (PTAs) and school management committees (SMCs), with only minimal effects on student learning. Support by projects to PTAs and SMCs was designed to focus on two areas: 1) school improvement and 2) student learning. The PTAs were more successful in school improvement, with interventions such as small grants to support girls’ education (Djibouti Projet AIDE [International Assistance for Education Development]) and monitoring student and teacher attendance (Ghana Partnership for Accountable Governance in Education (PAGE), Nigeria RARA [Reading and Access Research Activity], and Rwanda L3 [Literacy, Language, and Learning]). These interventions were related to a learning goal, i.e., by expanding time on task with greater attendance, but effects on learning outcomes were not examined and the success of the interventions on student achievement was unclear.

Promotion of reading outside of the school day was implemented on a small number of projects, with some limited results showing positive effects on learning outcomes. Parents were encouraged to support reading activities outside of the classroom, including reading as homework, on several projects. On a small number of projects, such as Kyrgyz Republic QRP, Philippines Basa, and Malawi EGRA, evaluations showed positive effects of community activities in reading and better student learning outcomes. However, on other projects such as Ethiopia IQPEP (Improving Quality of Primary Education Program) and Guatemala REAULA, evaluators did not examine these interventions except through some anecdotal evidence.
**A. Gender**

Gender-related activities were implemented on a substantial number of projects. However, across the evaluation reports reviewed, reporting of gender-related issues was inconsistent and most evaluations mentioned limitations to attribution due to other donor activities also focusing on girls’ education, often in collaboration with USAID implementing partners. Ghana PAGE, for instance, conducted a complete gender analysis, but it was only cited in the evaluation report as an annex, with no discussion of findings in the main body of the report.

**B. Disability**

Interventions in targeting students with disability were only found in a small number of projects. Around three-fourths of the evaluations did not mention disability in the reports. Of the project approaches involving disability, there were some innovations, such as conducting classroom assessments for both gender and disability issues at the same time and then revising materials with the Ministry to increase gender sensitivity and improve services to students with disabilities.

**C. Information and communications technology**

Interventions that used ICT in classrooms were only found in a small number of projects. Several projects – Ethiopia IQPEP, Rwanda L3, and Tanzania Bridge-IT – intervened with IT in classrooms but all reported that much of the equipment stopped working due to lack of maintenance. Conversely, use of cell phones by teachers to access instructional materials (Rwanda L3) and use of tablets by district supervisors to collect reading data (Malawi EGRA) were shown to have strong possibilities, particularly as the apps became more user-friendly.

**D. Innovative finance**

Interventions that explored innovative financing were only found in a small number of projects. Innovative financing was only found on one project, and that was simply an analysis of unit costs associated with reading improvements with no financing initiatives.

**E. Scaling and sustainability**

Capacity-building and scaling up were features of most projects but the resulting ability of countries to sustain and expand successful interventions was often lacking. Nigeria NEI had a clear objective of capacity strengthening, but the evaluation stated that the institutionalization and sustainability of activities was still at an incipient stage at the time of the project. Similarly, Mozambique ApaL planned for sustainability, but there was no evidence in the evaluation report that the implementation took place. An exception was Djibouti Projet AIDE, where capacity building – at the central and decentralized levels – was an explicit part of the project, and the evaluation report stated that the Ministry was expected to be capable of managing the reforms themselves at the end of the project.

**5. Systems and Policy Strengthening**

Support for government monitoring and evaluation (M&E) systems was generally unsuccessful, as were interventions using data in reforming policies around reading. The evaluations found that building M&E systems in support of reading initiatives with ministries of education was rarely done. In Liberia LTTP II, systems were established to monitor reading materials at the school level, but the reporting on the monitoring largely failed. As with the district coaching models described above, some projects found that ministry officials were not able to travel to the field to collect data, such as in the case of Malawi EGRA. In addition to a lack of data, when data were available, there was only minimal data sharing for policy reform and decision making. In Mozambique ApaL and Rwanda L3, extensive data were collected but the Ministry had difficulties in effectively using it.

**Crosscutting**
Limitations

USAID’s success in promoting best practices for improving children’s reading skills under the education strategy allowed for some generalizations in the synthesis of the findings. However, there were the following difficulties in obtaining information on the effects of intervention components from the evaluation studies:

- Since each of the activities in the evaluation reports had combinations of at least two different components of intervention models – i.e., from among classroom instruction, teacher training, community engagement, and policy and systems – the activity designs did not allow for separating out component effects.
- Each component had sub-components – e.g., within classroom instruction, activities had some combination of pedagogy, time on task, and materials – so gathering quantitative evidence on the effects from activities related to sub-components was equally difficult to obtain.
- Evaluations were often able to identify the intended dosage of interventions by activities and collected a variety of information on fidelity of implementation, but the information was not clear or consistent enough to permit analyses relating actual dosage to intervention effects across activities.
- Information was available on student learning outcomes from most activities involving teaching and learning, with consistent measurement methods used for most activities, but it was not possible to capitalize on this information to evaluate the effects of intervention components due to the issues described above.
- Given these activity and evaluation designs, qualitative studies were the only avenues for gathering and analyzing information on the relative benefits of different components, e.g., anecdotal information showed that interventions related to classroom instruction and teacher training were more effective in improving children’s learning than interventions in community engagement and policy and systems.

Large-scale quantitative information on the relative (and more cost-effective) effects of the components of the intervention packages would be highly valuable for programming purposes. The synthesis findings suggest the following recommendations for obtaining additional quantitative information on interventions from evaluation studies:

- Develop better theories of change in activity designs with comprehensive descriptions of how and why a desired change in a particular context is expected to happen, so that evaluations can be designed to determine whether that change has occurred.
- Continue to improve on the application of methods already in place to accurately and consistently measure student learning outcomes, especially grade two students’ reading skills at baseline, midline, and endline at the same time point in the school year, so that outcome measures will be similar across all activities.

Promote comparability of results across activities by using standardized methods for calculating effect sizes as an important step in generalizing results across activities, given the contextualized packages of intervention components and dosages in the field, to make more informed programming decisions.
Recommendations

Ensure that pre- and post-test learning outcomes data are available from all activities and results are presented in a consistent manner across reports.

For reading outcomes, data from the evaluation reports were difficult to synthesize due to a lack of complete reading data. For example, Djibouti Projet AIDE had a baseline but no midline or endline. Nigeria NEI initially planned to measure learning outcomes but experienced delays and then the activity was not able to collect data. Jamaica ETP had some issues with baseline reading data collection and was unable to follow up with corrections. There was also substantial variation in terms of how the evaluations reported their overall results. Philippines Basa and Rwanda L3 provided complete results with raw scores, zero scores, and performance categories, while Mozambique ApaL and Nigeria RARA reported zero scores but no performance category percentages. Some activities reported effect sizes while others only provided statistical significance testing.

Conduct further analyses of learning outcomes, most of which can only be carried out through secondary analyses of the data sets.

While the goal of the synthesis was to examine results from the evaluation reports, some additional analyses took place through the Secondary Analysis and Results Tracking activity, which reported on results achieved by country and for the entire Agency. However, these results, while useful, are limited to descriptive analyses involving ORF. Much more could be done to address topics such as benchmarking and target setting, annual yearly growth, effect sizes, zero scores, performance by gender, fluency in different languages, reading comprehension, correlations between subtasks, and internal consistency reliability. Some of this work has taken place on individual activities, but meta-analyses across multiple activities would be valuable for drawing lessons learned and improving programming in a variety of contexts.

Conduct further studies of differences in reading scores between boys and girls, progress by girls in closing gaps with boys over time, and the reasons behind these issues.

Quantitative studies across activities could include more in-depth analyses of issues such as generally higher reading scores for boys compared to girls, narrowing of gaps between boys and girls (such as on Mozambique ApaL), and instances of girls outperforming boys. Quantitative studies could include issues such as bias in textbooks, preference by teachers for boys over girls, and differences between boys and girls on use of time outside of school. Girls' education programs would comprise an important part of these studies.

Study and prioritize avoiding production delays and distribution problems with teaching and learning materials, especially textbooks.

The extent of the problem of production delays across all activities could be studied, with themes drawn out on the extent of the problem and the reasons for why it happens. For example, it would be expected that activities that need to produce textbooks in multiple languages—such as Ethiopia READ II—would have more production delays than those producing textbooks for a single national or regional language, but this might or might not be the case. Similarly, there appears to be a need to study why the distribution of books is generally more efficient in some countries compared to others. A more detailed examination could be undertaken of how these delays and problems have affected teacher training and classroom instruction, particularly since many interventions are time-sensitive and the duration of the activity support for a set of schools is sometimes only two years.
Recommendations (continued)

Continue with strong emphasis on in-service teacher training workshops but strive for greater application of best practices in terms of duration, frequency, and follow-up.
For the activities that had success with the workshops, such as Indonesia PRIORITAS, as well as those activities that did not have strong success, several issues need to be studied so that lessons learned can be applied to other activities. These issues could include impacts from different doses and durations of training workshops, and the extent of follow-up after the workshops in the schools. There is a general lack of evidence on whether the concepts and skills in the workshops were adequately learned by participants, whether the new techniques were applied in the field, and whether the teachers continued to apply these methods after activity support ended. Other factors related to these issues, such as support from the districts and within the schools, should be studied in depth.

Promote the design and implementation of district-level teacher coaching and school-level mentoring models by building on the limited evidence of success and expanding their implementation to most activities.
Coaching and mentoring models, while not often implemented, have had success in a limited number of activities, such as in Kyrgyz Republic QRP, which offers an example of success for other activities. In some cases, the coaching and mentoring models were not implemented because in part because of the activity design. In other cases, the models were not implemented with fidelity due to issues such as overreliance on cascade training, lack of coordination at the relevant ministry, and activity levels for overseeing the coaches and mentors, overburdened ministry specialists without enough time to conduct field visits, and inadequate funding for logistical necessities such as transportation. Perhaps the most commonly cited problem with these initiatives was a lack of follow-up training and support for the coaches and mentors.

Sharpen community reading initiatives through greater clarity in designs and capitalize on evidence of successful implementation models – including training programs and post-training follow-up – and evaluations.
In addition to the community engagement issues, there were also problems with unclear objectives and lack of a clear relationship—or an indirect relationship—with children’s reading. However, a few of the initiatives—such as Philippines Basa, and Malawi EGRA—promoted community engagement in reading and provided evidence of reading outside of the school day, and improved student learning. More in-depth studies of these activities, with lessons learned, could provide information needed so that the models could be attempted and/or replicated in other countries. Quantitative data provided by relevant ministry and activity staff on numbers of participants affected, materials used, training sessions organized, and visits to the communities would be helpful from a programming perspective. Similarly, it would be important to gather quantitative information to examine the relationship between community reading initiatives and student learning.
Recommendations (continued)

Make greater efforts to support ministries of education in building M&E systems and coordinating with activity M&E systems so that useful indicators are developed, tracked, and used.

In some extreme cases, such as Liberia LTTP 2, the activity team made efforts to boost the Ministry of Education’s M&E system, which was nonfunctional. In the case of other activities—such as on Mozambique ApaL and Rwanda L3—ministry M&E systems existed, but the activities’ efforts did not result in building the capacity of ministry counterparts to use data collected in the field. Without relevant and up-to-date information collection and processing by the government, the possibilities to undertake evidence-based policy making and planning are very limited. One of the issues cited in the evaluation reports was logistics, not unlike the problem with district-level teacher coaching. It should be possible to study the development and implementation of government M&E systems and discuss ways of improving them so that activities are monitored and information is used.

Increase the focus on activities in the crosscutting areas of disability, innovative finance, and information technology.

Almost any additional analyses would add new information to the body of evidence on crosscutting issues, except gender (cited above) which has benefitted from several analyses. Some activities have made efforts, such as Indonesia PRIORITAS, to help increase coverage and the quality of services for students with disabilities. Similarly, Rwanda L3 had an IT component involving interactive radio instruction, but these initiatives have not had proper evaluations. The Tanzania Bridge-IT activity was retrofitted into a reading activity before it reached fruition, though perhaps there are some lessons learned from the IT-related work that took place. Much more could be done, especially through qualitative methods, to identify the reasons behind the relative lack of programming, and the quality of the initiatives, in these areas.

Give more attention to issues of capacity-building, sustainability, and scaling up in evaluations, including post-evaluation studies, and build on validated methods to improve performance.

Unanswered questions include the degree to which capacity building has led to sustainability, or have activities not focused enough on capacity-building so that initiatives will be sustainable after the end of the activity. With scaling up, there is almost no evidence that activities used an internationally validated method (such as MSI’s Scaling Up Management Framework and Scaling Up Toolkit). Additional studies on scaling up across activities—using internationally tested and validated tools as a reference point—would likely provide insights on whether activities are including a scaling up framework from the beginning of the activity. These studies would also help identify where there is room for improvement so that successful initiatives can be designed with scaling in mind, and whether those initiatives can then be expanded for broader coverage and benefit to populations.