Investing in Learning Outcomes: Findings from an Independent Evaluation of the Russia Education Aid for Development (READ) Trust Fund Program

Alan Ruby  
*University of Pennsylvania, alanruby@gse.upenn.edu*

Daniel C. Kent
Investing in Learning Outcomes: Findings from an Independent Evaluation of the Russia Education Aid for Development (READ) Trust Fund Program

Keywords
education, student assessment systems

Disciplines
Education | Educational Assessment, Evaluation, and Research

This report is available at ScholarlyCommons: https://repository.upenn.edu/gse_pubs/512
INVESTING IN LEARNING OUTCOMES

Findings from an Independent Evaluation of the Russia Education Aid for Development (READ) Trust Fund Program.

Alan Ruby and Daniel C. Kent

Prepared for the World Bank, December 2015
This report was prepared by Alan Ruby, Senior Scholar, and Daniel C. Kent, graduate assistant, of the Alliance for Higher Education and Democracy (AHEAD) at the University of Pennsylvania’s Graduate School of Education. AHEAD is a group of scholars, researchers, and practitioners who share a commitment to “advancing higher education policy and practice that fosters open, equitable, and democratic societies” (http://ahead-penn.org). The authors wish to thank the representatives of the Russian Government and current and former World Bank officials for their willing participation in the review of Russia’s significant investment in strengthening learning outcomes. The final report has benefited from constructive feedback and helpful comments and corrections from both the Russian authorities and the World Bank. We particularly wish to acknowledge the inputs of Emily Gardner who prepared the final report for the World Bank and Marguerite Clarke whose technical expertise in assessment and leadership were referenced by many we interviewed and were present in all aspects of the review.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>4</td>
</tr>
<tr>
<td>WHAT IS READ?</td>
<td>9</td>
</tr>
<tr>
<td>- OVERVIEW</td>
<td>9</td>
</tr>
<tr>
<td>- BASIC ORGANIZATION</td>
<td>9</td>
</tr>
<tr>
<td>- OPERATIONAL GOALS</td>
<td>10</td>
</tr>
<tr>
<td>- WHY EVALUATE READ?</td>
<td>10</td>
</tr>
<tr>
<td>NATURE OF THE EVALUATION</td>
<td>11</td>
</tr>
<tr>
<td>- OUR TASK</td>
<td>11</td>
</tr>
<tr>
<td>- REVIEW PROCESS</td>
<td>12</td>
</tr>
<tr>
<td>- DOCUMENT REVIEW</td>
<td>14</td>
</tr>
<tr>
<td>- LIMITATIONS</td>
<td>14</td>
</tr>
<tr>
<td>HOW MUCH MONEY WENT WHERE?</td>
<td>15</td>
</tr>
<tr>
<td>- GLOBAL-LEVEL ALLOCATIONS</td>
<td>15</td>
</tr>
<tr>
<td>- COUNTRY ALLOCATIONS</td>
<td>15</td>
</tr>
<tr>
<td>WHAT DID READ ACHIEVE AT THE GLOBAL LEVEL</td>
<td>16</td>
</tr>
<tr>
<td>- METHODS</td>
<td>16</td>
</tr>
<tr>
<td>- ACTIVITY TYPES AT THE GLOBAL LEVEL</td>
<td>17</td>
</tr>
<tr>
<td>- TRAINING CONSTITUENCY AT THE GLOBAL LEVEL</td>
<td>18</td>
</tr>
<tr>
<td>SCOPE AND DEPTH OF GLOBAL KNOWLEDGE PRODUCTS.</td>
<td>19</td>
</tr>
<tr>
<td>WHAT DID READ ACHIEVE AT THE COUNTRY LEVEL?</td>
<td>21</td>
</tr>
<tr>
<td>- METHODS</td>
<td>21</td>
</tr>
<tr>
<td>- ACTIVITY TYPES AT THE COUNTRY LEVEL</td>
<td>21</td>
</tr>
<tr>
<td>- TRAINING CONSTITUENCY AT THE COUNTRY LEVEL</td>
<td>22</td>
</tr>
<tr>
<td>HOW TO APPRAISE READ’S ACHIEVEMENTS?</td>
<td>26</td>
</tr>
<tr>
<td>- FIELD BUILDING OR FIELD SUPPORTING?</td>
<td>28</td>
</tr>
<tr>
<td>- ACTIVITY PURPOSE AT THE COUNTRY LEVEL</td>
<td>29</td>
</tr>
<tr>
<td>WHAT DID READ ACHIEVE FOR THE DONOR?</td>
<td>30</td>
</tr>
<tr>
<td>WHAT DID READ ACHIEVE FOR THE WORLD BANK?</td>
<td>30</td>
</tr>
<tr>
<td>WHAT DID READ ACHIEVE FOR THE FIELD?</td>
<td>32</td>
</tr>
<tr>
<td>WHAT WERE READ’S STRENGTHS &amp; WEAKNESSES?</td>
<td>34</td>
</tr>
<tr>
<td>IS THERE A NEED FOR READ 2?</td>
<td>38</td>
</tr>
<tr>
<td>SYNTHESIS</td>
<td>41</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>42</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>55</td>
</tr>
</tbody>
</table>
Executive Summary

The goal of the Russia Education Aid for Development (READ) Trust Fund program was to increase educational assessment capacity in developing countries. Funded by the Russian Government and overseen by a council made up of Russian Government appointees and World Bank officials, it operated from 2008 to 2015. The READ Council asked the World Bank to commission an independent review of the Trust Fund’s impact to identify strengths and weaknesses in design and implementation. The findings would be used to inform the development of other donor-led trust funds and to guide the design of any future iteration of the READ program.

The Trust Fund’s emphasis on the quality of education and focus on learning outcomes makes it an ideal candidate for evaluation. In addition to capturing knowledge about what worked and what was difficult, this evaluation is also an opportunity to document the experience of a large institution working with and adapting to a new, significant donor with distinct needs, priorities, and norms of practice.

The review explores eight broad themes: 1) Global Achievements of the READ Program; 2) Country-Level Achievements of the READ Program; 3) Donor-Side Achievements of the READ Program; 4) World Bank-Side Achievements of the READ Program; 5) Strengths of the READ Program; 6) Weaknesses of the READ Program; 7) Opportunities for More Global Work on Learning Outcomes; and 8) the Shape and Design of READ 2. It does not cover READ Reimbursable Advisory Services, which supported Russia’s Center for International Cooperation in Education Development (CICED).

In its final form, the READ Trust Fund held US$32 million. Just over a third of the fund was allocated to global activities and nearly 60 percent to country-level programs and projects; the balance was the World Bank’s standard trust fund management fee. The global allocations supported five main streams of work: new knowledge products about student assessment systems and techniques; disseminating this knowledge and country experiences to a global audience; building and participating in partnerships with other donors and development agencies; strengthening connections and networks in the assessment field by convening the READ country teams and other assessment practitioners and experts; and overseeing and monitoring the implementation of projects at the global and country levels.

The two most striking global achievements were a substantial addition to the knowledge base on assessment and the closer integration of the agencies, national officials, and academic experts working in the field of educational assessment. READ “created a lot of value through global public
goods”, deepening understanding of national and international assessment techniques and sharing instances of good practice. The SABER (Systems Approach for Better Education Results)-Student Assessment frameworks are a significant technological advance in how to diagnose the strengths and weaknesses of a national assessment system. The knowledge products produced under READ cover different assessment traditions, different institutional approaches to quality assurance, and different governance structures. The READ global products were “Russian funded and READ branded, but the quality is World Bank assured.”

READ initially focused on seven countries: Angola, Ethiopia, Kyrgyz Republic, Mozambique, Tajikistan, Vietnam, and Zambia. In 2011, Armenia was invited to join the program at the request of the donor. The eight target countries received US$18.7 million over the life of the program. The individual country allocations were based on national action plans and ranged from just over US$1 million in Ethiopia and Armenia to a little more than US$4 million in Tajikistan.

The scope of READ-financed activities at the country level was wide and deep, addressing topics as diverse as early grade reading in Mozambique and Angola to exit examinations and school inspections in Ethiopia. All eight nations invested significantly in developing local expertise in assessment using short courses, degree programs, study tours, and workshops covering topics like item banks, test integrity, formative assessment, and interpreting national and international survey results.

At the start of the READ program, the World Bank team used key indicators from the SABER-Student Assessment framework to establish baseline values for each target country in the four main domains of a national assessment system. Repeating the exercise in 2014 allowed the team to determine the progress each country had made in these key indicator areas and domains. In all, six of the initial seven READ countries made progress on at least one of the main domains. The exception, Tajikistan, showed progress on some of the sub-indicators and is well placed to continue to develop. Armenia, a later addition to the READ target group, had less time to gain momentum.

While the eight target nations benefited substantially from READ, its activities also reached many other nations. For example, the World Bank team used some of the READ-financed tools to benchmark countries in East Asia and the Middle East and North Africa. In all, 51 countries were involved in READ activities or were the subject of READ case studies or reports.

One way to capture what READ produced is to view its outputs through a “field-building” versus “field-supporting” framework. This language from the philanthropic community makes a distinction between interventions which support a range of activities and organizations in a strategic or innovative fashion and activities which take a proven technique to a wider audience.
Field-building activities tend to address a specific opportunity in a domain; create new knowledge or technologies; invest in actions, networks, and agencies that apply new policies or practices; develop and adopt the right standards; build up infrastructure; and share knowledge.

Field-supporting activities scale up an existing process or program, reaching more people or sites. They tend to be activities that are or will become the ongoing responsibility of an agency or organization.

More than half of READ’s global activities were oriented towards field building while three quarters of the country-level activities were field supporting.

The field-building work is exemplified by the SABER-Student Assessment framework, the country case studies, the investment in PISA for Development, and the alliances and partnerships developed through READ.

In addition to its contributions to the field of educational assessment and to education in the target countries, READ’s successes can be viewed in terms of what it achieved for Russia and for the World Bank.

For Russia, the donor, READ established its reputation as a global leader committed to the improvement of education in developing countries. It also established Russia as a source of expertise in the field of educational assessment and proved that Russia was committed to working collaboratively with the World Bank to improve educational achievement, equity, and quality across the world.

READ allowed the World Bank to leverage its world-class assessment expertise by creating accessible materials in the domain, including a global template for diagnosing assessment systems. READ’s activities and products shifted the focus of the development community towards issues of quality and learning outcomes, equipped World Bank operations staff to lead policy discussions with country counterparts, and established a basis for new investments in educational quality.

Some of READ’s successes come from the way it was designed and implemented. READ focused on a specific field, assessment of learning outcomes. It concentrated its resources on a small number of countries and on a suite of knowledge products. As a result, at both the country and global levels, work programs were relatively well resourced. This enabled countries to carry out substantial initiatives.

Overall, READ was purposefully and thoughtfully planned. The program channeled a new international donor’s funds into an area where they could have some observable impact in the short term and lay the basis for sustained improvement. The READ Council was also flexible,
responding positively to proposals from countries that took different approaches to building better assessment systems.

The design and implementation of READ also struck an appropriate balance between creating global public goods (the READ knowledge products) and supporting country-specific activities that were grounded in a good appreciation of the needs and priorities of a particular nation.

The tight geographic focus of the country-level activities was in part a product of the donor predetermining the beneficiaries. The selection of the target countries by the donor concerned many World Bank staff who felt that it created a sense of entitlement rather than an incentive to improve or innovate. Many of them preferred a competitive funding process where countries applied for support. Other criticisms of the program include its support for large-scale training and the limited opportunities for cross-national exchange of ideas and approaches in the initial years.

These design issues from READ 1 – the pre-determined target nations and sense of entitlement; the balance between innovative, experimental, and reform activities and large-scale, recurrent training; and national insularity in the initial planning phases should be addressed in the design of subsequent trust funds. At the heart of these concerns is the selection of eligible countries. There is a difference between a donor-determined closed set of target countries and a “donor informed” list of eligibility criteria for applicants. The first has the problems of entitlement and reduced motivation. The second raises country expectations and tends to diffuse resources and diminish donor commitment. The middle ground may be in determining a list of eligible countries with the active engagement of the donor and inviting applications in a form akin to the action plans from READ 1. Eligible countries might request through a “letter of intent” a modest base allocation to underwrite the preparation of a national improvement plan that is assessed competitively by a central team, with allocation decisions made by the READ Council.

There is still much to be done in the broad area of learning outcomes globally and within the eight countries in the initial READ program. At the global level, the international development community has reaffirmed the importance of learning outcomes as a cornerstone of quality education. The “Incheon Declaration” and the associated framework for action stressed the importance of stronger evaluation and measurement processes. This is aligned with the Sustainable Development Goal of ensuring “inclusive and equitable quality education” and promoting lifelong learning for all (United Nations, 2015).

At the cross-national level, the international trend surveys continue to be powerful tools for mobilizing resources and political interest in better quality learning for all. PISA and TIMSS are
useful comparative tools that constitute “public goods” by making data about learning outcomes openly accessible and useable by all.

There is much to be gained from using the SABER-Student Assessment framework to guide and monitor development in the area of student assessment systems and to identify areas where good practices should be documented and disseminated. The continued investment in the creation of knowledge products that illuminate different ways of measuring learning outcomes to guide resource allocation and appraise strategies is a worthwhile strategic investment. A lot of the knowledge products created through READ concentrated on national and international large-scale assessments and produced admirable studies of different practices. There is a need for a similar investment in the field of high-stakes student examinations where READ 2 could “create a critical mass of products” that could inform practice and policy at the national and provincial levels and support a global conversation about integrity, fairness, and transparency in the design, conduct, and uses of exit and completion examinations. These knowledge products should also examine how educational leaders can use and interpret data from different types of assessments to inform policy and practice and to communicate to stakeholders.

“The global public goods element is an essential part of READ. The products gave shape and coherence” to many country activities and underpinned the successful international partnerships.
What is READ?

Overview
In October of 2008, the Government of Russia and the World Bank agreed to collaborate, for an initial period of five years, on the Russia Education Aid for Development (READ) Program to facilitate Russia's effort to expand its role as an emerging donor, and to focus that effort on the education sector. The original conception was to pursue two high-level objectives:

(1) To assist countries with well-designed plans for improving access to basic education to also improve education quality and learning outcomes; and

(2) To build Russia’s capabilities as a provider of robust and reliable technical assistance on education issues to low-income countries.

The first objective would be addressed by establishing “the READ Trust Fund,” a single-donor trust fund in the World Bank to finance analytical work, build capacity, and fund countries’ efforts to measure and improve learning outcomes.

These activities were aligned with the World Bank’s education strategy established in 2005, which included an emphasis on “more results-oriented” projects and products that maintained the quality of educational services as participation in primary and secondary schooling increased under “Education for All” initiatives.

This linkage ensured that the READ activities would be aligned with the multi-donor partnership, “The Education for All - Fast-Track Initiative” (FTI), which was supposed to help countries make rapid progress towards the Millennium Development Goal of universal primary school completion by 2015. To complement countries' efforts to increase enrollment and successful school completion, READ funds would enable the World Bank to deliver research and analytical services, technical assistance, training, and materials development which would support the “design, implementation, and evaluation of interventions to improve education outcomes in developing countries.” As well as aligning with global education priorities and with the World Bank sector strategy, the READ program was connected to the Russia/World Bank Country Partnership strategy, which envisaged co-operative endeavors that would help establish Russia’s Official Development Assistance program.

READ constituted a multi-year, multi-million-dollar endeavor. It was envisioned as a five-year, US$50 million commitment that would be relatively tightly focused to ensure that there was a greater chance of real, observable impact.

Basic Organization
The funds for the READ program were provided by the Russian Government. The Trust Fund program was overseen by the READ Council, which was made up of Russian and World Bank
leaders. The Council set broad strategic directions for the program and established the leadership of the program. The program was implemented at the global level by World Bank central staff who worked on many of the global initiatives discussed in this report. In the target countries, programs were overseen by World Bank education specialists and operations officers who worked closely with country officials and staff. The evaluation team was not directly involved in the implementation of the program.

**Operational Goals**
The READ program had different goals and activities at the global and country levels.

At the global level, the primary goals of the program were to contribute to the overall knowledge of countries in the field of educational assessment through the development and dissemination of tools, analytical reports, and case studies on ideal assessment practices and strategy.

At the country level the goal was to address key gaps in assessment capacities that were identified in countries’ initial action plans. These plans and activities varied significantly as countries were at different stages along the continuum of assessment development.

Seven countries were initially selected to participate in the READ Trust Fund program. These countries were Angola, Ethiopia, Kyrgyz Republic, Mozambique, Tajikistan, Vietnam, and Zambia. In 2011, Armenia was invited to join the READ Trust Fund group of countries at the request of the donor. Although Armenia was added late in the program, it was still able to access significant technical services and financial resources to build up its assessment capacity and expertise. Armenia also had immediate access to the suite of global knowledge products.

Each country was assigned at least one World Bank Task Team Leader for the execution of the funds. These staff worked closely with national ministry of education officials to determine needs and priorities, understand the local context, and identify appropriate interventions or activities to build up existing capacity and improve assessment quality. Most of the funds were disbursed through procedures requiring direct authorization by World Bank officials, making READ a predominantly “Bank-executed” rather than recipient- or client-executed program.

**Why Evaluate READ?**
The scale and length of the commitment is sufficient justification for an external review, but a more compelling reason is the emphasis on the quality of education and an attendant focus on learning outcomes. This emphasis was an important addition to the global focus on the goal of universal basic education, which rightly invested resources in increasing access and participation. READ’s attention to quality and measurement of student learning presaged future development goals of high-quality educational outcomes and greater attention to evidence-based policy.
It is equally important to document the experience of a large institution working with and adapting to a new, significant donor with its own needs, priorities, and norms of practice. What lessons can be learnt for future relations between the Government of Russia as an active participant in Official Development Assistance and development agencies like the World Bank? Might the experiences of both parties be informative for other emerging donors, both governmental and private?

Finally, it is of value to observe and learn from how this well-conceived and thoughtfully-designed program was implemented, noting how it evolved and developed and what was more successful and what was problematic. This would inform the development community in general, but would also offer guidance to the donor and the World Bank on the merits of further investments in this area through a similar mechanism.

**Nature of the Evaluation**

**Our Task**
With these rationales in mind, we were asked to review the READ program, document its strengths and weaknesses, and offer advice on the need for further work in the area of learning outcomes and the shape and direction of a possible second phase of the READ program. The review explores eight broad themes:

- Global Achievements of the READ Program;
- Country-Level Achievements of the READ Program;
- Donor-Side Achievements of the READ Program;
- World Bank-Side Achievements of the READ Program;
- Strengths of the READ Program;
- Weaknesses of the READ Program;
- Opportunities for More Global Work on Learning Outcomes; and
- Shape and Design of READ 2.

The *Terms of Reference* for the work is in Appendix 1. The review does not cover READ Reimbursable Advisory Services, which supported Russia’s Center for International Cooperation in Education Development (CICED) and offered opportunities for Russian nationals to share their expertise in assessment and quality management. There was some interconnection between the Advisory Services and the READ projects supported through the World Bank-administered Trust Fund, particularly in the use of two assessment tools; one addressing technology use by secondary students and one that assesses mathematics, science, and language in primary schools. The funds for the development of these tools were administered separately and fall outside the scope of this survey.
Review Process

Development projects of this kind are not amenable to quantitative analysis, and causality between investment and verifiable learning outcomes is difficult to establish. The time scale of the READ Trust Fund is relatively short, especially in a policy environment where design and implementation takes years. In some cases, the full effects of a change in assessment technology can only be discerned after a full cycle of schooling has been completed. Even then, the performance of a single cohort is not enough to determine unequivocal success or failure.

Aware of these limitations, the review relied principally on qualitative methods, collecting data and observations through interviews and document review. Where possible, individual’s experiences and opinions were triangulated with the views of others and with formal records. The work plan for the review is attached as Appendix 2.

As with any project involving a number of countries with differing capacities in data collection and presentation, it is difficult to standardize data across countries. This issue is especially pertinent when working with dispersed teams and countries with different priorities and traditions. For example, there are at least four different approaches to the assessment of learning embedded in the educational history of the eight READ nations. The Russian scientific and positivist approach shapes practice in Armenia, Kyrgyzstan, and Tajikistan. The Russian traditions have influenced assessment in Vietnam, which also draws on Chinese and French traditions. Examination and inspection regimes are legacies of British colonial rule in Zambia. Mozambique inherited a weak education system from Portugal, with significant regional and gender imbalances, which catered for a minority of the population (Eduardo, 2012, 26). Assessment of learning was fragmented and inspection systems needed strengthening. Angola had a similar legacy from its years of Portuguese rule; low participation, limited infrastructure, and no national examination after any level of education (World Bank, 2013, 4). Ethiopia’s education system was initially influenced by the French model and French was the language of instruction up to 1935. The Italian occupation (1935-1941) had little impact on the structure of the school system, but the British model was important in the first decades of national independence, with English becoming the mode of instruction and English approaches shaping the evaluation system (Bishaw & Lasser, 2012).

Conscious of these differences in context and traditions, the interviews conducted during the review were semi-structured. A list of general themes or questions to be explored in interviews was generated from the Terms of Reference, and further informed by a review of the key program documents and the prior experience of the lead reviewer. This was discussed with the World Bank’s team leader, and then refined and finalized. These questions and themes (see Box 1) were shared with potential participants prior to interview or email exchange, with the proviso that other issues were likely to be explored and that participants could raise any topic of interest.
Box 1: List of Questions and Themes

1. What worked well in the READ program?
2. What did not and why not?
3. What gaps have been identified in the program coverage that should be addressed in future READ-funded work?
4. What processes and procedures should be refined to increase effectiveness?
5. How do the participant countries perceive the READ work and how have they integrated it into ongoing activities?

The reviewers formally interviewed 20 people in total over a three-month period. Interviews were conducted in-person at the Washington, D.C. offices of the World Bank, remotely via phone, and via email. These included members of the READ Council, members of the Global Program team, Country-Level Program team members, and other experts.

There are certain analytical benefits to interviewing an administratively diverse group of people related to a program. Current and former members of the READ Council provided strategic, transversal perspectives across the program and covered its inception, execution, and completion. They also offered insights into the initial organizational strategies of the program and pointed to the different motivations and goals of the various actors in the development and delivery of the READ program.

The Global Program team members provided insight into the design and structure of the cross-national activities and the challenges of coordinating a program that crosses different regions and deals with countries in different parts of the World Bank’s organizational structure. They also contributed to an understanding of how the program differed within the eight target countries, including differences in timing, initial entry points, and client relationships. There was an unusual level of staffing continuity or stability among the Global Program staff in the later years of the program, which enabled them to act as an institutional memory.

Country-level teams dealt most closely with the implementation of the programs, and shared highly-specific knowledge about the effectiveness and outcomes of the program at the country level and at the institutional and sector levels. Some program staff members were engaged with only one nation while others had experience in two READ target nations. Some had relatively short engagements while others initiated the first READ activities in a nation and saw the program
through its first cycle. These differences in intensity and length of engagement enrich the qualitative data while adding complexity to the task of analysis.

A full list of those interviewed can be found in Appendix 3.

**Document Review**

Various documents were reviewed as part of the formative aspect of this evaluation. This yielded a seemingly objective viewpoint on official READ Trust Fund outcomes. The principal documents ranged from initial concept notes and a cross section of knowledge products to successive READ Annual Reports. These reports provided a comprehensive listing of the tangible activities conducted during the course of the program. Other documents reviewed included the READ Trust Fund Administrative Agreement; the READ Trust Fund Concept Note and Results Framework; similar Trust Fund program documents separate from the READ program; a benchmarking assessment created by Julia Lieberman; a report on “Strengthening Education Quality in East Asia” produced by the World Bank and UNESCO; and the Statement by the head of the delegation of the Russian Federation at The Third International Conference on Financing for Development, Deputy Finance Minister S. Storchak.

**Limitations**

While the analysis presented in this report is independent, it is not without its shortcomings. These shortcomings are related to a reliance on qualitative data drawn in large measure from participants in the READ program who might be expected to see only the successes of their work. This potential for respondent bias is tempered by the professionalism of the participants and by the reviewers’ assurance that respondents would not be identified with particular observations or critical comments.

In addition, there were inconsistencies in the type and quality of data collected across the different countries and different activities. These differences are apparent both at a global level and at the country level. As a result, it is difficult to make quantifiable statements about the outcomes of READ that go beyond overall trends. For example, a metric that would have been highly useful as a proxy for measuring program reach is the number of people trained through READ funds. However, the data available on this topic varied greatly across countries and READ Annual Reports. This makes it difficult to reliably compare training outputs across countries. Another limitation in this area is likely differences across countries in training quality. It is difficult to say that one specific instance of training would be analogous to training in another country since standards for such activities are highly varied. An obvious illustration is the differences between training that led to the award of Master’s degrees and a short intensive course, tailored to needs of a particular agency, which did not result in a formal credential.
These differences in part reflect the country-specific customization of many of the READ-funded activities. The benefits of crafting activities based on individualized needs analyses of the assessment systems in nations with different challenges and priorities can limit ready comparisons and constrain the accuracy of summary reviews of different types of activities like the number of people trained in student assessment issues. Nonetheless, we have developed an approach to aggregate and report on the scope and reach of READ-funded activities, which we present below.

**How Much Money Went Where?**

In its final form, the READ Trust Fund held US$32 million. The final financial commitment to global activities was US$11.6 million (36.4 percent) and the country-level programs and projects had a total financial commitment of US$18.7 million (58.5 percent), with the balance of US$1.6 million (5.1 percent) being applied to the World Bank’s management fee.

**Global-Level Allocations**
The global allocations supported four main streams of work: new knowledge products about student assessment systems and techniques, disseminating this knowledge and country experiences to a global audience, building and participating in partnerships with other donors and development agencies, and convening the READ country teams and other assessment practitioners and experts to strengthen connections and networks in the assessment field.

The global allocation was split between SABER framework development costs (US$5.4 million) and knowledge sharing and program coordination (US$6.2 million).

**Country Allocations**
Individual country allocations were based on action plans, which countries submitted to the Bank early in the program. Allocations ranged from just over US$1 million in Ethiopia and Armenia to a little more than US$4 million in Tajikistan. In all cases, most of the allocation was disbursed by the time the program was completed. There were some adjustments between initial and final allocations as needs changed and as some countries were slower to disburse. These adjustments indicate prudent management and sound fiscal oversight and also show a welcome element of flexibility in the implementation of the Trust Fund.
These global-level and country-specific allocations do not show the diverse array of activities supported through the READ program or the detail of disbursements by year or final expenditures by country or global program. More detailed financial reporting and commentaries illustrating the diversity of country activities are set out in the final report of the READ team. Here our interest is in understanding the nature of the activities that were financed and the geographies and topics covered to get a sense of what was attempted and achieved.

What Did READ Achieve at the Global Level

While a significant percentage of the money allotted to the global program was allocated for certain objectives, it is important to analyze where this money ended up in terms of the different types of activities that were funded. What follows is an analysis of the READ-funded global activities.

**Methods**
All of the activities listed in the READ Annual Reports were appraised independently by the two reviewers and categorized by type. The two categorizations were compared and any differences were discussed and resolved.
The categorization is independent of the magnitude of expenditure as the countries vary significantly in scale and size. Rather, it attempts to capture the array of discrete activities or accomplishments. The categorization does not necessarily reflect the time, effort, or funding put into a specific area.

The activities were categorized using the following definitions:

- **Training, Workshops, and Conferences**: Activities in which information was disseminated in person. These activities are broken down later in the report by constituency.

- **Materials Development**: Activities that created discrete tools, frameworks, usable data, and knowledge that could inform assessment practice. Rather than overall studies and “lessons learned” types of analysis, these are usable tools that can be implemented.

- **Analysis**: Includes activities such as country and regional reports and empirical research. This category covers the development of knowledge in the field of educational assessment as a whole, rather than the creation of any specific frameworks or tools.

- **Partnerships and Convening**: Activities that were sponsored or initiated through the creation of a partnership with a non-READ organization.

- **Implementation and Execution**: Activities in which assessment systems and examinations were implemented as a result of READ-based work.

**Activity Types at the Global Level**

This meta-analysis of activities revealed a number of interesting insights into the workflow of the READ program at the global level.

As can be seen in Chart 2, activities were not uniformly distributed across the categories. The largest category is analytical activities. These made up nearly one third of all READ activities at the global level. The next biggest category is materials development, 29 percent of all activities. These findings are in line with the stated global-level goals of the READ program. The basic analytical work, which was provided at the global level, underpins the development of a global dialogue about educational assessment. This intellectual work adds to the body of knowledge and the global pool of expertise on student assessment. These increases are of strategic importance as demand for these skills and a deeper knowledge base is increasing as attention is turning to evaluating reforms and policies in terms of learning outcomes.

The materials development work was also strategically important; it increased the set of internationally-validated tools available to country policymakers. This increase in the repertoire of materials suitable for use in a range of contexts and assessment regimes is one of the more significant outcomes of the READ program and contributed to the overall success of the program.
The third biggest category of READ activities globally was training, workshops, and conferences. This category comprised a fourth of the total number of completed activities. This is a direct result of the program’s goal of building up educational assessment capacity and expertise globally and legitimizing it as a field that can lift educational quality.

Chart 2: Activity Type at the Global Level

![Activity Type at the Global Level](chart2.png)

Training Constituency at the Global Level
Although training made up a relatively small portion of activities at the global level, it is useful to analyze who was being trained at this level, as demonstrated in Chart 3. All of the training activities listed in the READ Annual Reports were appraised independently by the two reviewers and categorized by type. The two categorizations were compared and any differences discussed and resolved. The categorization does not capture where the training took place; nor does it capture the quality of the training delivered.

There are three discernible approaches to READ-funded training activities.

- **Large-Scale Training**: general efforts to either disseminate information, spread best practices, or facilitate trainings that involve large groups of practitioners.
• **Specialized Technical Expertise**: trainings for small numbers of individuals who need to acquire distinct skills necessary for the successful conduct of assessment activities in their home countries.

• **Policymakers and Opinion Formers**: trainings geared towards elite policymakers and other stakeholders to foster the use of learning outcomes data to improve education quality.

Large-scale training activities and specialized technical expertise trainings make up 92 percent of all training funded under READ at the global level. Only 8 percent of activities were solely for policymakers and opinion formers. Practitioner-focused training was by far the most common activity; aimed at equipping local people with assessment skills as well deepening understanding of the importance of assessment as a way to lift educational quality. Many of the global READ events served multiple constituencies and in these cases, they were simply labeled as large-scale training.

**Chart 3: Training Constituency at the Global Level**

![Chart showing the distribution of training constituencies at the global level.](image)

**Scope and Depth of Global Knowledge Products.**
The global activities funded by the READ Trust Fund reached well past the eight target countries and covered many aspects of assessment. The geographic sweep of activities covered countries at all stages of economic and educational development. For example, the World Bank team used READ-financed tools to benchmark countries in East Asia and the Middle East and North Africa. These activities also involved UNESCO and ALESCO, further extending the scope of READ’s activities. This work is captured in two regional publications covering thirteen economies in Asia and seventeen in the Middle East and North Africa.
The World Bank team also studied and produced county reports on 29 nations, in addition to the core group of eight. These country studies included nations from Africa, Central Asia, and the Pacific.

To capture the lessons of practice across a range of assessment traditions, the World Bank team also commissioned thirteen case studies of distinctive elements of assessment policy and practice in eleven nations ranging in size from under 10 million people in New Zealand and Singapore to over 100 million people in Brazil and the Russian Federation.

Regional and global meetings also engaged a broader group of countries to further build up assessment expertise and quality. In all, 51 countries were involved in READ activities or were the subject of READ case studies or reports. A visual illustration of the scope is given in Figure 1.

Figure 1: Countries Participating in SABER-Student Assessment Activities (as of 6/30/2015)

The “Global knowledge products,” covered a range of policy and technical issues. The country case studies, for example, were commissioned to deepen understanding of how nations have purposefully gone about the complex task of improving their assessment systems. One set of studies examined what sets of policies and processes are likely to create more effective systems and another set looked at how to use assessment information to improve instructional design. The studies draw on a range of different assessment practices and strategies in an array of contexts, ensuring that the READ countries had access to a diverse set of good practices.
Collectively, the global knowledge products are a significant net addition to the knowledge base on assessment. This was a recurring theme in our interviews. READ “created a lot of value through global public goods,” deepening understanding of different approaches to national and international assessment techniques, and documenting and sharing instances of good practice. The SABER-Student Assessment framework is widely regarded as a useful tool for diagnosing the state of a national assessment system. The “generalizability and reliability of the knowledge products” are enhanced by the breadth of countries, assessment traditions, approaches to quality assurance, and governance structures they cover. The READ global products were “Russian funded and READ branded, but the quality is World Bank assured”.

What Did READ Achieve at the Country Level?

Methods
Activities at the country level were categorized using the same process that was applied to the global activities. All of the activities listed in the READ Annual Reports were appraised independently by the two reviewers and categorized by activity type and training constituency. The two categorizations were compared and any differences discussed and resolved.

Again the categorization is independent of scale of expenditure as the countries vary significantly in size and available resources. Rather, it attempts to capture the array of discrete activities or accomplishments. The categorization does not necessarily reflect the time, effort, or funding put into a specific area.

Activity Types at the Country Level
At the country level, READ activities were focused on practitioners (see Chart 4). Trainings, workshops, and conferences made up the bulk of READ-financed work at the country level. This is as expected given the hands-on nature of country-level activities. Also to be expected, materials development, analysis, and partnerships and convening activities were less prevalent.
Training Constituency at the Country Level

The same definitions and categorization process used to show the distribution of global training activities were applied to the country-level training activities. This revealed a slightly different pattern (see Chart 5). While large-scale and specialized training still made up the majority of country-level activities, 29 percent of activities were directed at policymakers and stakeholders. This possibly reflects the need for more broadly-based policy dialogue within countries to build a wider base of support for change in assessment practice.

Chart 5: Training Constituency at the Country Level
The categorization of the different activity types and the distribution of training opportunities across different populations are good descriptive indicators of inputs and processes that might contribute to improvements in a national assessment system. To gauge the general direction of country-specific developments, we looked at the movement of each country along the continua in the SABER-Student Assessment Framework. The World Bank team used the key indicators in the framework to make baseline determinations of where each target country was on the four main domains of an assessment system when the READ activities were being initiated.

They repeated the exercise in 2014 to determine the extent of development of these countries’ assessment systems over the course of the READ program. These changes should be interpreted with care and a realistic sense of how much can be achieved in this field in a short time. Establishing a high-performing, reliable assessment system takes years. The best-performing systems are based on years of experience and are underpinned by significant financial and personnel resources. They have usually been in place and operating in a stable policy and resource environment for multiple cohorts of students so that benefits and changes are discernible and enduring. The second caveat is the complexity of the field; there are multiple variables or processes that need attention before significant change is realized. The SABER-Student Assessment framework seeks to account for complexity by breaking key stages into definable, discrete tasks. This allows progress to be acknowledged within a stage even when action is not complete. With these caveats noted, the final report of the READ Trust Fund captures the country-level changes as shown in the summary table below. In all, six of the initial seven READ countries made significant progress on at least one domain. The exception, Tajikistan, showed progress on some of the indicators and is well placed to continue to develop. Armenia, as a later addition to the READ target group, had less time to gain momentum and was also already operating a relatively well-established assessment system.

Two nations, Vietnam and Ethiopia, made progress on more than one domain; both strengthening classroom assessment and one of the large-scale assessment domains. Vietnam also made progress with its examination system, an area where Ethiopia was already well established.

The Tajikistan team built an ambitious program, blending READ money with funds from the Open Society institute, USAID, Education for All, and a World Bank credit to support a “whole reform agenda that included building a secure testing facility. It was a coordinated approach, which leveraged the READ funds to great effect” to create a better university entrance examination and improve the school curriculum.
Notably, Kyrgyzstan elected to step back from participation in International large-scale assessments, a decision that is understandable given resource constraints and the efforts required in developing other aspects of its student assessment system.

Overall, the changes made in a short time constitute a very substantial achievement. Compared to the baseline survey, six of seven cases made measurable improvement against clearly-defined criteria with a relatively modest but necessary financial investment.

The fact that the SABER-Student Assessment framework was able to track progress and capture movement in both directions is testament to the quality and utility of the framework itself. It is a good approximation of the steps or actions that constitute the evolution of an assessment system. It is a considerable intellectual achievement in its own right and a valuable outcome of the READ Trust Fund program. The framework can be easily applied to national and sub-national systems in countries at all stages of economic and social development, increasing its significance still further.
Table 1: Summary of the Benchmarking Results for Student Assessment Activities in READ

<table>
<thead>
<tr>
<th></th>
<th>Classroom Assessment</th>
<th>Examinations</th>
<th>NLSA</th>
<th>ILSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angola</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Emerging</td>
</tr>
<tr>
<td>Armenia</td>
<td>Established</td>
<td>Established</td>
<td>Established</td>
<td>Established</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Latent</td>
<td>Emerging</td>
<td>Established</td>
<td>Established</td>
</tr>
<tr>
<td>Mozambique</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Established</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Emerging</td>
</tr>
<tr>
<td>Vietnam</td>
<td>Emerging</td>
<td>Established</td>
<td>Emerging</td>
<td>Established</td>
</tr>
<tr>
<td>Zambia</td>
<td>Emerging</td>
<td>Emerging</td>
<td>Established</td>
<td>Established</td>
</tr>
</tbody>
</table>
Illustrative Country-Level Activities
While the benchmarking results summarize the system-level changes in the various student assessment programs, they do not offer a picture of the variety of activities developed and implemented at the country level. These are described in appropriate detail in the final report from the World Bank READ team. Interviews with country team members highlighted some of the national achievements. These ranged from piloting new assessment exercises to building local technical capacity and training large numbers of classroom teachers.

These brief observations do not capture the detailed design and development that went into the country activities. For example, in Mozambique, the Ministry of Education developed a reading comprehension test that teachers could administer at the start and end of the academic year. It was aligned with the national curriculum and designed to show progress over time. The strategy helped create robust pedagogical programs by giving teachers a basis for setting annual goals, identifying strategies for various learning levels, and grouping students within grades. End-of-year results were shared with parents and the wider community, increasing transparency and strengthening accountability. After a successful pilot in 68 schools, the strategy was rolled out in two stages to cover over 200 schools in five provinces. READ supported the design of the intervention, the development of materials, and the pilot phases. It also funded the evaluation of the pilots and joined with the Strategic Impact Evaluation Fund to assess the impact on student learning using a control group of schools and a quarter of the schools that participated in the pilot project.

Rather than recount all the details of countries’ achievements reported in the various READ Annual Reports, brief vignettes of these country-level activities, drawn from the various documents and our interviews, are included as Appendix 4. Some of the vignettes also include illustrative quotes from interviewees.

How to Appraise READ’s Achievements?
It is difficult to assess innovative development strategies like the READ program. READ and similar programs are not amenable to large randomized controlled trials or even to simpler measures of those served or supported. The goals can sometimes be underspecified, making them difficult to measure. One way to capture what the READ Trust Fund has tangibly produced is to examine its
outputs through a “field-building” versus “field-supporting” framework. This language comes from the philanthropic community where it has been used to describe efforts that go past funding individual projects or interventions and instead look to support a range of activities and organizations in a more strategic fashion. A popular example is the standardization of U.S. medical education in the early 1900’s, which followed from the financial support of the Flexner report. Philanthropies see a number of benefits in “field building.” It tends to reduce duplication and inefficiencies while drawing attention to a particular issue or legitimating efforts to improve a situation. Ideally, it creates new knowledge, encourages the exchange of information between domains, and provides incentives for collaboration. By its very conception, it tends to involve multiple actors or agencies and takes a holistic approach to problem solving.

Drawing on the writings of Bernholz, Seal, & Wang (2009), we define field building as those actions that meet many or all of these characteristics:

1. They recognize and address a specific gap or development opportunity in a domain;
2. They establish a research base, creating new knowledge or advancing thinking in that domain;
3. They prioritize sets of actions and networks that apply or lay the basis for application of new policies or practices;
4. They develop and adopt the right standards;
5. They build a network infrastructure; and
6. They share knowledge.

These steps do not have any particular order and each serves a unique purpose, but one of the common first steps in field building is the development of common terms and standards. This, for example, has been true of various attempts to improve initial teacher education and the certification of teachers.

Similarly, field-supporting activities are those that scale up an existing process or program by, for example, increasing frequency of action or reaching more people or sites. They tend to be activities that are or will become the ongoing responsibility of an agency or organization.

Our assessment reveals that much of the READ global-level activity is field building. This is not surprising given the theory of change implicit in the original conception of the Trust Fund and in how it was expressed in the SABER-Student Assessment Framework, with its field-building continuum; a four-step process of latent, emerging, established, and advanced. The first three phases are essentially field-building phases. It is not until the processes and policies are in place, supported by sufficient, appropriately-trained people and repeated routinely that we can move
into field-supporting activities. These might include spreading expertise to provincial and local levels, training additional cadres of experts, or reviewing and updating existing programs.

Much of the READ-funded work at the global level was oriented towards field building while many of the country-level activities, especially in the more established schools systems, were field supporting. This is corroborated by a codification of the various READ activities.

Field Building or Field Supporting?
All of the activities that were categorized by type were further categorized as either field building or field supporting, using the same process of independent coding, comparisons, and reconciliation and the definitions set out above.

It is not surprising then that a majority of READ activities were related to field building (Chart 6). This is one of the trademarks of a high-impact, non-profit program. Activities that are field building add intellectual capacity and technical capabilities that will have a longer-term impact, reaching people and systems beyond those directly involved in the project. Field building also strengthens the foundations underpinning new approaches to persistent problems; in this case, it added credence and credibility to a focus on learning outcomes as a way of assessing educational quality. It broadened the repertoire of development assistance professionals, allowing them to intervene in the processes of teaching and learning and in the design and calibration of assessment strategies.

The READ Trust fund was a “knowledge builder,” investing in the development of a diagnostic framework which could be applied to many different national assessment systems.
Activity Purpose at the Country Level
The activities at the country level skewed towards field-supporting. Whereas more than half of the activities related to field-building at the global level, only a quarter were related to field-building at the country level. This shows the different nature of activities at the country level. Seventy-five percent of the activities were field-supporting. Clearly, the goals of the READ Council for global activities to be focused on knowledge creation and dissemination were carried through all the way down to the country level.

Chart 7: Activity Purpose at the Country Level
The lens of field building also lets us look at what has still to be done to make assessment a more robust part of the education systems in many developing economies. Field building also underscores the need for sustained investment – the length of time it takes to see the results of this sort of intervention.

**What Did READ Achieve for the Donor?**

While READ was certainly a program that accomplished the goals it was established to achieve in relation to strengthening developing countries’ educational assessment systems, it also produced a number of benefits for the donor. Firstly, it established the donor as a global leader committed to the improvement of education in developing countries. This is evidenced by the continued monitoring and interaction between the senior leadership of the READ Council and the rest of the program team.

It further established the donor as an expert in the field of international educational assessment. As a key and necessary part of the program, the donor expended resources in developing this specialty in order to disseminate it to member countries through the fee-based services component of READ (which is not included in this review). It drew on the body of expertise and specialist knowledge that Russian academicians and experts have in the area of test development and assessment and, on occasion, used materials and testing strategies developed by these Russian experts that looked at different skill domains. It also linked those Russian technical experts to the wider international assessment community, enriching both sets of professionals.

Lastly, it proved that Russia was committed to working collaboratively with the World Bank to improve educational achievement, equity, and quality across the world.

**What Did READ Achieve for the World Bank?**

READ financed important knowledge-creating activities that contributed to the intellectual foundations of the World Bank’s wide-reaching SABER initiative. SABER now covers over 100 countries and helps national decision makers and advisers benchmark their education policies and evaluate key policy choices.

Three specific themes stand out in the READ-funded global activities. First, and most enduring, is the creation of a world-class set of knowledge products about identifying and evaluating policy
choices in the assessment of learning outcomes. These products cover assessment of the individual student, assessment to inform and improve teaching practices, and national and international assessment strategies and opportunities. Looked at as a set, the READ publications are a very substantial and significant contribution to the field of student assessment.

The most important contribution has been creating and codifying a repertoire of accessible terms and concepts for the evaluation of national assessment regimes. It has developed a common language for talking about the systematic appraisal of learning outcomes. This has facilitated the exchange of ideas and experiences among nations, which is the second distinctive theme of the global activities funded by READ. The eight recipient nations learnt from each other and shared their experience more broadly. READ activities generated data on fifty-nine countries, there are 29 national studies and two important regional surveys.

This intellectual productivity laid the basis for knowledge sharing with donor groups, development agencies, and professional associations of education assessment experts.

The global knowledge products also had a practical outcome. They equipped the World Bank’s education specialists with an array of tools and with a level of confidence to initiate a policy dialogue about student assessment with their country counterparts. The World Bank staff was also given renewed confidence that they could sustain a technically-substantive relationship in this area and support important development work in the field. In the relatively short life of READ, this led to new relationships with Ministries of Education in some countries in sub-Saharan Africa and to new lending and credit opportunities that addressed learning outcomes. These constitute net additions to resources applied to lifting the quality of education for thousands of young people.

The World Bank’s stewardship of the READ Trust Fund illustrated that the World Bank is able to be a good intellectual partner and reliable fiscal agent for a “new donor.” Its staff made sure the program activities were well focused and of the highest quality. The funds were prudently and appropriately managed and when necessary re-allocated to meet new country priorities or to support those nations that were moving ahead more quickly than others. The Trust Fund closed with less than US$ 20,000 unexpended.

At the country level, READ investments offered the World Bank the opportunity to open a policy dialogue with the Ministry of Education in Zambia, developing a relationship that been dormant for some time. READ investments also generated new lending and credits for the education sector in Angola and saw assessment components included in Ethiopia’s General Education Quality Improvement Project (2014-2018). In both cases, the goal was to bring new resources to improve learning outcomes.

In summary, the READ Trust Fund allowed the World Bank to:
1. Leverage its limited supply of world-class assessment expertise by creating accessible materials in the domain, including a global template for diagnosing the strengths and weaknesses of national and provincial assessment systems, by developing training programs, and by marshalling an international network of assessment experts;

2. Shift the focus of the development community towards issues of quality and learning outcomes;

3. Diversify the donor community interested in global education;

4. Equip World Bank operations staff with skills and products for more effective policy dialogue with country counterparts about measuring learning outcomes; and

5. Establish a basis for new lending operations and, in at least one case, opened a new relationship with the education ministry.

What Did READ Achieve for the Field?
The field-building activities conducted with support from the READ Trust Fund were a significant contribution to the intellectual architecture in the discipline or science of educational assessment. As we noted above, the substantial corpus of knowledge products from READ was shared with many actors and agencies in the field of educational development and assessment of learning outcomes. The World Bank team was very active in disseminating reports and publications as the basis for knowledge sharing with other donors, multilateral and national development agencies, and technical experts. Some of this knowledge management was embedded in the existing global partnerships and alliances and these were strengthened by relatively modest allocations of READ funds to support meetings and to convene technical and policy experts. In addition to modest financial contributions, the rich body of analytical materials and case studies as well as the expertise and international reputation of key World Bank READ team members made READ a welcome partner and participant in regional and global forums and networks.

READ contributed to large-scale, cross-national assessment exercises, such as PISA for Development, providing technical expertise to the project’s steering group, and funding country participation in technical workshops as well as Zambia’s overall participation in the pilot survey. READ also commissioned a background paper on the cumulative fifteen years of PISA activities and another paper on ways of aligning different international assessments to maximize their utility for developing nations.
The READ Trust Fund supported and co-sponsored important international assessment forums with the Global Partnership for Education, the US Agency for International Development, the UK Department for International Development, and UNESCO.

World Bank officers involved in READ also developed and maintained effective collaborative relationships with staff at the OECD, at UNESCO’s Institute for Statistics, and with assessment experts at Brooking’s Center for Universal Education. They also contributed to important regional forums in East Asia and Africa.

These partnerships, sponsorships, and collaborative endeavors added to READ’s visibility in the international aid community and its importance in the field of educational assessment. It added to the reach and depth of READ’s contribution to the field and further legitimated the strategic investment in the global knowledge products. The involvement of officials from the READ target countries in some of these activities also helped maintain reform impetus at the county level as well as inform the design and development of national activities.

READ contributed to an improved policy dialogue about the importance of assessing learning outcomes in the global development community. This is reflected in the work leading up to the newly-adopted Sustainable Development Goals. It also fostered policy dialogue at the national and regional levels about evidence-based decision-making and the value of a more systematic approach to measuring and comparing.

Further, READ generated greater investments in the development of professionals in the field. These took place both at the global knowledge leader-level and at the country practitioner-level as READ financed training opportunities for national officials working directly on assessment practices and for teachers grappling with notions of formative assessment. These investments strengthened the technical capacity in the target nations.
What Were READ’s Strengths & Weaknesses?

THE READ Trust Fund program had important strengths. Three stand out for simplicity and importance. First, READ focused on a specific field, assessment of learning outcomes. This helped target financial resources from READ, the country, and other donors, and direct intellectual capital and time at factors instrumental in improving student achievement. Second, much of READ’s resources were concentrated on a small number of countries. This created a reasonable span of control for the small number of central staff who were managing and coordinating the program. It also ensured that the available financial resources were not too widely and thinly disbursed. Similarly, it enabled the few World Bank experts in this area to focus their expertise on a set of countries and on the creation of global products. Third, both country and global-level work programs were relatively well resourced. This enabled countries to undertake substantial initiatives, like conducting a large-scale Early Grade Reading Assessment exercise in Mozambique that when proven could be built into the national assessment program financed by the government. But the total funds available both globally and at a county level were still constrained, which ensured that discipline and rigor were involved in deciding what to fund.

READ had other strengths. It was purposefully and thoughtfully planned. The initial design took into account existing international development and aid initiatives and targeted an area that needed strengthening in many countries as they worked towards reaching Millennium Development goals. The program design channeled a new international donor’s funds into an area where they could have some observable impact in the short term and lay the basis for sustained improvement. The READ Council was also flexible, responding positively to proposals from countries that took different approaches to building better assessment systems. It supported large-scale short courses for teachers as well as intensive formal degree programs for technical experts. It supported pilot projects as well as infrastructure and equipment. The unifying theme was better assessment of learning outcomes.

READ is also unusual in its support for global public goods that serve broad and somewhat diffuse audiences as well as financing practical activities that help practitioners. The design and implementation struck an appropriate balance between creating global public goods, the READ knowledge products, and supporting country context-specific activities that were grounded in a good appreciation of the needs and priorities of a particular nation.

“READ activities provided a focal point so that other donors can channel their money to greater impact, knowing that the intervention had been well designed and would be effectively monitored”.
The early identification of target countries was seen by some as an advantage. It allowed for a quick start in planning and implementing activities in those countries where there was already an effective relationship between the education ministry and the World Bank. This welcome speed also carried some costs. When some nations are moving ahead quickly, the task of developing and coordinating appropriate and consistent monitoring procedures can be overlooked or left for later stages. When that happens, it can produce uneven data sets because information on, for example, training participants was not collected at the time an activity took place. The monitoring and data collection processes in READ were not ideal, especially in the early stages of implementation. Subsequent efforts to establish good baseline data and collect information on training participants overcame some of these shortcomings. The existence of the robust SABER-Student Assessment framework should assist future efforts to quickly establish a baseline diagnosis of a nation’s assessment system and to ensure that data collection commences with the first funded activities.

The pre-selection of the target countries by the donor concerned many World Bank staff. It was an issue raised in most interviews. Some expressed the view that it created a sense of entitlement to resources in the national Ministry of Education. This impeded serious dialogue about what changes needed to be made with the application of READ funds. The apparent promise of funds, implied by being identified as a READ target country, acted as a disincentive to search for and identify actions or strategies that would strengthen the national assessment system. A number of World Bank education specialists suggested there should be a more competitive approach to distributing READ resources to countries as this would produce better plans with more highly-committed local partners. Others argued that the donor’s intention in nominating eligible recipients was akin to other donors identifying particular themes, like clean water or communicable diseases, which should be addressed through a particular trust fund.

This issue is more complex than that. The donor, in choosing countries to receive READ funding, is attempting to ensure that the influence of the program will maximize its geo-political impact. Countries are selected by the donor using a process that balances variables such as the donor’s strategic interests, its history of economic and diplomatic relations with the other country, and some cultural affinity. As a newer member of the international development assistance community, it may also have been seeking to ensure that the visibility of its efforts was safeguarded by concentrating on old friends and allies.
It is highly likely that the donor will continue to select a list of “eligible countries”, or at least a majority of them, that can seek support from any future Russian-financed trust fund for education. It is not likely to adopt an open application process, making the funds available on a competitive basis to any developing country as this would either dilute the value of support to a level where it would have little country-level impact or leave many applicants unfunded.

The fact that READ was almost completely a “World Bank executed” trust fund where commitments and disbursements required specific approvals by a World Bank official contributed to the tension around the sense of entitlement. Some World Bank operational staff also caviled about the delays and impracticability of “World Bank execution” arrangements for activities like short training courses in remote locations with few vendors. Others saw direct involvement in approvals as a way of ensuring the appropriate quality of materials and timeliness of events. This is a longstanding area of debate about ideal modes of implementation in development and is not particular to READ. In the context where it is the first Trust Fund for a new donor targeted at countries with a range of managerial capabilities, “World Bank execution” seems prudent.

While there was a sense amongst some country team members that READ was established and became operational relatively quickly, aided in part by the early identification of target countries and the tight thematic focus on assessment, some of the World Bank’s managers and representatives of the Russian Government felt that READ had a slow start. This perception comes in part from the time it took for the parties to develop and agree on the shape and structure of the Trust Fund and to sign the necessary legal and financial instruments.

The complexity of the READ program, the involvement of countries in three different regions, the need to establish operational procedures and to determine initial national and global financial allocations all contributed to the gradual start of the READ program. While a faster start would have been welcomed by all parties, there are reasonable causes for the actual rate of implementation. With procedure and protocols established the start of a second READ program should be smoother and a little faster.

There were also some staffing discontinuities in the early years of the World Bank’s administration of READ and there was some staff turnover in the READ country teams. These are hard to avoid in large organizations where personnel are reassigned to meet changing circumstances and where career advancement is often dependent on mobility. That aside, there were noticeable increases in the quantity and quality of global activities once the leadership was
stabilized and a core team of people with solid assessment expertise was in place. The delays in determining the future of a second cycle of READ will see some of this expertise dispersed to other immediate projects.

On the donor side, participation from some key players was sporadic although other senior figures were clearly very committed and applied great time and energy to steering READ’s activities to a fruitful end for all parties. It also took time for both parties to adapt to differences in operating style, and to recognize that shifts in priorities and changes in the wider geo-political context would influence the program’s operations. These were largely resolved by time and experience. Nonetheless, the Russian Government’s representatives commented that the flow of information between the World Bank and the donor could be improved in both quality and quantity with “more detailed information about the implementation process”.

A few people raised some implementation concerns that are related to the design of READ. One or two people observed that training activities dominated a lot of the country-level activities. **Chart 4** above shows that in terms of frequency, training was nearly half of all activities at the country level. Some countries did offer many training events and aimed to reach significant numbers of classroom teachers, especially when the express purpose was to promote formative assessment at the classroom level. If training seems to be consuming too much of a Trust Fund’s resources, limiting opportunities for innovation, piloting, and evaluation of initiatives, the proportion of funds for some types of activities could be “capped.” Alternatively, if there is a competitive or semi-competitive process, the selection criteria can be weighted to favor development and experimentation rather than recurrent or large-scale training, which would be more appropriately seen as a national budget item. Another design constraint was that the initial seven target countries developed national “action plans all at the same time so they didn’t get to learn about” innovative or creative elements in other plans until the global READ conferences, by which time plans had been approved, expectations established, and resources committed. For example, a number of country teams admired Mozambique’s investment in Master’s degrees for a small number of assessment specialists rather than numerous short courses and workshops that did not result in a formal, internationally-recognized credential. Any subsequent cycle of READ investments could ameliorate this insularity by sharing the action plans from “READ 1” and by investing a little more in communication across national teams.

These design issues, the pre-determined target nations and sense of entitlement; the balance between innovative, experimental, and reform activities and large-scale, recurrent training; and national insularity in the initial planning phases should be addressed in the design of subsequent trust funds. At the heart of these concerns is the selection of eligible countries. There is a difference between a donor-determined closed set of target countries and a “donor informed” list of eligibility criteria for applicants. The first has the problems of entitlement and reduced...
motivation. The second raises country expectation and tends to diffuse resources and diminish donor commitment. The middle ground may be in determining a list of eligible countries with the active engagement of the donor and inviting applications in a form akin to the action plans developed to commence READ. Eligible countries might request through a “letter of intent” a modest base allocation to underwrite the preparation of a national improvement plan that is assessed competitively by a central team, with allocation decisions made by the READ Council.

Is There a Need for READ 2?
There is still much to be done in the broad area of learning outcomes globally and in the eight countries in the initial READ program. At the global level, the international development community has reaffirmed the importance of learning outcomes as a cornerstone of quality education. The “Incheon Declaration” and the associated Framework for Action stressed the importance of stronger evaluation and measurement processes: “We commit to quality education and to improving learning outcomes, which requires strengthening inputs, processes and evaluation of outcomes and mechanisms to measure progress” (Item 9 Education 2020).

This is aligned with the Sustainable Development Goal of ensuring “inclusive and equitable quality education” and promoting lifelong learning for all (United Nations, 2015). The Education 2030 Framework (paragraphs 97-103) sets out the importance of “effective monitoring and accountability mechanisms” in the systematic pursuit of Sustainable Development Goal 4. Recurring themes in the Framework for Action include the importance of “evidence based policies,” of “quality frameworks in national education plans,” and the “centrality of teaching and learning quality.” All are embedded in the work of READ to date and in the SABER framework. They form a solid foundation for further investments in developing better national practices and creating more global knowledge products that continue to advance our understanding of what works in national and international assessment systems. These shifts in the broad architecture of international development assistance reinforce the centrality of good assessment systems, policies, and practices in the pursuit of better learning outcomes and high-quality education for all.

At the cross-national level, the international longitudinal surveys continue to be powerful tools for mobilizing resources and political interest in better-quality learning for all. PISA and TIMSS are useful comparative tools that constitute “public goods” by making data about learning outcomes openly accessible and useable by all (Wagner 2012). But the high entry costs of
international studies can exclude less-well-developed nations and limit involvement in the design and calibration of assessment instruments. This, in turn, orients the content covered by such assessments towards the concerns and norms of more advanced economies. This reinforces the importance of READ investments in the work of international assessments to ensure that the needs of emerging nations are addressed (Carton & Jakovleski, 2015). The OECD’s education director is conscious of the need for PISA to evolve its methods and surveys to “cater for a larger and more diverse set of countries” if PISA is to be a relevant “global yardstick for measuring success in education” (Schleicher, A & Costin, C., 2015). Adjusting the PISA instruments, revising the context surveys, and piloting the modified tools in a cross-section of middle- and lower-income economies requires intellectual and financial resources, some of which could come from Russian experts and a second READ Trust fund.

While these types of international comparative studies are important in providing reference points and aligning national performance with the performance of others, they do not in themselves point to desirable changes in policy or practice. To identify and design effective intervention strategies, they need to be supplemented by more finely-grained and more sharply-focused case studies and analytical work. As Fang and Gopinathan (2009: 569) observe, “the research frameworks guiding large-scale international studies cannot reveal the elements that lead to differences in student performance. More fine-grained discourse analysis...is needed.” This argues for continued investments in illustrative case studies of good assessment practices in developed and developing nations as well as continued engagement with the agencies and partnerships involved in international and regional comparative studies of student performance.

There is much to be gained from using the SABER-Student Assessment Framework to guide and monitor development in the area of student assessment systems and to identify areas where good practices should be documented and disseminated. The continued investment in the creation of knowledge products that illuminate different ways of measuring learning outcomes to guide resource allocation and appraise different strategies is a worthwhile strategic investment. A lot of the knowledge products created through READ concentrated on national and international assessments and produced admirable studies of different practices. There is a need for a similar investment in the field of school examinations where READ 2 could “create a critical mass of products” that could inform practice and policy at the national and provincial levels and support a global conversation about integrity, fairness, and transparency in the design,
conduct, and uses of exit and completion examinations. These knowledge products should also examine how educational leaders can use and interpret data from different types of assessments to inform policy and practice and to communicate to parents and the public about the learning outcomes of different levels and modes of schooling.

At the country level, there is still much to be done, especially in those initial READ countries that had “latent” assessment systems. There are two views about the direction of future activities. Some favor a “future-oriented” approach, with READ 2 supporting innovation and experimentation, acting as an “incubator for new ideas” and evaluating pilot projects. READ 2 in this model is a “dedicated budget for innovation.” Others favor a more linear approach, with READ 2 supporting the roll out or scaling up of practices and policies that have been developed with support from READ 1. In the case of the Kyrgyz Republic, this would mean a suite of “activities to support the spread of formative assessment techniques, capacity building for staff at the National Testing Center, modelling improved instructional practices, participation in an international comparative survey like PISA or TIMSS or even possibly PIRLS.” The aim would be to “maximize the return on investment from both the initial READ Trust Fund and round two.”

The two approaches are not mutually exclusive and both reflect different approaches adopted by different ministries and agencies in READ 1. Some, like Mozambique, piloted new approaches. Others, like Tajikistan, invested in more comprehensive, integrated strategies which took longer to come to fruition. Both have merits and the choice of appropriate approach is shaped by capacity, local priorities, and need and by the presence or absence of other donors and the policies being pursued by those donors. Given the value many actors found in READ’s flexibility, both approaches should be able to attract support from READ 2.

The geographic reach of READ 2 was a recurring theme in interviews. The most forceful comment was that the eligible or target counties should not be confined to Eastern Europe and Central Asia. A few people noted that the inclusion of Francophone nations might further enrich the conceptual work and provide new opportunities for innovation and experimentation. One of the benefits of READ 1, only partly realized, was that the involvement of countries from very different assessment traditions increased opportunities for cross-national learning. Another is the wider visibility and recognition of the donor’s commitment to improving education. These are valuable and should be continued.

Similarly, the scope and reach of READ at the global level has come in considerable measure from the breadth of countries at all levels of economic development and with various forms of governance covered by the knowledge products and included in the SABER database. This is most clearly illustrated by the field-building characteristics of READ’s public goods, the suite of intellectual materials produced, validated, and disseminated in a few years. In short, there are
very real benefits in terms of effectiveness, visibility, and impact from breadth and engagement at the country and global levels in programs like READ.

**Synthesis**
The READ Trust Fund was a surprisingly good investment; “it surpassed expectations.” It built up the field of educational assessment and laid the foundation for a more rigorous approach to learning outcomes in the international development assistance community. One recent expression of the importance of measuring learning outcomes is in the latest United Nation’s Sustainable Development Framework. Another is the OECD’s PISA for Development project.

READ supported a range of activities in eight targeted countries, with discernible positive effects in most cases. This is well illustrated by the progress in the development of their national assessment systems, captured in the SABER-Student Assessment framework.

The activities and materials supported by READ investments reached well past the eight target nations. There were more than 50 countries directly involved.

The global educational assessment community benefited from READ activities, from the creation and dissemination of new knowledge to widening and strengthening of alliances and professional partnerships. This is illustrated by the involvement of various national aid agencies, international agencies, cross-national organizations, and some private donors in READ-related activities that focus on measuring and improving learning outcomes.

The donor nation benefited by showcasing its unique capabilities and techniques in assessment, by displaying its willingness to be a member of good standing and repute in the international development community, and building closer links between its own experts and the international assessment community.

The World Bank benefited by showing that it is a good intellectual partner and reliable fiscal agent for a “new donor as well increasing its investments in learning outcomes and supporting its operations staff in policy dialogue around monitoring student achievement to inform quality improvement.”
Appendices

Appendix 1: Terms of Reference

TERMS OF REFERENCE

Consultancy to Conduct Final Evaluation of the Russia Education Aid for Development (READ) Trust Fund Program

BACKGROUND

Established in October 2008, the Russia Education Aid for Development (READ) program is a collaboration of the Government of the Russian Federation and the World Bank that focuses on improving education quality in low-income countries. The READ Trust Fund is part of this program, with an amount of US$32 million to be executed over a seven-year period, 2008 to 2015.

The World Bank’s Education Sector Strategy 2020 highlights the importance of “Learning for All.” The proven economic gains and poverty reduction tied to education are only obtainable when children actually learn. All actors in an education system need to know whether or not learning is taking place so that they can use this information to improve education quality.

The READ Trust Fund’s main purpose has been to help low-income countries improve their student learning outcomes through the design, implementation, and use of robust systems for student assessment. It has supported analytical work and technical assistance to help countries:

- establish systems or institutions—or strengthen existing ones—that formulate learning goals and carry out assessments of student learning;
- improve existing or develop new instruments to measure student learning outcomes; and
- strengthen existing or develop new mechanisms (policies) to use learning outcomes data to improve teaching and learning.

The main outcome of the READ Trust Fund is expected to be increased institutional capacity of countries to develop, carry out, and use data from student assessments to improve education quality and student learning. Armed with information on how well students are performing, teachers, policy makers, and international donors alike will be better able to determine where to focus their energy and resources for the greatest improvement in learning outcomes.

The READ Trust Fund has been operating under the guidance of the READ Council, a group of key Russian and World Bank officials (Annex 3). This group has provided strategic direction for the work carried out at the global and country levels. At the global level, the focus has been primarily on generating and sharing knowledge and good practices in the form of tools, analytical...
At the country level, World Bank teams and country stakeholders have worked together to develop and implement a set of READ Trust Fund-supported activities that address gaps in the country’s learning assessment system. Ultimately, the READ Council selected eight countries to be direct beneficiaries of READ Trust Fund support – Angola, Armenia, Ethiopia, the Kyrgyz Republic, Mozambique, Tajikistan, Vietnam, and Zambia.

The READ Council adopted a results framework to systematically monitor progress at the country level (see Annex 1). The indicators used in the results framework draw heavily on knowledge and tools developed at the global level under the program. The READ Trust Fund program has been using this results framework to monitor implementation and results at the country level on an annual basis. After the country-level programs ended in October 2014, the READ Trust Fund team in Washington, DC collected additional data on each country to ensure that progress against this framework was fully captured.

2015 is the final year of the READ Trust Fund program. It was agreed at the outset that an external evaluation of the program’s activities and outputs would be conducted upon its completion. Based on the Russian Government and World Bank’s satisfaction with the program to date, negotiations are already underway for a second READ Trust Fund program. Hence, a primary purpose of this final evaluation will be to capture major lessons learned under the first READ Trust Fund so that the READ Council can consider them in the design and implementation of a second program. The proposed evaluation will therefore be both summative (looking back at READ 1) and formative (looking forward to READ 2) in nature.

GOAL, SCOPE, AND PRODUCTS OF PROPOSED WORK

The goal of this consultancy is to undertake an external evaluation of the READ Trust Fund program as it nears its completion date of June 30, 2015. The primary audience for the evaluation report is the READ Council. Secondary audiences include World Bank management and staff, and select government/local officials responsible for implementing the program in their respective countries. Key questions to be addressed by the evaluation include the following:

1. To what extent did the READ Trust Fund program achieve its stated purpose (i.e., to help low-income countries improve their student learning outcomes through the design, implementation, and use of robust systems for student assessment)?
   a. To what extent did global-level activities contribute to this result?
   b. To what extent did country-level activities contribute to this result?

2. What are some lessons learned that can be applied to the design and implementation of a future READ Trust Fund program?

The Consultant is expected to undertake the evaluation in a rigorous manner, and to make recommendations based on valid, reliable, and sound data collection and analysis. It is
expected that the Consultant will conduct the evaluation through a combination of (i) desk review of various materials and (ii) interviews with those actively involved in the work of the trust fund. Existing program documents and deliverables will be shared with the Consultant, in addition to a list of possible contacts to be interviewed (see Annexes 2 and 3).

The final product will be a well-designed and detailed evaluation report (50-70 pages, 12-point font, excluding annexes) that captures lessons learned under the first READ Trust Fund and which may be used to inform the design and implementation of a second program.

To accomplish these tasks the Consultant will:

- Assess the program’s overall success in achieving its purpose, drawing on key stakeholder views, completed products, activity records, and other evidence sources to make this determination;
- Assess the global knowledge products produced under the program with the intent of making recommendations for future work in this area;
- Assess overall program performance in terms of the relevance of results, sustainability, appropriateness of design, resource allocation, and general structure and organization;
- Identify lessons learned and provide recommendations for guiding a second READ Trust Fund;
- Based on feedback from the READ Council and READ Management Team, finalize and submit the evaluation report.

OUTPUTS

The Consultant will prepare: 1) a draft and final evaluation work plan, and 2) a draft and final evaluation report. The report will describe the purpose, methodology, and findings of the evaluation and offer evidence-based recommendations and lessons learned.

The Consultant will begin by preparing an evaluation work plan that provides a conceptual, methodological, and logistical framework for carrying out the evaluation. The work plan will provide sufficient detail in each of these areas so that it is possible for the READ Trust Fund Program Manager to make a solid determination as to the likelihood of the evaluation producing sufficient evidence to make the aforementioned ‘assessments’ as to the success of the program, overall and at the country and global levels. The work plan must be approved by the READ Trust Fund Program Manager and will act as the agreement for how the evaluation is to be conducted.

The evaluation work plan (5-10 pages) will, at a minimum, address the following key elements:

- Primary objective(s), questions, and audience(s) for the evaluation
- Program evaluation methodology to be used – e.g., logic model, theory-based model, objectives-oriented, CIPP (context/input/process/product), IPO (input/process/output), discrepancy, other
Data sources and data collection methods to be used
Analytical procedures to be used, including approaches to ensuring validity and reliability
Detailed Table of Contents for the evaluation report
Suggested communication/dissemination strategies for final report
Work schedule, including dates for delivery of draft and final evaluation report
Any issues or concerns regarding the evaluation, along with suggestions for how to address them

Four phases are proposed for delivery of the final product:

1. Draft Evaluation Work Plan: To be submitted by the Consultant to the READ Trust Fund Program Manager within one (1) week of signing the contract.

2. Evaluation Work Plan: Within three (3) days of receiving comments on the draft work plan, the Consultant will produce a final evaluation work plan.

3. Draft Evaluation Report: The Consultant will submit a draft evaluation report for review on a date jointly agreed in the evaluation work plan.

4. Evaluation Report: Within ten (10) days of receiving comments on the draft evaluation report, the Consultant will submit a final evaluation report, including an abstract/executive summary.

REPORTING ARRANGEMENTS

The Consultant will complete this assignment under the overall direction and guidance of the READ Council. Day-to-day supervision and questions will be coordinated by the READ Trust Fund Program Manager, Marguerite Clarke (mclarke2@worldbank.org). The Consultant will submit draft and final versions of the work plan and the evaluation report to the READ Trust Fund Manager for review and comment/approval. The draft and final versions of the evaluation report also will be shared with the READ Council for their comment/approval. The Consultant will use feedback from these groups to revise and finalize the evaluation report.

REQUIRED SPECIALIZED SKILLS OF THE CONSULTANT

The Consultant should have the following qualifications:

- At least five years of practical and successful work experience in the area of program evaluation;
- Demonstrated technical skills in the collection and analysis of qualitative and quantitative data, including qualitative interviewing;
- Prior successful experience working with World Bank projects;
- Knowledge and expertise in the area of student assessment;
- Fluency and excellent writing skills in English;
- Excellent interpersonal skills and ability to work constructively and productively with others.

LOCATION OF ASSIGNMENT, TIMING, DURATION, AND PAYMENT

The Consultant will conduct interviews with relevant World Bank and Russian representatives remotely or in person, based upon the location of the Consultant. This may involve travel to Washington, DC to interview relevant World Bank staff as well as possible travel to Moscow, Russia to interview relevant Russian representatives and stakeholders. Any travel is subject to prior discussion with, and approval by, the READ Trust Fund Program Manager. Remuneration will be based on satisfactory deliverables and the READ Trust Fund Program Manager’s approval of Requests for Payment submitted by the Consultant and processed by the World Bank in line with World Bank policy. Travel expenses will be handled separately.

SAFEGUARD

The Consultant will make evidence-based recommendations without fear or favor as the Consultant is independent of the World Bank.
ANNEX 1: READ Trust Fund Results Framework

**GOAL:** Support the improvement of student learning outcomes through the design, implementation, and use of robust student assessment systems

**KEY INDICATORS:**

<table>
<thead>
<tr>
<th>Enabling Context</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EC1 – Setting clear policies</strong></td>
<td>There is a formal document(s) that provides guidelines for assessment activities</td>
</tr>
<tr>
<td></td>
<td>The formal document(s) is available to key stakeholders and the public</td>
</tr>
<tr>
<td><strong>EC2 – Having strong leadership</strong></td>
<td>Key stakeholders support the assessment activities</td>
</tr>
<tr>
<td></td>
<td>There is key stakeholder support for continuous improvement of assessment activities</td>
</tr>
<tr>
<td><strong>EC3 – Having regular budget/funds for assessment activities</strong></td>
<td>There is a line item in the government education budget for assessment activities</td>
</tr>
<tr>
<td></td>
<td>The budget provides adequate funding in major areas -- design, administration, reporting</td>
</tr>
<tr>
<td><strong>EC4 – Having strong organizational structures</strong></td>
<td>There is an agency, institution, or unit with the mandate to carry out assessment activities</td>
</tr>
<tr>
<td></td>
<td>The assessment agency, institution, or unit is accountable to a clearly recognizable body</td>
</tr>
<tr>
<td><strong>EC5 – Having effective human resources</strong></td>
<td>There is a team of people with the requisite skills/capacity to carry out assessment activities</td>
</tr>
<tr>
<td></td>
<td>There are opportunities available to build assessment capacity; e.g., courses/training</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System Alignment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SA1 – Aligning the assessment with learning goals</strong></td>
<td>There is a clear, common understanding among key stakeholders of what the assessment activities measure</td>
</tr>
<tr>
<td></td>
<td>Assessment activities are aligned with an official curriculum/learning standards that outlines what students are expected to learn</td>
</tr>
<tr>
<td><strong>SA2 – Providing opportunities to learn about assessment activities</strong></td>
<td>There are training sessions/courses for teachers to learn about the assessment activities</td>
</tr>
<tr>
<td></td>
<td>Teachers are involved some aspect(s) of assessment-related activities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessment Quality</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AQ1 – Ensuring quality</strong></td>
<td>There is formal documentation about the technical aspects of the assessment activity</td>
</tr>
<tr>
<td></td>
<td>Assessment results are deemed by key stakeholders to be valid and reliable</td>
</tr>
<tr>
<td><strong>AQ2 – Ensuring effective use of assessment results</strong></td>
<td>Assessment results are disseminated in meaningful ways to key stakeholders</td>
</tr>
<tr>
<td></td>
<td>Assessment results are used to promote and inform students’ learning</td>
</tr>
</tbody>
</table>
ANNEX 2: List of Documents To Be Reviewed

1. READ Trust Fund Concept Note
2. READ Trust Fund Administrative Agreement with Amendments
4. READ Trust Fund Comprehensive Report 2012
5. READ Council Meeting Aide Memoires
7. SABER-Student Assessment Conceptual Framework Paper – What Matters Most for Student Assessment Systems
8. SABER-Student Assessment Benchmarking Tools – Questionnaires, Rubrics, and Country Reports
9. National Assessments of Educational Achievement Series (Volumes 1-5)
10. READ Working Papers (1-12)
11. READ Video
12. Country-level Products/Reports (10-12 in total)
13. READ Website
14. READ Global Conference and Regional Workshop Materials
Appendix 2: Evaluation Work Plan

Overview

The goal is to evaluate the overall impact of the seven years of work under the READ Trust Fund program. The evaluation will focus on what has been achieved through the READ-funded program. The first audience for the evaluation’s findings is the READ Council, to inform it of accomplishments achieved and challenges still to be addressed. In addition to assessing the impact of work to date, the evaluation will make observations and suggestions for future activity in this domain. The World Bank’s senior leaders and technical experts may also find the evaluation informative and instructive for ongoing work in education and in co-financing and donor relationships. Government officials from participating countries and from other donor governments may also find the evaluation useful as a tool to reflect on their own activities and programs.

Operationalizing the Terms of Reference

The Terms of Reference helpfully set out the following “key questions to be addressed”;

1. To what extent did the READ Trust Fund program achieve its stated purpose (i.e., to help low-income countries improve their student learning outcomes through the design, implementation, and use of robust systems for student assessment)?
   a. To what extent did global-level activities contribute to this result?
   b. To what extent did country-level activities contribute to this result?
2. What are some lessons learned that can be applied to the design and implementation of a future READ Trust Fund program?

These questions and the detail in the Results Framework for READ point to the need to evaluate the array of funded activities against the explicit purpose of the program, which is to support the development and operation of better student assessment systems through national and global projects and activities. It also distinguishes two levels of activity - global and country; a distinction which we will attend to in data collection, in the analysis, and in reporting. This is essentially the summative part of the evaluation.

The Terms of Reference are also forward looking, envisaging a future READ program in broadly the same domain of activities. To guide the design and operation of a second program of work,
the evaluators are asked to identify “lessons learned”. This is the formative piece of the evaluation.

These key questions will be used to structure the final report and to organize the main conclusions. This will guide interviews with the main informants and enhance consistency in data collection and analysis. Similarly, these questions need to be broken into some smaller discrete themes to provide a common framework for the review and analysis of the program documentation, which runs for many pages.

**Some Open-Ended Guiding Questions**

We see five overarching questions shaping the study and guiding data collection and analysis. We believe that they can be used with most informants and in reviewing the READ program products and documents. They should also allow us to test the viability of the theory of action, the logic model that was embedded in the Results Framework and administrative agreements. That model assumed that providing accurate, timely, and accessible information on how well students are performing to teachers, policy makers, and international donors alike would lead them to focus their energy and resources for the greatest improvement in learning outcomes. The seven-year span of READ-supported activities, while commendable, is still too short to see the full effect, or even the full realization, of all of the steps in the logic model. We will endeavor to lay out the steps and offer an assessment of how far the READ Trust Fund’s efforts have progressed to date. This should also point to areas for further work or greater investment.

The five broad questions we will pursue are:

1. What worked?
2. What did not and why not?
3. What gaps have been identified in the program coverage that should be addressed in future work?
4. What processes and procedures should be refined to increase effectiveness?
5. How do the participant countries perceive the READ work and how have they integrated it into ongoing activities?

Framing questions in this way allows the same themes to be pursued across different types of artefacts, in interviews with people from various institutional and cultural contexts, and across different levels of authority. They should generate insights into broader issues like sustainability and impact as well as specific topics like increasing institutional capacity, data usage, and improvements in systems of student assessment without leading informants to address particular topics. The use of open-ended questions in interviews also allows participants to offer fresh insights into the topic by raising subjects not anticipated by the evaluators.
Should additional interview prompts be required, we will use some or all of the following set of questions to deepen our coverage of the issues. They will be adjusted depending on the particular responsibilities of the interviewee when they were part of the READ program. Prompts are by nature a little more directive. We might ask;

(a) What were your expectations for the READ-funded work?
(b) How did it add to the effectiveness of your (assessment system/technical assistance/Bank-client dialogue)?
(c) What is your assessment of the value and impact of READ-funded work (nationally/regionally/globally)?
(d) With hindsight, what would you have done differently?
(e) What would you set for the national or global priorities for a further cycle of READ programming?

To encourage candor, we will not record the interviews but take contemporaneous field notes. We will also advise informants that the final report will only identify respondents by position or area of responsibility. In the case of a direct quote, our intention is to give the informant the draft of the relevant section to see if it was interpreted accurately. This should only be an issue when the individual's comment is an outlier, an N of 1, or particularly apt for the point being made。

We will of course make this clear in requests for interviews and at the beginning of each interview. We will also ensure that our notes are held securely and destroyed after the formal completion of the project.

**Audience**

The audience of this evaluation will primarily be the READ Council. They have a natural interest in the summative aspects of the study and how effective it has been in supporting the development of better student assessment systems. And as the entity likely to be responsible for any further similar program, the Council will benefit directly from the forward-looking, formative work.

Other audiences include Bank staff, policy makers in client countries, and technical experts in the areas of student assessment and examination systems. It may also be of interest to people interested in international development assistance and donor-led programming.

A subsidiary audience might include third parties interested in organizing similar programs, and academics conducting research related to development or education policy.
Program Evaluation Methodology

The study will utilize a qualitative methodology based on the triangulation of documentation and interviews. We will begin with a desk review of the materials stipulated in Annex 2 of the Terms of Reference. We will take the READ Trust Fund Administrative Agreement, Concept Note, and Results Framework as the foundation documents and use them as an organizing framework for the analysis of the materials and products generated throughout this cycle of the program. We will look for alignment between the main elements of these documents, particularly the Results Framework and subsequent activities, noting variations, discontinuities, and where observable, the reasons for changes in approach and how they contributed to or influenced program effectiveness. This will reveal how robust the initial foundation documents were and also provide a sensible basis for identifying and acknowledging the challenges of implementation that arise in any cross-national, large-scale, multi-year, multi-actor program.

Using this organizing framework for reviewing the program’s documentation should also strengthen the reliability and consistency of analysis. There is a common set of terms and concepts, a common technical language to guide the reviewers. More importantly, the Results Framework uses a set of terms and concepts that are known by the potential interviewees and has been used to report the progress of the program for many years. This should increase the quality of the data collected.

The document review will also look for common themes in the content of program activities, common challenges in design and delivery of program activities, and common constraints on implementation. It will also identify successes, intended and unintended benefits, and any notable shortfalls. Finally, the desk review will note any distinctive or singular country or global activity that yielded significant impact or was particularly disappointing.

We would hope to interview many, if not all, of the actors identified in Annex 3 using the framework described above and augmented where relevant by specific questions about products or programs that may emerge from the document review. Consistent with our plan we will start with the foundation documents, the READ Trust Fund Concept Note, Administrative Agreement, and Results Framework and then follow the cascade of materials.

Our intention is to start with at least some of the core team program managers and READ Council members and then move on to actors on the various programs and country teams.

Analytical Procedures

Evaluative work of this kind faces significant challenges in ensuring that conclusions and judgments are grounded on good data. This challenge is heightened by the diversity of data sources, especially given the different vantage points of the various actors: some with a global perspective and some with a single-country focus. To address this, we will follow the basic principle of triangulation: “Good research practice obligates the researcher to triangulate, that
is, to use multiple methods, data sources, and researchers to enhance the validity of research findings” (Mathison 1988:13).

We will also follow a “systematic and essentially taxonomic process of sorting and classifying data” as it is being collected as this plus immersion in the data, coding data, categorizing and identifying themes lead to good evidence in qualitative research (Green et al, 2007:546). We will start coding using an inductive method as we undertake the interviews and review documents. This is in the grounded theory tradition of qualitative research, but its benefits are well summarized by Thomas (2003): “The primary purpose of the inductive approach is to allow research findings to emerge from the frequent, dominant or significant themes inherent in raw data, without the restraints imposed by structured methodologies. Key themes are often obscured, reframed or left invisible because of the preconceptions in the data collection and data analysis procedures imposed by deductive data analysis such as those used in experimental and hypothesis testing research.”

The inductive approach is well suited to both the formative and summative aspects of the evaluation. While we are looking at the effectiveness of a particular theory of change, we are not elevating it to the level of an empirically-testable hypothesis.

We will also pay careful attention to the views of different groups, noting that global team members, country team members, and country client participants are likely to have different perspectives and experiences.

**Communication/Dissemination Strategies**

While the main audience for the report is the READ Council, we would also propose that the report be readily accessible to all those interviewed. This will encourage co-operation in the data collection process and possibly add to the fulsomeness and veracity of responses.

Given the range of audiences we identify above (pages 1 & 4) and the fact that most of them would have access to the World Bank’s home page, an e-publication strategy is likely to reach all cost-effectively. Reach would be maximized if all those interviewed and any other relevant stakeholders were contacted by the READ Trust Fund Program Manager to inform them the report is available for download. The report should be available for download in a variety of file types and, if appropriate, in languages other than English.

It may be valuable if the lead evaluator orally briefed the READ Council, but that is a matter for the Council’s leadership.

There are additional qualifications and addendums that may be necessary to include depending on how public the READ Council decides that the report will be.
**Work Schedule**

We plan to present a draft evaluation plan by 26 May and finalize it by 1 June 2015.

We will then begin the document review, starting with extant documents and taking into account key documents that will be completed during June. This analysis will continue into July.

Our intention is to begin interviews with key program staff in the first half of July.

This will be followed by interviews with World Bank technical and program staff and any client country informants they identify. These will run through August given likely constraints of individual’s travel and vacation schedules.

A brief progress report on emerging themes will be submitted by 21 August for comment and identification of any areas for further inquiry or analysis.

The suggested delivery date for the draft final report is 30 September, with a final report due after discussion and necessary revision by 21 October.

**Report Outline**

While the structure of the report will evolve with the marshalling of evidence, analysis of data, and the emergence of themes, we expect it to contain the following:

1. Executive Summary
2. Overview of the READ program (to ensure the evaluation report is self-contained.)
3. Purpose of the Evaluation
4. Evaluation Methodology
5. What was Achieved Globally
6. What was Achieved Nationally
7. Overall Conclusion: a summation
8. Looking Ahead
   a. Lessons learned about design and implementation
   b. Lessons learned about the domain of student assessment
   c. Opportunities for further work in this domain
   d. The scale and direction of such further work.

**Annexes**

A. TOR
B. Documents examined
References


Appendix 3: List of Interviewees

The review team would like to thank all those who were interviewed as a part of this project. Their insights and observations, coupled with experience and candor, added greatly to the team’s understandings of the complexities and successes of the READ project. Errors of fact and failure to appreciate the nuances of the various national circumstances are our responsibility, not theirs.

READ Council
Cristian Aedo
Andrey Bokarev
Luis Benveniste
Robin Horn
Alberto Rodriguez
Andrei Volkov

Program Management/Global Team
Marguerite Clarke
Emily Gardner
Julia Liberman

Country Teams
Michael Crawford
Dingyong Hou
Sophie Nadeau
Jem Heinzel Nelson
Anush Shahverdyan
Mai Thanh
Girma Woldetsadik

Others
Andrei Markov
Tigran Shmis
Anna Valkova
Appendix 4: Vignettes of READ Country Activities and Achievements

Angola

After a 25-year civil war, Angola had to re-build in 2009. Angola had no culture of assessment in the education system and no staff dedicated to assessment. READ activities began with two objectives: 1) to improve the assessment capacity of Angola’s Ministry of Education, and 2) to develop a culture of evidence-based decision making. Initial interest in READ activities was muted. Angola was benefiting from high oil prices and there was no World Bank education program in the Ministry. READ helped open a policy dialogue on the importance of measuring learning outcomes. By the end of the READ program, the Ministry had created clear policies mandating regular Early Grade Reading Assessments and a National Longitudinal Survey of student achievement. The technical expertise of Ministry staff had been strengthened; a regular budget line-item for assessment, in the amount of US$1 million, had been created; and a National Assessment System Technical Group (NASTG) had been formed with permanent staff. At the same time, there is still much to be done and the nation would benefit from READ 2 resources and technical assistance.

“Angola needs more education data so the policy makers, teachers and parents can talk about education outcomes. They didn’t want to share their assessment results in the very beginning”. They need a “plan for the future related to national assessment data, and that will not happen well without outside help.”

Armenia

Education has been a priority of the government in Armenia since Independence. Despite significant advances in the education system, issues persist about reliable measurement of learning outcomes and inequities in educational outcomes between rich and poor. The goal for READ in Armenia was threefold: 1) enhance capacity to design and implement national and international assessment systems, 2) strengthen in-classroom assessment exercises, and 3) provide channels for feedback, policy analysis, and recommendations for actions to improve student learning outcomes. The 2012 SABER-Student Assessment baseline study showed a pressing need to improve national assessment capability.

In Armenia, “there is still a lot to be done, to follow up, to consolidate, to strengthen the in-country educational opportunities for teachers and officials to learn about assessment.”
In a relatively short time, Ministry officials were trained in aspects of educational assessment; a Master’s degree program in education at State Linguistic University with specific courses on assessment was developed, and formative and summative assessment strategies were introduced into pre-service and in-service teacher education programs.

**Ethiopia**

Quality is the biggest issue facing the education system. The primary goal for READ in Ethiopia was to improve educational assessment to provide reliable information to guide the design and implementation of strategies to improve student learning outcomes. The main strategy was to strengthen existing institutions and develop some new agencies and processes by coordinating READ activities with work financed by other donors. Major accomplishments included creating an autonomous National Educational Assessment and Examinations Agency, training its staff in various aspects of assessment, and supporting the development of sustainable item banks for both National Longitudinal Assessments and national examinations. By the end of READ 1, a Directorate of School Inspection had been established and national and regional officials had been trained in how to use the framework. Implementation commenced with a pilot phase reaching 28,000 schools. School inspection activities are continuing with support from a large World Bank loan. With technical assistance, self-diagnosis in each area was undertaken, which identified gaps and areas of improvement. There has been a lot of progress in building up the assessment system in Ethiopia, but “while a lot has been achieved, there is still a lot to be done and a READ 2 would accelerate progress. Main three areas for future work are international benchmarking - TIMSS or the African cross-national program; school graduate employability; better analysis and use of national assessment results and Grade 10 and 12 exam results.”

**Kyrgyz Republic**

The low quality of education in the Kyrgyz Republic was highlighted by the 2006 and 2009 PISA surveys. The Kyrgyz Republic ranked last out of all participating countries. The government responded with reforms aimed at strengthening the system and improving the monitoring of student learning outcomes, especially through better formative assessment at the classroom level, with over 6,000 primary school teachers being trained. At the system level, the government
developed and adopted a new national assessment strategy that includes a reliable and secure summative assessment process. This led to the creation of a new Grade 11 exit examination and a Grade 4 large-scale assessment. The independent Center for Educational Assessment’s work on improving the quality of national sample surveys was supported by READ funds and the results of the 2014 survey were disseminated to key stakeholders. This was a marked increase in public availability of assessment findings, with greater transparency on test specifications and scoring criteria, which lays a basis for improving teaching methods. Overall the “READ Trust Fund was useful in building capacity -- both technical and physical”. The results have shaped and informed future investments in learning. “New grant projects will target preschool because of low access, and the another grant is focusing on high schools and improving student learning.”

**Mozambique**

National examinations, NLSA, and ILSA were all occurring in Mozambique when READ activities began. However, they were disconnected parts of an inchoate system and were not used effectively to improve the quality of learning outcomes. The primary goal of READ activities in Mozambique was to improve the technical and institutional capacity of the Ministry of Education to assess student learning outcomes at all levels of schooling. READ financed training activities in various forms, ranging from study tours and conferences to formal degree programs. Further, READ funds “lifted the profile of classroom assessment”. One of the most visible results of the READ work was the funding of Provinha, an early grade reading assessment which included communication materials for teachers, school administrators, and parents. After an initial pilot was successful, the program was expanded to other provinces. Results of Provinha assessments were used at national policy forums and increased public debate about learning outcomes. Other achievements attributed in some measure to READ activities include an annual budget allocation for a national learning assessment program and wider and more meaningful engagement of civil society in the development of the national education strategy. READ activities also attracted the attention of other donors who found the assessment projects solid and safe investments.

**Tajikistan**

In Mozambique “READ produced some very good outcomes, it built the internal capacity on assessment in the exam council, in universities and in the Ministry of Education; a cohort of trained people who are still in government service.”
Educational quality and access to higher levels of education are the most pressing education issues in Tajikistan. The primary goal of READ activities was to increase capacity for assessment in the country.

Training events and study tours supported by READ funds enabled key Ministry of Education officials to develop a critical understanding of best practices and issues in assessment. The Tajikistan country project team built an ambitious program, blending READ money with funds from the Open Society Institute, USAID, Education for All resources, and a World Bank credit to support a “whole reform agenda that included building a secure testing facility. It was a coordinated approach, which leveraged the READ funds to great effect” to create a better university entrance examination and improve the school curriculum.

A centerpiece of the strategy was the new University Entrance Examination (UEE). The READ program assisted in the development of clear policies, in workshops for leaders on how to administer the new examination procedures, in setting a new budget for the UEE, in overhauling the organization of the National Testing Center (NTC), in capacity building for NTC and Ministry of Education and Sciences officials, aligning of UEE policy with university goals, in communicating the new UEE plan to a broad range of stakeholders across the country, and in ensuring the integrity of the new UEE.

Vietnam

With near-universal primary education achieved, officials in Vietnam had begun to focus on educational quality. With a system of nearly 40,000 schools and 20 million students, designing and implementing assessment activities is challenging in terms of scale alone. An additional and equally significant challenge is to transform the country’s curriculum and assessments from a focus on the recall of knowledge to a focus on the application of knowledge: a competency-based system. The goal of READ activities in Vietnam was to improve the effectiveness and scalability of the nation’s assessment activities. READ-funded activities were a vital element in attempts to strengthen the assessment system. These activities included training programs, study tours, and national and regional workshops for Ministry of Education and Training officials and specialists, and conferences, coursework, study visits, and workshops for regional and

“In Vietnam the World Bank’s curriculum loan was improved by the READ project. It helped the client think through standards, draw from global expertise and build capacity nationally and locally- the school level champions – aligning trained people with message and materials.”
district administrators. Thus, in the area of national examinations, NLSA, and ILSA, READ was critical in building up technical capacity, knowledge of the system’s needs, system alignment, quality assurance, and clear policy practices. Vietnam was successful to the point where READ funds were reallocated to other countries in order to make a larger impact. Thus, it is likely that if the education and assessment systems continue to improve at this rate, Vietnam will be fully prepared to enter the global knowledge economy. READ funds were used to explore Vietnam’s participation in PISA 2012 and the findings from this work provided the basis for the Asia Development Bank financing the full cost of the survey in Vietnam. Other assessment activities were supported by a World Bank school quality assurance project and by the Global Partnership for Education, adding to the impact of the READ Trust Fund’s investment in the country.

Zambia

While Zambia has nearly universal primary education, there were still significant gaps in education quality and learning outcomes that could be addressed by the READ program. The 2012 national assessment study found that only four out of ten fifth graders were meeting basic literacy and numeracy standards. READ-funded activities concentrated on strengthening the institutions responsible for monitoring student outcomes and devising strategies to improve classroom practice. Much of the capacity building involved conferences, trainings, workshops, study visits, and university courses and reached over 1,600 stakeholders. Other activities were undertaken in order to ensure system alignment in classroom assessment and national examinations, ensuring quality of assessments of all types, and in disseminating knowledge of assessments to the general public. READ also financed the initial costs of Zambia’s participation as one of seven pilot countries in the PISA for Development project. Ongoing costs of participation are being financed by the national budget and the United Kingdom’s governmental aid program.
Appendix 5: References


