USAID | KENYA
FROM THE AMERICAN PEOPLE

EdData II

The Primary Math and Reading (PRIMR) Initiative

Endline Impact Evaluation

EdData II Technical and Managerial Assistance, Task Number 13
Contract Number AID-623-M-11-00001
Strategic Objective 3
Date: April 2014

The Primary Math and Reading (PRIMR) Initiative
Endline Impact Evaluation

Prepared for
Dr. Teresiah Wambui Gathenya, Contracting Officer's Representative (COR),
United States Agency for International Development (USAID)/Kenya

Prepared by:
Dr. Benjamin Piper & Dr. Abel Mugenda
RTI International
3040 Cornwallis Road
Post Office Box 12194
Research Triangle Park, North Carolina 27709-2194
USA

Table of Contents

Section Page

List of Figures

List of Tables

Abbreviations

CDE	County Directorate of Education
CEMASTEA	Centre for Mathematics, Science and Technology Education in Africa
clpm	correct letters per minute
COR	Contracting Officer's Representative
cpm	correct per minute
cspm	correct syllables per minute
cwpm	correct words per minute
DFID	UK Department for International Development
DID	differences-in-differences
DQASO	District Quality Assurance and Standards Officer
EdData II	USAID Education Data for Decision Making II project
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
ICT	information and communication technology
KEMI	Kenya Education Management Institute
KES	Kenya shilling
KG	kindergarten
KICD	Kenya Institute of Curriculum Development
KISE	Kenya Institute of Special Education
KNEC	Kenya National Examinations Council
LTTP	Liberia Teacher Training Program
M&E	monitoring and evaluation
MoEST	Ministry of Education Science and Technology
NESSP	National Education Sector Support Programme
NGO	nongovernmental organization
OLS	ordinary least squares
PDIT	Program Development and Implementation Team
PRIMR	Primary Math and Reading Initiative
RTI	RTI International (trade name of Research Triangle Institute)
SAGA	Semi-Autonomous Government Agency
SD	standard deviation
SSME	Snapshot of School Management Effectiveness
TAC	Teachers' Advisory Centre
TSC	Teachers' Service Commission
US	United States
USAID	United States Agency for International Development

Acknowledgments

We appreciate the hard work and commitment of the Ministry of Education, Science and Technology (MoEST), led by the Cabinet Secretary, Professor Jacob Kaimenyi, and the Principal Secretary, Dr. Richard Belio Kipsang'. We also appreciate the support of the Directors of the MoEST and the relevant semi-autonomous government agencies, and the Programme Development and Implementation Team housed at the MoEST that included members from across the Ministry and other organizations. Funding is provided from the American people through USAID/Kenya. The USAID/Kenya education team, specifically Dr. Dwaine Lee, Dr. Christine Pagen, and Dr. Teresiah Wambui Gathenya designed a research study worth analysis. The PRIMR M&E team is a very strong team focused on collecting valid data. From a technical perspective, Jessica Mejia, Joseph DeStefano and Melinda Taylor's leadership and support were instrumental in this research and in the PRIMR Initiative. We also acknowledge the excellent editorial work of Erin Newton.

Executive Summary

The Primary Math and Reading (PRIMR) initiative is a partnership between the United States Agency for International Development (USAID/Kenya) and Kenya's Ministry of Education, Science and Technology (MoEST), with technical implementation by RTI International. PRIMR works closely with the MoEST and several of its Semi-Autonomous Government Agencies: Kenya National Examinations Council (KNEC), Kenya Institute of Curriculum Development (KICD), Kenya Education Management Institute (KEMI), Kenya Institute of Special Education (KISE), and Teachers' Service Commission (TSC). PRIMR is managed by the MoEST's Program Development and Implementation Team (PDIT). Under their direction, PRIMR worked with 547 formal and low-cost private schools (LCPS) from Nairobi, Murang'a, Kiambu, and Nakuru counties. A PRIMR study on the use of information and communications technology (ICT) for instructional support was implemented in Kisumu County in 2013 (Piper & Kwayumba, 2014).

PRIMR used evidence-based instructional improvement methods to increase the fundamental literacy and numeracy skills of students in grades 1 and 2 (Classes 1 and 2). Key components of the intervention have been innovative teaching methods, new materials based on the Kenyan syllabus, and professional development and coaching for educators related to actual teaching practice.

PRIMR has had three cohorts of schools. The first began the full scope of interventions per the project design in 2012, continuing through 2013; the second began the interventions in 2013. Cohort 3 was a control group that began receiving the interventions in 2014, after the endline assessment. PRIMR used the Early Grade Reading Assessment (EGRA) and the Early Grade Mathematics Assessment (EGMA) instruments to assess the impact of project activities. In January 2012, a baseline evaluation was administered to a random sample of pupils from all three school cohorts, and in October 2012, a midterm evaluation was conducted (Piper & Mugenda, 2013). This report is based on the endline evaluation conducted in October 2013. The basic question that this study is organized to answer is ***whether PRIMR had an effect on pupil achievement in reading and math.***

PRIMR's Impact

The randomized controlled trial design of PRIMR made it feasible to estimate the impact of PRIMR on learning. ***Table ES1,*** organized according to selected subtasks from the English EGRA instrument, shows the mean scores at the endline for pupils in the PRIMR treatment schools (Cohorts 1 and 2) and those in control schools.

For letter-sound fluency, treatment pupils in PRIMR identified 47.0 correct letters per minute (clpm) correctly, compared to 25.7 letters per minute among the control pupils. PRIMR's causal effect was 21.3 clpm, or 0.73 standard deviations (SD). In oral reading fluency, the PRIMR effect was 13.7 correct words per minute (cwpm) overall. If Cohen's effect size research says that .50 SD is a large impact, these are very large. This equates to more than 1 year of gain for pupils in control schools. Reading comprehension scores were more than twice as high in PRIMR (21.1%) as they were in control schools (9.8%) in Class 1, and the absolute gain in comprehension attributed to PRIMR in Class 2 was 17.3%. Although the

number of pupils supported in 2013 nearly tripled from the year before, the proportion of pupils reading at benchmark by the time of the endline assessment was more than twice as high in PRIMR (28.3%) than control schools (12.6%). The impact of PRIMR also was felt on the proportion of pupils reading at the KNEC benchmark for English (65 or more wpm), with more than twice as many treatment pupils reading at benchmark in both Classes 1 and 2. Effect sizes were moderate to large across the English subtasks, with an average overall effect size of 0.46 SD.[1]

Table ES1. Endline impact of PRIMR on English outcomes (selected measures)

	Overall			Class 1			Class 2		
English EGRA subtasks	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*
Letter-sound fluency (correct letters per min.)	**47.0**	25.7	*0.73*	**43.5**	24.6	*0.68*	**50.8**	26.8	*0.78*
Oral reading fluency (correct words per min.)	**45.1**	31.4	*0.40*	**32.2**	20.1	*0.44*	**58.9**	42.8	*0.45*
Reading comprehension (% correct out of 5 questions)	**34.3**	19.4	*0.38*	**21.1**	9.8	*0.38*	**48.4**	29.1	*0.44*
Reading at benchmark (% of pupils reading 65 cwpm+)	**28.3**	12.6	*0.36*	**14.0**	4.0	*0.32*	**43.7**	21.3	*0.45*
Average effect size			***0.46***			***0.47***			***0.49***

Table ES2 presents PRIMR's impact on Kiswahili, as measured by selected Kiswahili EGRA subtasks.[2]

For letter-sound fluency, the results show that the PRIMR effect was 15.6 clpm for Class 1 and 22.1 clpm for Class 2. The overall effect size for letter-sound fluency was 0.63 SD. Surprisingly, while the control classrooms were spending a great deal of time on learning syllables, the PRIMR program still showed a 0.41 SD effect on syllable fluency. In Class 2, that equates to 11.9 cspm. PRIMR effects on oral reading fluency were 7.0 cwpm (0.41 SD) in Class 1 and 6.7 cwpm (0.35 SD) in Class 2. Gains were also identified for reading comprehension, with a 0.45 SD effect in Class 1 and a 0.32 SD effect in Class 2.

For the proportion of pupils reading at the Kiswahili benchmark (45 wpm), scores in PRIMR were nine times larger in Class 1 (0.28 SD) and two times larger in Class 2 (0.30 SD). Overall, the effect of PRIMR in Kiswahili was 0.39 SD in Class 1 and 0.36 SD in Class 2. Outcomes were higher than those presented in the midterm analysis report (Piper & Mugenda, 2013), and higher than might have been expected given the school closures and

[1] An effect size is calculated by dividing the causal program effect by the pooled standard deviation. It is a measure of the effectiveness of an intervention that can be compared against the effects in other programs.
[2] Note that the Kiswahili EGRA varied from the English version to accommodate characteristics of the language.

other distractions during the March 2013 national election and the five-week teacher strike of June–July 2013.

Table ES2. Endline impact of PRIMR on Kiswahili outcomes (selected measures)

	Overall			Class 1			Class 2		
Kiswahili EGRA subtasks	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*
Letter-sound fluency (correct letters per min.)	**47.5**	28.8	*0.63*	**42.4**	26.8	*0.57*	**52.9**	30.8	*0.70*
Syllable fluency (correct syllables per min.)	**45.7**	34.6	*0.41*	**38.4**	27.6	*0.42*	**53.3**	41.4	*0.45*
Oral reading fluency (correct words per min.)	**27.4**	20.6	*0.35*	**20.9**	13.9	*0.41*	**34.0**	27.3	*0.35*
Reading comprehension (% correct out of 5 questions)	**35.9**	25.8	*0.34*	**25.6**	14.9	*0.45*	**46.6**	36.5	*0.32*
Reading at benchmark (% of pupils reading 65 cwpm+)	**15.9**	6.7	*0.27*	**7.2**	0.8	*0.28*	**24.9**	12.5	*0.30*
Average effect size			***0.35***			***0.39***			***0.36***

As explained in the midterm report, given the very limited amount of time that the math learner books and teachers' guides were in classrooms before the midterm assessment, PRIMR was not convinced that the positive effect identified in the midterm assessment was due to the program (Piper & Mugenda, 2013). In 2013, however, the math materials were in schools on time when the school year began in January.

Table ES3 presents the impact of PRIMR on mathematics outcomes on selected EGMA subtasks at the October 2013 endline. It shows a moderate effect of PRIMR on math overall of 0.16 SD for Class 1 and 0.26 SD for Class 2. PRIMR seemed to improve outcomes on the number identification (0.27 SD) and missing number (0.29 SD) subtasks, but had no effect on quantity discrimination (0.03 SD).[3] The computational measures showed some effect, with higher outcomes in addition fluency or subtraction fluency. The impact was consistently larger in Class 2 than it was in Class 1. Word problems showed a small impact (0.13 SD).

[3] The PRIMR mathematics program showed small or moderate impacts on all areas, except quantity discrimination. This task requires a developed number sense which remains difficult for many learners. In 2014, the program in this area was simplified as PRIMR moved away from scripted lesson plans to teachers' guides.

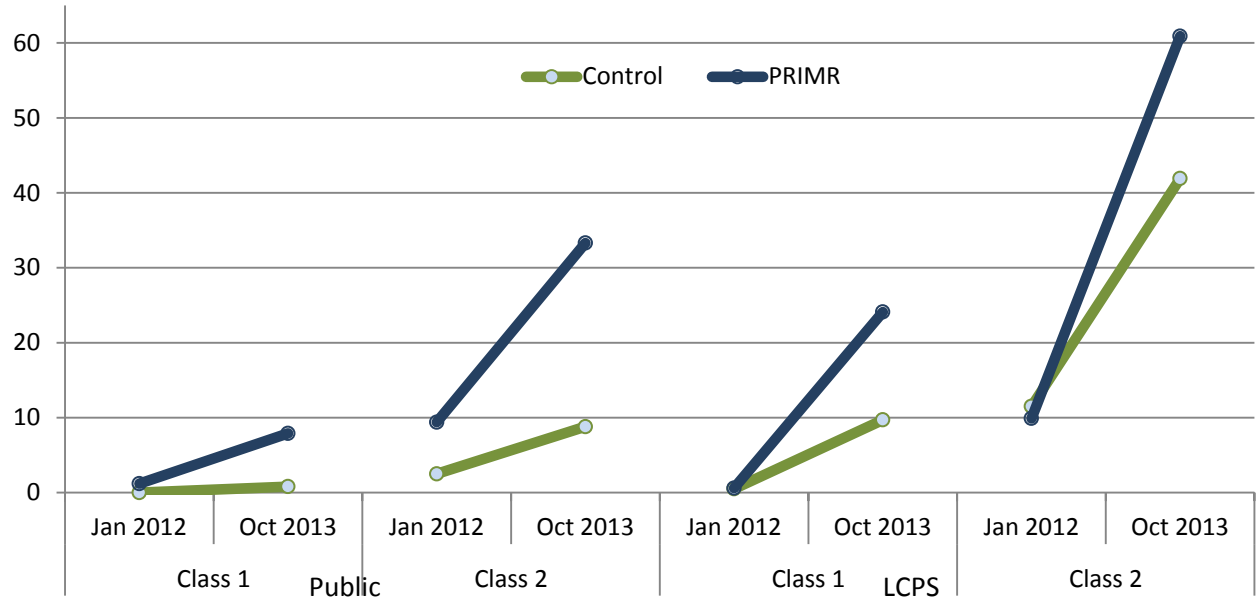
Table ES3. Endline impact of PRIMR on mathematics outcomes (selected measures)

	Overall			Class 1			Class 2		
EGMA subtasks	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*
Number identification (correct numbers per min.)	**24.5**	21.3	*0.27*	**19.6**	16.7	*0.31*	**29.6**	25.7	*0.33*
Quantity discrimination (% correct comparisons)	**59.9**	59.2	*0.03*	**48.4**	44.6	*0.16*	**72.0**	73.0	*-0.04*
Missing number (% correct)	**43.5**	36.8	*0.29*	**32.8**	28.6	*0.23*	**54.7**	44.6	*0.45*
Addition fluency (correct items per min.)	**10.1**	9.3	*0.17*	**7.9**	7.5	*0.10*	**12.4**	10.9	*0.33*
Subtraction fluency (correct items per min.)	**7.1**	6.2	*0.21*	**5.4**	4.7	*0.18*	**8.9**	7.5	*0.34*
Word problems (% of 5 items correct)	**40.7**	37.4	*0.13*	**33.9**	31.6	*0.10*	**47.8**	42.9	*0.18*
Average effect size			***0.20***			***0.16***			***0.26***

PRIMR and KNEC Benchmarks

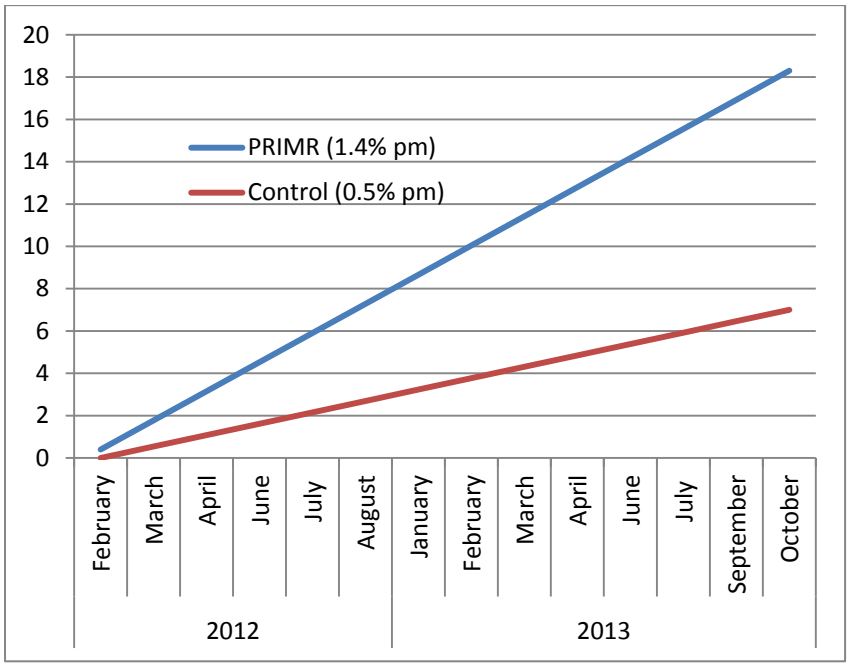
Figure ES1 is a graphical representation of the impact of PRIMR on the percentage of pupils reading at the KNEC English benchmark for Class 2. These results indicate that treatment pupils in Classes 1 and 2 and in public and LCPS were making significant gains in literacy. The rates of increase between PRIMR and control schools were dramatically different, and in short, PRIMR was helping these pupils become literate much faster than the control public or LCPS were able to. For this figure, Class 1 was measured against a Class 2 benchmark, so gains were expected to be modest. Similarly large gains were found in Kiswahili.

Figure ES1. Proportion of PRIMR and control pupils reading at English benchmark



Reading comprehension has proven difficult to improve in literacy programs (Piper & Mugenda, 2012). ***Figure ES2*** presents the increase in the proportion of Class 2 pupils able to comprehend at 80% or above, based on the reading comprehension subtask scores on the Kiswahili EGRA. The rate of increase in comprehension was nearly three times higher in PRIMR treatment schools than in control schools. These pupils were learning how to read *and* comprehend.

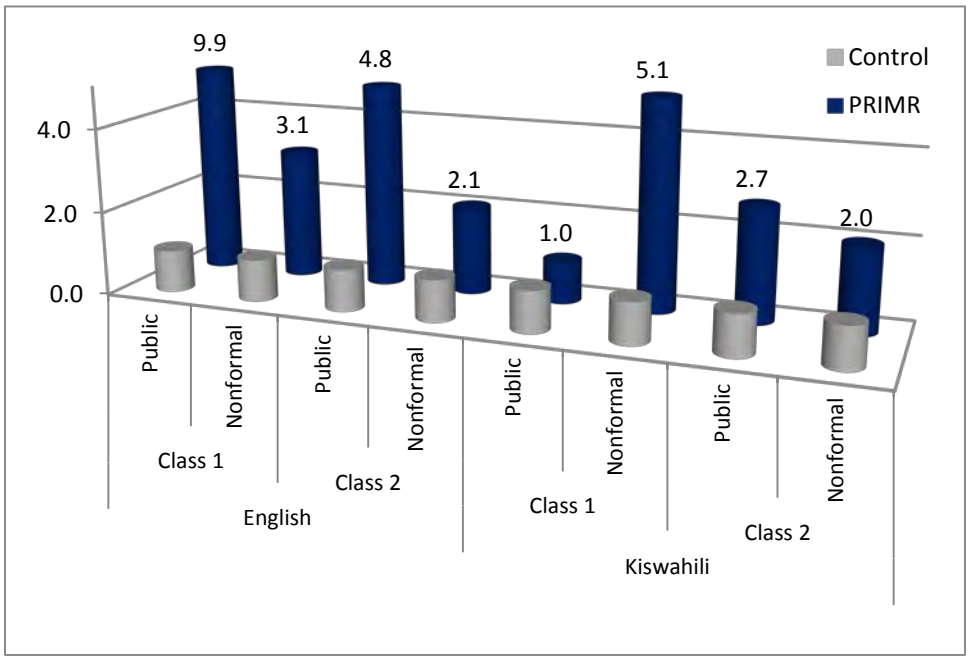
Figure ES2. Rate of increase in the proportion of pupils reading with 80% or higher comprehension: Kiswahili, Class 2



Likelihood of Reading at KNEC Benchmark

In order to assess whether PRIMR made it more likely that pupils would be able to read at the benchmark—controlling for key variables—we fit logistic regression models. The results (see ***Figure ES3***) show the odds that pupils in treatment schools would achieve the benchmark compared with pupils in control schools. The gray bars represent control schools, and always show 1. The blue bars show the likelihood of being able to read at the KNEC benchmark if a pupil was in a PRIMR school. The interpretation of 9.9 in public Class 1 for English is that pupils in PRIMR were 9.9 times more likely to be able to read at benchmark than pupils in control schools.

Figure ES3. Logistic regression results on the likelihood of reading at KNEC benchmark



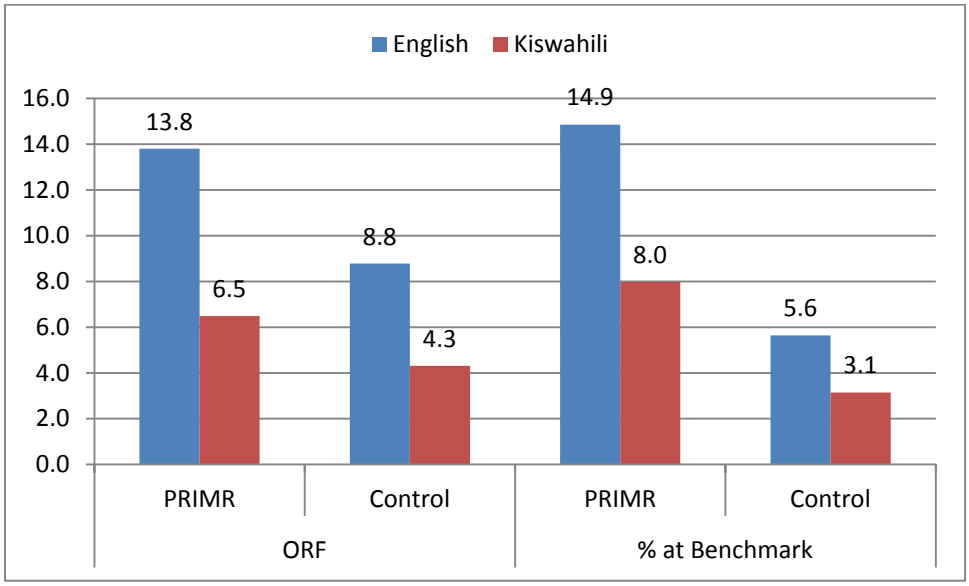
Cost-Effectiveness Analysis

Determining whether the impact of a program is worth the cost is essential for an impact evaluation to be truly useful. ***Figure ES4*** compares the cost of the PRIMR program to the estimated cost of the current system in the 2013 academic year, against the gains in outcomes achieved per 1 U.S. dollar. In order to understand how this works, recall that above we presented the gains between baseline and endline, comparing treatment and control. For Figure ES4, these gains were divided by the per pupil costs to create a cost-effectiveness estimate. The costs were derived from the expenditures[4] in PRIMR and the current system on key categories per pupil. The figure shows that PRIMR was much more cost-effective than the current system, at least against the estimated cost from available data. For example, PRIMR increased oral reading fluency for English by 13.8 cwpm per U.S. dollar, while the

[4] Expenditures include those accrued on the cost of pupil books, teachers' guides, teacher training, classroom observations, and TAC tutor training. The basic unit cost for PRIMR is therefore a ratio of the sum total of the total cost for program materials and training, divided by the number of pupils. This basic cost translated to an average $2.28 per pupil per subject, and is uniform for all zones implementing the full PRIMR.

current system increased oral reading fluency by 8.8 cwpm per dollar. The cost-effectiveness benefit of PRIMR is even more evident for the measure looking at the percentage of pupils reading at the KNEC benchmark. This shows that for both English and Kiswahili, the PRIMR treatment was more than two and nearly three times as cost-effective as the current system.

Figure ES4. Cost of PRIMR vs. the current system, 2013: Comparisons based on gains in oral reading fluency and pupils reading at benchmark



Lessons Learned

This section presents key lessons learned from PRIMR in a variety of key areas focused on quality improvement in Kenya's primary schools.

1. **Training for TAC tutors:** As the results show, TAC tutors' visits to schools are critical for supporting teachers and improving pupils outcomes. Proper training of TAC tutors is essential so that they can effectively support teachers. The results also indicated that schools that are visited frequently are likely to have stronger pupil performance; hence, TAC tutors should focus on making frequent and consistent classroom observations, even in the face of their heavy workload.
2. **Travel reimbursement structures:** PRIMR successfully facilitated TAC tutors to visit classrooms. This utilized a modest reimbursement that incentivized TAC tutors to visit classrooms consistently.
3. **Teacher training:** Training of teachers is a complex task that must assume that teachers are adult learners who learn best by doing and interacting with other professionals. This implies that teacher training should be organized around modeling and practice, and that having brief trainings with follow-up and refresher meetings is more effective than longer trainings.
4. **Distribution of classroom materials:** Distribution of materials to schools is a complex task. It requires accurate school enrollment data, prior planning, and a sophisticated distribution network. Ensuring that materials reach the schools on time was an essential PRIMR task.

5. **Priorities in the school calendar:** During the implementation of PRIMR it became apparent that certain times of the academic year required that the TAC tutor spend significant time away from the classroom. This occurred primarily during the extracurricular activity periods. These are clearly important for a balanced learning experience for pupils, but better understanding how these extracurricular activities could be organized so that they do not impede the TAC tutors' ability to support instruction is important.
6. **In-service training:** During PRIMR assessments and implementation, the evidence suggested that most of the teachers supported by PRIMR had not attended professional development courses or in-service courses for several years since leaving college or becoming teachers. The PRIMR Initiative's regular professional development through training and other activities filled a demand for increased instructional practice and support.
7. **Changes in instructional approaches:** Old habits take time to change, and the shift from traditional teaching to more active, sequenced, pupil-focused approaches was the central focus of PRIMR. Some teachers continued to use the two approaches concurrently at the beginning of PRIMR, in part because of concern about whether the lessons properly covered the material that would appear in the national end-of-year examinations. Advocacy was needed to change the mindset of some teachers.

Recommendations

Some recommendations from the endline assessment have already informed the program's final year of implementation. Others should be considered prior to a scale-up of PRIMR.

1. **Results and scale-up:** PRIMR's results to date have shown remarkable improvements in pupils' literacy and numeracy abilities, especially for pupils starting at the lowest levels of literacy and numeracy. The MoEST should therefore consider scaling up PRIMR activities to improve the quality of instruction in Class 1 and 2.
2. **Girls' performance:** The results indicated that, overall, girls were performing at the same level as—if not better than—boys, especially in literacy. Teachers should be trained in strategies for motivating girls so that they remain competitive as they move to upper primary.
3. **Zonal size:** The results showed that TAC tutors in large zones were less likely to have a significant an impact on pupil outcomes than those in smaller zones. Considerations should be made to limit the number of schools that the TAC tutors are responsible for. This would make TAC tutors more effective in supporting teachers frequently.
4. **Textbook ratio:** Provision of books to pupils at a 1:1 ratio is paramount in improving pupils' literacy and numeracy. The PRIMR analysis suggested that the government's current allocation would be enough to have a 1:1 ratio of books for all pupils in Kenya at low cost, if the cost of the books was more competitive.
5. **Advocacy and uptake:** There should advocacy of PRIMR's success through sharing of research results with a wider circle of stakeholders, including the MoEST and SAGAs.

6. **Language of instruction:** The language of instruction remains a complex issue for the Kenyan education system. Any attempt to scale up PRIMR activities without resolving this issue is likely to increase complexity during the implementation. The DFID PRIMR study, which is funding instructional materials and support in two mother tongues, will provide evidence as to the effectiveness of mother tongue compared with a basic instructional support program.
7. **Textbook policies:** The findings on cost and impact suggest that there is a need to consider the guidelines regarding vetting and selection of textbooks for use in schools. The complexity of multilingual literacy and numeracy instruction requires vetting guidelines that are tailored to the instructional characteristics of Kenya's system.
8. **Daily literacy and numeracy instruction:** Lesson time could be revised to accommodate more literacy and numeracy instructional time during the week. This is true not only because Kenya's literacy and numeracy allocations are paltry compared to the rest of East Africa, but also because of the evidence that in control schools, pupils spent very little time actually reading texts.
9. **Teacher assignments:** The transfer of teachers trained in PRIMR should be minimized to avoid the need for repeated onboarding and introductory training on a rolling basis. The TSC worked tirelessly to ensure that transfers were kept to a minimum, and we hope that can continue in the future.

1. Introduction

1.1 Background on the PRIMR Initiative

Since independence in 1963, the Kenyan government has identified education as a basic human right and the bedrock of national development (Ministry of Education, Science and Technology [MoEST], 2008). Key policy documents in Kenya, including the Education Act, the Sessional Paper #14 of 2012, and the draft National Education Sector Support Programme (NESSP) document make explicit references to improving literacy and numeracy outcomes at the lower primary levels. Similarly, the government has guaranteed every child the right to a free and compulsory basic education under section 53 of the new constitution. The government's commitment to education is also reflected in the relatively large proportion of funding that is set aside for education every year (MoEST, 2008).

Despite the achievements in educational equity and access in Kenya, previous research has consistently indicated that children in lower primary school do not have the requisite skills in literacy and numeracy (Piper, 2010; Piper & Mugenda, 2012; Uwezo, 2012). To achieve acceptable levels of literacy and numeracy among young learners in Kenya, the MoEST and the United States Agency for International Development (USAID) collaborated in 2011 to design the Primary Mathematics and Reading (PRIMR) Initiative. During the past three years, PRIMR has focused on improving numeracy and literacy outcomes in grades (Classes) 1 and 2 using a data-driven strategy in selected schools. Specifically, the initiative has aimed at helping teachers provide pupils with fundamental skills in literacy and numeracy.

PRIMR was designed to achieve the following objectives, the first two of which are directly evaluated in this endline report:

- Grade-appropriate reading fluency and comprehension increased for children in Classes 1 and 2;
- Grade-appropriate mathematical abilities increased for children in Classes 1 and 2;
- MoEST equipped and prepared to scale up successful features and approaches from the Early Grade Reading and Mathematics Assessments (EGRA/EGMA).

PRIMR has been fortunate to have worked closely with the MoEST, teachers and head teachers, pupils, and civil society. PRIMR has also depended on the expertise and leadership of experts in several organizations, including the Kenya Institute of Curriculum Development (KICD), Kenya National Examinations Council (KNEC), Kenya Education Management Institute (KEMI), Kenya Institute of Special Education (KISE), and Teachers' Service Commission (TSC). Fruitful collaboration with these organizations has been essential for PRIMR to have an impact on outcomes.

1.2 Program Components

In order to meet the objective of providing useful and actionable advice to the MoEST, RTI International used a randomized controlled design to rigorously measure the impact of PRIMR. Key activities focused on building teachers' capacity to deliver high-quality instruction in literacy and numeracy. This support included (1) providing teachers with models and practice using new and effective instructional strategies, (2) having coaches and TSC Teachers' Advisory Centre (TAC) tutors provide ongoing instructional support and

follow-up, (3) providing carefully designed learner books with a balance of literacy activities in a structured manner at a 1:1 ratio, and (4) providing teachers with teachers' guides with specific activities that match with pupil books. These components were the core elements of PRIMR's theory of change, and are essential to understanding the effectiveness and cost of PRIMR.

More broadly, the core activities during the implementation of the PRIMR initiative have been:

- Developing an experimental research design that included rigorous baseline, midterm, and endline assessments administered to randomly assigned treatment groups.
- Designing a scope and sequence of KICD curriculum-based content in Kiswahili, English, and mathematics.
- Preparing teachers' guides for Kiswahili, English and mathematics for Classes 1 and 2.
- Training teachers and head teachers to implement PRIMR lessons and use teachers' guides.
- Supporting regular supervision and monitoring of teachers by coaches and TAC tutors.
- Providing literacy and math books for pupils at a 1:1 ratio, matched with teachers' guides.
- Revising learning and teaching materials regularly, based on teachers' feedback, to make them more relevant and user-friendly.
- Training teachers to employ continuous assessment methods.
- Using EGRA and EGMA results to revise and update program materials.
- Carrying out policy studies to inform the MoEST on issues related to education quality and the policy reforms revisions needed to improve student outcomes.

1.3 Overall PRIMR Implementation Design

Learning materials: distribution to schools

As noted above, PRIMR is a partnership between USAID/Kenya and the MoEST, with technical implementation by RTI International. At the inception of PRIMR, a Program Development and Implementation Team (PDIT) was formed and charged with making key decisions and managing the overall direction of the program. The PDIT is led by Mrs. Margaret Murage of the MoEST, with other members drawn from key Semi-Autonomous Government Agencies (SAGAs), which include TSC, KISE, KICD, KNEC, and KEMI.

PRIMR is organized to test a cost-effective and scalable model's ability to improve literacy and numeracy among Class 1 and 2 pupils in Kenya. The design has the following elements:

- **Inexpensive books:** The project team made several decisions that would ensure savings on book purchases. For example, all elements of the literacy program—which for English and Kiswahili include phonics activities, illustrations, and decodable

stories—are embedded in the books. The books have attractive illustrations, which were initially in black and white, but have now been produced in color following KICD recommendations. The books utilized in the 2013 academic year and evaluated here cost US$0.75 each.

- **Basic instructional aids:** PRIMR provided simple instructional aids, including an A3-sized pocket chart and a set of letter and numeral flashcards. For the 2014 academic year that began in January, the teachers have been supported to make their own letter cards from heavy (manila) paper.
- **Self-contained teachers' guides:** The main resource for teachers was the teachers' guide. The math teachers' guides were produced in one volume while the English and Kiswahili materials were in two guides. Teachers were also given an assessment manual of less than 30 pages, some supplementary readers, a sheet of training tips, and a two-page document to track pupil progress; and pupils were given a single B5 sheet to track reading at home. These additional materials were discontinued in late 2013 but were part of the PRIMR program evaluated here. In the 2014 version of the materials, the first few lessons are scripted but the rest of the teachers' guide gives lesson outlines only, and the teachers' guides are contained in one volume.
- **Modest teacher training:** PRIMR decided to invest proportionally more resources in follow-up and observation than in traditional training, so the entire standard training program was only 10 days for the three subjects. This was allocated as five days at the beginning of Term 1 (January–April), three days at the beginning of Term 2 (May–August), and two days at the beginning of Term 3 (September–November).
- **Focused observations:** Much of PRIMR's attention and energy was spent in supporting TAC tutors and instructional coaches to visit schools and observe classrooms. Project funds reimbursed coaches' and TAC tutors' travel based on the proportion of teachers observed twice per month, to ensure that coaches and tutors had an incentive to provide equal support to distant or remote schools. The reimbursements were based on detailed observation forms that gave PRIMR the information needed to make program course corrections, matched with school logs signed by the head teacher. Like District Quality Assurance and Standards Officers (DQASOs), PRIMR's technical team spent time accompanying TAC tutors and coaches on their visits. In the 2014 academic year, these observation notes are recorded on tablets and the data shared to the project's cloud-based database.

If these individual elements were successful, PRIMR's theory of change suggested that program success was likely and should be identifiable in learning outcomes. This endline analysis evaluates this hypothesis.

1.4 Implementation in 2012

This section highlights specific aspects of program implementation during the first year, 2012, when the intervention details, logistics, and materials were being established. A core technical team comprising Kenyan PRIMR staff, the PDIT team, and subject specialists developed teachers' guides for Kiswahili, English, and math. The documents took into account the results of scope-and-sequence workshops that had taken place earlier to confirm the most pedagogically efficient chronology for introducing students to various skills and

concepts, and also involved regular consultations with the PDIT. The teachers' guides for English and Kiswahili were printed and distributed to schools and were ready for use in the intervention schools by January 2012, with the math teachers' guides and learner books distributed in July 2012.

The technical team that was working on the math teachers' guides drew on inputs from the ongoing USAID Liberia Teacher Training Program (LTTP). PRIMR began using the math teachers' guides in the latter part of Term 2 of 2012. This allowed the teachers to become acclimated to the PRIMR methods using the English and Kiswahili plans first, without being overwhelmed by content. Due to a three-week teacher strike during Term 3, however, the math program had been implemented for less than a full month when the midterm data collection teams began assessments in October 2012.

Based on feedback from PRIMR teachers, a revision of teacher and pupil materials in all three subjects took place in September through December 2012. A workshop held in December 2012 brought together a subject panel of MoEST officials, subject experts, consultants, and USAID representatives to finalize the material. The revised materials were ready for use at the beginning of the 2013 academic year.

The first coaches' and TAC tutors' trainings were held in early January 2012, followed by teacher trainings at the cluster and zonal level. The trainings in math took place in June and the teachers started using math teachers' guides at the end of the second term. The coaches and TAC tutors continued to support the teachers at the school level and collected observation data. Each coach or TAC tutor was required to visit each teacher twice a month to observe a lesson. The observation would be followed by a reflection discussion with the teacher on what went well, what should be improved, and how randomly selected pupils were performing.

Baseline assessments comprising EGRA (Kiswahili and English) and EGMA were also conducted in January 2012. A total of 4,385 pupils (2,199 girls and 2186 boys) randomly selected from 230 schools were assessed. The PRIMR midterm evaluation took place in October 2012, the end of the first school year of the intervention.

1.5 Implementation in 2013

In January 2013, PRIMR expanded from 126 treatment schools in 2012 to 311 treatment schools, with 121 schools remaining as control schools until January 2014. PRIMR was able to feed the results of the midterm assessment back into the intervention design to improve the program before implementation in January 2013.

Eight more coaches were hired to provide support to the additional 120 LCPS Cohort 2 schools that joined the program in January 2013. The coaches—as well as five additional TAC tutors—were inducted into PRIMR activities and the expected level and standard of support to teachers. Training was held during the second and third weeks of January 2013 for the new TAC tutors and coaches. It was emphasized to the new TAC tutors and coaches that supporting the teachers to deliver instruction accurately and efficiently to improve pupils' outcomes in both literacy and math was the coaches' and TAC tutors' key responsibility. Refresher trainings for the new and continuing TAC tutors and coaches took place in April and August 2013.

Reading Contest: winners receive prices

Other activities in 2013 included open-to-the-public reading and math contests within the PRIMR school clusters; teachers' monthly reflection meetings; PRIMR material review, revision, and development; and preparations for the endline assessment, which took place in October 2013. The revision of learner books and teachers' guides for English, Kiswahili, and math was undertaken alongside the development of mother-tongue pupil books and teachers' guides for Kikamba and Lubukusu. The mother-tongue books were developed under the PRIMR Rural Expansion program funded by the UK Department for International Development (DFID). The review of the materials in 2013 focused on rewriting stories; developing comprehension questions; introducing creative writing, listening, and speaking exercises; and enhancing various components of existing literacy and numeracy materials.

1.6 Challenges to Implementation of PRIMR

The PRIMR Initiative has had to mitigate a number of challenges since its inception in 2011. Among the most significant were the following.

1. The process of organizing the LCPS schools into clusters was quite complex and time consuming. The MoEST mapping documents did not include LCPS schools, and documents from other entities were not updated to include the entirety of the LCPS school population in Nairobi. PRIMR physically mapped over 1,000 schools on the ground using lists from other organizations and the snowball sampling method (i.e., expanding the knowledge base by networking and requesting additional contacts). This was followed by geographical clustering of the schools as a precursor to random selection and assignment.
2. As indicated earlier, public school teachers went on extended strikes twice during 2012–2013. The first instance was in September 2012, disrupting learning at the start of the third term. Thus, the public treatment schools received only limited PRIMR support before the midterm data collection. Teachers went on strike again in July 2013 and PRIMR activities were disrupted in public schools by nearly a month prior to the endline data collection.
3. In March 2013, learning was interrupted during the first term because of the national elections and the political activities that preceded it. Schools were closed for one week in March to allow the elections to take place, as most voting is done in schools. In addition, many of the PRIMR teachers and TAC tutors were heavily involved in politics prior to the election.
4. The initial uptake of PRIMR by some Class 1 teachers, particularly in public schools, was slow. It took the PRIMR team time and effort to convince the teachers, head teachers, and TAC tutors of the benefits of PRIMR.

5. TAC tutors found it difficult to support their assigned teachers while also handling the other duties for which they were responsible.
6. The program faced the challenge of significant teacher turnover. This usually happened in LCPS., where turnover of teachers typically is very high. The PRIMR team followed up such cases, sometimes training new teachers during the school term so that PRIMR activities could continue successfully.
7. PRIMR had difficulty organizing some of its activities that required the same time slots as MoEST and TSC extracurricular activities. This led to reduced lesson time for the pupils.

In spite of these challenges, over the three years, PRIMR saw high levels of uptake by teachers and head teachers, an increased demand for PRIMR, increased enrollment in PRIMR schools, and an ongoing enthusiasm for the program from the County Directorate of Education (CDE) and TSC offices. The PRIMR implementation team is thankful for the support of USAID/Kenya, the MoEST, and the SAGAs, as well as the county teams (TSC and CDE) that embraced a radically new approach to education quality in lower primary.

1.7 Assessment Tools

1.7.1 Early Grade Reading Assessment

The EGRA is an instrument for measuring learners' fundamental prereading and reading skills. The test is administered orally to pupils one on one, and takes approximately 15 minutes per child. Since EGRA was first piloted in 2007, RTI International, funding agencies, and experts in the field of early literacy have administered the EGRA in nearly 60 countries and over 100 languages. Pupils are selected randomly and assessed in various subtasks. In Kenya, the EGRA was administered to all students in both English and Kiswahili. ***Table 1*** shows the EGRA English subtasks utilized at the baseline, midterm, and endline.

Table 1. EGRA English subtasks implemented in PRIMR

English EGRA subtasks	Baseline, Jan. 2012	Midterm, Oct. 2012	Endline, Oct. 2013
1. Letter-sound fluency: ability to identify the sounds of the letters fluently	Done	done	Done
2. Decoding fluency: ability to decode new words fluently	Done	done	Done
3. Segmenting: ability to identify and sound out each sound present in a word	—	—	Done
4. Vocabulary: ability to tell the meaning of words	—	—	Done
5. Oral reading fluency: ability to read a story fluently	Done	done	Done
6. Reading comprehension: ability to comprehend reading passages associated with a timed reading assessment	Done	done	Done
7. Untimed reading fluency: ability to read a story fluently without timing	Done	—	—
8. Untimed reading comprehension: ability to comprehend reading passages associated with an untimed reading assessment	Done	—	—
9. Maze: ability to determine which of three words best fits as the missing word	—	done	—

As shown in ***Table 2,*** the EGRA Kiswahili subtasks were similar to the EGRA English subtasks, except that some subtasks differed based on the language assessed.

Table 2. EGRA Kiswahili subtasks implemented in PRIMR

Kiswahili EGRA subtasks	Baseline, Jan. 2012	Midterm, Oct. 2012	Endline, Oct. 2013
1. Letter-sound fluency: ability to identify the sounds of the letters fluently	Done	done	Done
2. Syllable fluency: ability to understand letters joined to form part of a word and to sound them appropriately	—	—	Done
3. Decoding fluency: ability to decode new words fluently	Done	done	Done
4. Oral reading fluency: ability to read a story fluently	Done	done	Done
5. Reading comprehension: ability to comprehend reading passages associated with a timed reading assessment	Done	done	Done
6. Untimed reading fluency: ability to read a story fluently without timing	Done	—	—
7. Untimed reading comprehension: ability to comprehend reading passages associated with an untimed reading assessment	Done	—	—
8. Listening comprehension: ability to understand a simple oral story read to the learner	Done	done	Done
9. Maze: ability to determine which of three words is the missing word	—	done	Done

1.7.2 Early Grade Mathematics Assessment

The EGMA focuses on measuring basic mathematical skills. This includes computational and number concepts and number sense. "Mathematics" is taken to be broader than arithmetic; it also encompasses non-operational number concepts. ***Table 3*** lists the PRIMR EGMA subtasks and their administration history.

Table 3. EGMA subtasks implemented in PRIMR

EGMA subtasks	Baseline, Jan. 2012	Midterm, Oct. 2012	Endline, Oct. 2013
1. Rational counting: ability to count accurately and fluently	Done	—	—
2. Number identification: ability to fluently identify numbers	Done	done	Done
3. Quantity discrimination: ability to fluently determine which of two numbers are larger, testing place value and number sense	Done	done	Done
4. Missing number: ability to identify missing numbers using knowledge and application of number pattern skills	Done	done	Done
5. Addition fluency: ability to add simple sums fluently, at lower levels of complexity	Done	done	Done
6. Addition level 2: ability to add simple sums fluently, at higher levels of complexity	Done	done	Done
7. Subtraction fluency: ability to subtract simple differences fluently, at lower levels of complexity	Done	done	Done
8. Subtraction level 2: ability to subtract simple differences fluently, at higher levels of complexity	Done	done	Done
9. Word problems: ability to solve basic word problems	Done	done	Done

1.7.3 Snapshot of School Management Effectiveness

The SSME contains a checklist that a trained observer uses to collect a wide range of classroom and school information. For PRIMR—at all three administrations—this included robust data on the languages used by the teacher during instruction and the nature of a teacher's interaction with students (e.g., was the teacher speaking to the class, a group, or a single student?). Additional information was obtained from Class 1 and 2 teachers and head teachers using SSME interview guides, inventories of classrooms and schools, and pupil questionnaires. Results from all of these variables were merged with the pupil outcome data to estimate the relationships between these school and classroom factors and pupil achievement, and, specifically, to determine whether those factors attenuated the relationship between the PRIMR intervention and pupil learning gains.

2. Research Design

2.1 Overall Research Design

Table 4 graphically presents the research design of the EGRA/EGMA assessments. The PRIMR design was organized to test the impact of a variety of literacy and numeracy strategies. This included an analysis of whether PRIMR was more effective in public or LCPS., what the most cost-effective ratio of schools to TAC tutors or instructional coaches was, and the relative impact of three different ICT interventions (described in Piper & Kwayumba, 2014).

Recall from the previous section that in 2012, 126 schools (Cohort 1) began the PRIMR intervention. In 2013, 311 schools (Cohorts 1 and 2) implemented the USAID-funded non-ICT aspects of PRIMR. In this report, we compare the outcomes for Cohorts 1 and 2 against those of the Cohort 3 control group as of October 2013.

Table 4. Implementation of PRIMR and EGRA/EGMA assessments

	2012		2013		2014
Cohort 1	125 schools 66 public, 59 LCPS				
Cohort 2			185 schools 65 public, 120 LCPS		
Cohort 3 (Control)					101 schools 51 public, 50 LCPS
Pupils Assessed	Baseline, January 2012	Midterm, October 2012		Endline, October 2013	
Cohort 1	1,335	1,320		1,300	
Cohort 2	1,860	1,850		1,876	
Cohort 3 (control)	1,190	992		1,046	
Total	4,385	4,162		4,222	

2.2 Treatment Groups

The USAID-funded PRIMR initiative is supporting 547 schools during the period 2011–2014. These schools were randomly selected and assigned to treatment and control groups

following strict statistical procedures. In the first sampling stage, zones were selected from counties, stratified by district. These zones were then randomly assigned to treatment groups. For the impact evaluation, for all selected zones, schools were randomly selected from within the selected zones for the assessments. It should be noted that, for the endline study, the schools sampled during the midterm study were maintained. This was determined to be appropriate after an investigation of the classroom visits of TAC tutors to schools selected for the baseline, and visit frequency did not differ. The third sampling stage involved systematic sampling of Class 1 and 2 pupils in the selected schools, stratified by gender and class.

The LCPS were divided into two categories. Half of the randomly selected schools were assigned to clusters with a school-to-coach ratio of 10:1 and the other half were assigned to clusters with a school-to-coach ratio of 15:1. Hence, in 2012, three treatment clusters had a school-to-coach ratio of 10:1 and two clusters had a school-to-coach ratio of 15:1. In 2013, 8 clusters were at 10:1 and 7 clusters were 15:1. Comparing the outcomes of pupils in these two groups was expected to help PRIMR advise the MoEST on the most cost-effective school-to-TAC tutor ratio that would significantly improve learners' outcomes in literacy and numeracy.

The PRIMR program also includes an ICT component in Kisumu County. The ICT component uses a randomized controlled design to compare the effectiveness and cost-effectiveness of three different ICT interventions. The Kisumu program started with a baseline survey that was undertaken in January 2013. The selection and assignment of zones followed generally the same procedure as it did for the formal schools in the basic PRIMR model. Stratifying by urbanicity, zones were randomly selected and then assigned to the three ICT treatment conditions (pupils with an e-reader, teachers with a tablet loaded with relevant resources, TAC tutors with a tablet). The results of the evaluation of the PRIMR ICT intervention are presented in a separate report (Piper & Kwayumba, 2014).

PRIMR's design treats control schools ethically. A control group of 51 government schools in Nairobi, Nakuru, and Kiambu counties was selected, as was a control group of 50 LCPS.. To meet established research ethical standards, in January 2014, all schools selected for the control groups at each intervention level started receiving PRIMR activities that had been recognized as most cost-effective.

2.3 Formal and LCPS Schools

Formal schools in Kenya are supported by the government through provision of learning materials, teachers, and infrastructure. In each district within a county, schools fall within zones. The TSC's TAC tutor system was established as a support system for teachers within each zone. The TAC tutors exist to provide instructional support to teachers, to improve the quality of instruction at the classroom level. The number of schools in each zone for which an individual tutor is responsible ranges from about 8 to 20 schools. The distances between most of the schools in rural areas are substantial, and with limited resources, the TAC tutor system struggles to support teachers at full capacity, such that very little classroom instructional support occurs in most zones. PRIMR has been working through this TAC tutor system to determine how it can focus more on instruction, and to determine the impact on instruction of individuals already within the formal system.

A critical issue is the lack of resources within the public sector to provide education for school-age children. The private sector has therefore supplemented the government's efforts through private schools, but the cost of education in private institutions is too high for the majority of Kenyans. This has led to the mushrooming of low-cost private schools, especially in urban informal settlements. Generally, low-cost private schools target primary-school-age children, using the KICD curriculum with the support of nongovernmental organizations (NGOs), communities, faith-based organizations, and individual investors. These schools are characterized by relatively low tuition, poor infrastructure, frequent transfer of pupils from one school to another, lack of adequately trained staff, and high teacher turnover. The managerial operations of these schools are not standardized by any government policy or agency. Nevertheless, the data from this evaluation showed that LCPS in Nairobi were performing at the same level as, if not better than, public schools. Learning typically is more focused in these schools despite deplorable conditions, and in recent times, the poor often have chosen these schools over public schools. PRIMR recently undertook a survey of parents in urban Nairobi who send their pupils to public and low-cost private schools to better understand the mechanisms for their schooling choices.

2.4 Sample

At the baseline assessment, zones and clusters were randomly selected from the counties or regions that were agreed upon among the MoEST, RTI, and USAID. Approximately 50 percent of the schools were picked for assessment in the selected zones and clusters. At midterm, the same procedures were followed and approximately 50 percent of the schools were again randomly selected, but only for Cohort 2. It is noted here that schools selected for assessment in Cohort 1 and 3 were maintained at midterm because the pupils selected from these schools formed a longitudinal sample of specific students who would be traced and assessed at all three time points. For the endline study, it was decided that the same sample of schools used at midterm should be maintained for endline, given the failure of the school selection process to influence the number of classroom visits made by TAC tutors and coaches.

The total number of schools assessed during the endline was 214. In all, 4,222 pupils were assessed at the endline compared to 4,166 pupils at midterm and 4,385 pupils at baseline. Based on PRIMR's power calculations, this sample size was considered sufficient to detect an impact of at least 0.20 standard deviations (SD) or more. ***Table 5*** shows the pupils assessed at endline disaggregated by gender and class.

Table 5. Endline sample size by gender and class

Gender	Class 1	Class 2	Totals
Girls	1,052	1,054	2,116
Boys	1,058	1,058	2,106
Totals	2,110	2,112	4,222

2.5 Training and Data Collection

Assessing a pupil using a tablet

RTI has a corps of experienced data assessors in Kenya who have been engaged in collecting assessment data in schools using EGRA and EGMA tools since 2007. Since the midterm in October 2012, PRIMR assessors have been using RTI-developed open-source Tangerine® data collection tools on tablets. For the endline assessment, assessors were trained for a week during September 2013. A total of 124 assessors were trained in Nairobi (52 for the USAID PRIMR and 77 for the parallel DFID PRIMR Rural Expansion study). During the training, the assessors were instructed in how to administer the EGRA (Kiswahili and English) and EGMA, and also evaluated on their ability. Assessor reliability tests were done for both EGRA and EGMA. The interrater reliability scores for all three subjects (English, Kiswahili, and math) were above 93% at the endline. This shows that the assessors' consistency in assessing pupils was very high.

The assessors were grouped into 12 teams of three assessors and one supervisor. On each team, the most experienced assessor was appointed as the supervisor. The supervisors were given further training in how to conduct classroom observations and teacher and head teacher interviews, which were always carried out at each school along with the student assessments. At the end of each data collection day, data were uploaded from tablets to the project's cloud-based database. Missing data were identified and, where necessary, the missing assessments repeated at the school level.

3. Reliability Estimates

A reliability analysis was conducted to determine the appropriateness of the subtasks within the EGMA and EGRA tools. Pearson correlation coefficients were computed among the subtasks in each tool, as discussed in the sections that follow. Ideally, strong correlations among subtasks are preferred because they indicate consistency in the performance of the sampled learners across the subtasks.

3.1 English Tool Analysis

All the correlations for the EGRA (English) subtasks were statistically significant ($p < 0.001$), as shown in ***Table 6.*** It is interesting to note that the correlations between the segmenting subtask (identifying the sounds in words, read to the pupil twice) and all other subtasks were somewhat lower, which indicates that this subtask was assessing a different set of skills compared to the other subtasks. There was a moderate correlation between letter sounds and invented words; letter sounds and oral reading; and invented words and vocabulary. The results show relatively strong correlations between invented words and oral reading; invented words and reading comprehension; oral reading and vocabulary; oral

reading and reading comprehension; and vocabulary and reading comprehension. Given the design of these subtasks, these correlations were logically expected. Moreover, the results indicate that the pupils who could read invented words also were able to read passages and comprehend to a reasonable degree.

Table 6. Pearson correlations for EGRA subtasks in English

	Letter-sound fluency	Decoding fluency	Segmen-ting	Oral reading Fluency	Vocabulary	Reading compre-hension
Letter-sound fluency	1.00					
Decoding fluency	0.56***	1.00				
Segmenting	0.47***	0.34***	1.00***			
Oral reading fluency	0.50***	0.86***	0.30***	1.00		
Vocabulary	0.40***	0.57***	0.37***	0.63***	1.00	
Reading comprehension	0.41***	0.62***	0.29***	0.74***	0.63***	1.00

*p < 0 .05, **p < 0.01, ***p < 0.001.

Table 7 shows the internal consistency of each EGRA English subtask, as assessed using Cronbach's alpha. The overall Cronbach's alpha for the EGRA (English) tool was 0.86, while the coefficient for all the subtasks was above 0.80, which is considered high. These results show that the English subtasks were able to distinguish poor-performing and high-performing pupils to a high degree.

Table 7. Cronbach's alpha for EGRA subtasks in English

Subtask	Item-test correlation	Item-rest correlation	Alpha
Letter-sound fluency	0.72	0.59	0.85
Decoding fluency	0.85	0.77	0.82
Segmenting	0.60	0.43	0.88
Vocabulary	0.78	0.67	0.84
Oral reading fluency	0.87	0.79	0.81
Reading comprehension	0.80	0.69	0.83
Totals			**0.86**

3.2 Kiswahili Tool Analysis

Reliability analyses were conducted for the Kiswahili subtasks and the results are presented in ***Table 8.*** Results show that the correlations were all statistically significant (p < 0.001). There were strong correlations between syllable and invented words; syllable and reading fluency; invented words and reading fluency; invented words and reading comprehension; and reading fluency and reading comprehension. The results show moderate correlations between letter sounds and invented words; letter sounds and reading fluency; invented words and maze; reading fluency and maze; and reading comprehension and maze. Moderately strong correlations were observed between letter sounds and syllable fluency and between

syllable fluency and reading comprehension. However, the listening comprehension subtask showed weak correlation with all the other subtasks, which is evidence that this subtask assessed a different skill set from that of the other subtasks.

Table 8. Pearson correlations for EGRA subtasks in Kiswahili

	Letter-sound fluency	Syllable fluency	Decoding fluency	Oral Reading fluency	Reading compre-hension	Listening compre-hension	Maze
Letter-sound fluency	1.00						
Syllable fluency	0.65***	1.00					
Decoding fluency	0.55***	0.82***	1.00				
Oral reading fluency	0.53***	0.77***	0.86***	1.00			
Reading comprehension	0.49***	0.68***	0.75***	0.84***	1.00		
Listening comprehension	0.29***	0.39***	0.37***	0.42***	0.48***	1.00	
Maze	0.34***	0.46***	0.54***	0.57***	0.56***	0.30***	1.00

*p <0 .05, **p < 0.01, ***p <0.001.

Table 9 shows the Cronbach's alpha coefficients for the EGRA Kiswahili subtasks. All the coefficients are above 0.85, which is considered high. The overall alpha coefficient is 0.90. These results show a high consistency across the EGRA Kiswahili subtasks.

Table 9. Cronbach's alpha for EGRA subtasks in Kiswahili

Subtask	Item-test correlation	Item-rest correlation	Alpha
Letter-sound fluency	0.70	0.59	0.89
Syllable fluency	0.86	0.81	0.87
Decoding fluency	0.89	0.84	0.86
Oral reading fluency	0.90	0.86	0.86
Reading comprehension	0.87	0.81	0.87
Listening comprehension	0.59	0.45	0.91
Maze	0.69	0.57	0.89
Totals			**0.90**

3.3 Math Tool Analysis

The EGMA tool had eight subtasks: number identification, quantity discrimination, missing number, addition fluency and level 2, subtraction fluency and level 2, and word problems. The pairwise correlations for all subtasks were all statistically significant ($p < 0.001$), as shown in ***Table 10.***

Number identification was moderately correlated with quantity discrimination and other subtasks. Similarly, quantity discrimination was moderately correlated with other subtasks but strongly correlated with missing number. These results seem to imply that number identification and quantity discrimination were not strong predictors of pupil performance in higher-level subtasks such as addition, subtraction, and word problems. The correlations

between word problems and the other subtasks were weak, meaning that word problems require a different skill set from the other subtasks. Subtraction level 2 was moderately strongly correlated with addition level 2 and moderately correlated with subtraction fluency. Subtraction fluency was moderately correlated with the other subtasks, implying that pupils who could solve subtraction sums at that level could also perform the other subtasks at almost the same level. These findings are similar to the results from the midterm evaluation (Piper & Mugenda, 2013).

Table 10. Pearson correlations for EGMA subtasks

	Number identification	Quantity discrimination	Missing number	Addition fluency	Addition level 2	Subtraction fluency	Subtraction level 2	Word problems
Number identification	1.00							
Quantity discrimination	0.57***	1.00						
Missing number	0.55***	0.65***	1.00					
Addition fluency	0.57***	0.55***	0.58***	1.00				
Addition level 2	0.44***	0.45***	0.51***	0.59***	1.00			
Subtraction fluency	0.51***	0.50***	0.53***	0.69***	0.53***	1.00		
Subtraction level 2	0.36***	0.34***	0.41***	0.44***	0.63***	0.51***	1.00	
Word problems	0.30***	0.36***	0.40***	0.42***	0.43***	0.45***	0.42***	1.00

*$p < .05$, **$p < 0.01$, ***$p < .001$.

The Cronbach's alpha coefficients for the math subtasks ranged between 0.86 and 0.88 ***(Table 11).*** The overall Cronbach's alpha coefficient for the math tool was 0.88. Alpha coefficients that fall within this range are considered high and indicate consistency among the items in the subtasks. This means that the items within each subtask were able to effectively discriminate high-performing pupils from low/poor-performing pupils with regard to their numeracy and computational skills.

Table 11. Cronbach's alpha for EGMA subtasks

Subtask	Item-test correlation	Item-rest correlation	Alpha
Number identification	0.72	0.62	0.87
Quantity discrimination	0.75	0.65	0.87
Missing number	0.78	0.70	0.86
Addition fluency	0.81	0.74	0.86
Addition level 2	0.77	0.69	0.86
Subtraction fluency	0.79	0.72	0.86
Subtraction level 2	0.69	0.58	0.87
Word problems	0.63	0.51	0.88
Totals			**0.88**

3.4 Equating Procedures

The initial EGRA, EGMA, and SSME tools were discussed with the MoEST and piloted in November 2011. The tools were revised based on the results of the pilot tests prior to the January 2012 baseline. Eventually, PRIMR selected three EGRA assessment tools for each language (English and Kiswahili)—that is, the structure was exactly the same in each instrument, but the specific content within the subtasks differed across the three test forms. These were supposed to be administered at baseline, midterm, and endline. However, the original Kiswahili and English stories that were to be used at endline were later substituted with new stories that were judged to be more reliable measures.

The new stories in Kiswahili and English were picked from a pool of stories initially developed at the start of PRIMR. Although these stories were evaluated before they were used, it was still necessary to conduct an equating exercise that compared the new stories with the stories used at baseline. In other words, it was desirable to determine whether the baseline sample's scores on these new tests would have been similar to those of the endline population. If there were slight differences, those differences could then be adjusted statistically.

Nine experienced research assistants were trained for half a day on proper administration of EGRA (English and Kiswahili) tools using the tablets. Data collection was undertaken in one day with 126 pupils assessed in the two languages. The pupils were picked at random from three different PRIMR schools in Kisumu County. Data were analyzed and the results used to calculate equating coefficients for the endline oral passages. The equating coefficient was obtained by dividing the baseline oral reading fluency mean score by the endline oral reading fluency mean score for English and Kiswahili stories (i.e., *equating coefficient = baseline fluency mean score ÷ endline fluency mean score*). Each pupil's score in English and Kiswahili in the main endline assessment was multiplied by the respective coefficient to obtain the equated scores.

4. Endline Study Findings: Impacts of PRIMR

4.1 Descriptive Statistics: Treatment and Control Schools

In this section of the report, we present the mean scores for each subtask for the PRIMR and control groups. Analyses presented in ***Table 12*** through ***Table 14*** convert the differences between the PRIMR and control schools into program impacts, and then use the pooled standard deviation of each of the tasks to produce a PRIMR impact effect size.[5]

For the PRIMR English program, it is worth noting that the classroom materials were designed to support the pupils in their ability to understand how English and Kiswahili relate, such that pupils would be introduced to letters and sounds first in Kiswahili and then in English. This allowed the program to follow established research findings and the pupils to have an efficient manner of learning to read and comprehend English. In Class 1, the first

[5] This is a more conservative way of calculating effect size than the methods used by many other organizations in the education sector, and more conservative than some other programs in which RTI is involved, such as the Liberia Teacher Training Program. However, the PRIMR team is convinced that this is the correct way to measure program effect in Kenya.

seven weeks of English lessons were oral, and only in Week 8 were pupils expected to know the English sounds. The English program systematically taught pupils all of the basics of literacy according to the KICD syllabus, including listening, speaking, reading, and writing.

Table 12 shows that pupils in PRIMR achieved higher scores than those in control schools on each of the English tasks, in Class 1 and 2 as well as overall. For letter-sound fluency, pupils receiving the PRIMR intervention read 47.0 letters per minute (clpm) correctly, compared to a mean score of 25.7 letters per minute among control pupils. PRIMR's causal effect, therefore, was 21.3 clpm. This equates to an effect size of 0.73 standard deviations. Effects were somewhat higher in Class 2 (0.78 SD) than in Class 1 (0.68). Impacts were somewhat larger in Class 2, overall (0.49 SD) than in Class 1 (0.47 SD).

Substantively, in the area of oral reading fluency, the PRIMR effect was 13.7 cwpm overall, and 12.1 cwpm in Class 1, and 16.1 cwpm in Class 2. This equates to more than 1 year of normal gain in Kenya, and this size of a causal effect is some of the largest seen in any program that the authors are aware of. Reading comprehension scores were more than twice as high in PRIMR schools (21.1%) as in control schools (9.8%) in Class 1, and the absolute gain in comprehension attributed to PRIMR in Class 2 was 17.3%. Interestingly, even in items that were not previously assessed, such as segmenting and vocabulary, PRIMR effects were significant, and in the case of segmenting, quite large (0.62 SD).

The proportion of pupils reading at benchmark was more than twice as high in PRIMR (28.3%) as in control schools (12.6%). That figure held true for both Class 1 and Class 2. In short, the PRIMR Initiative had a significant impact on outcomes in English. Gains were higher in the endline assessment than they were at midterm (Piper & Mugenda, 2013), even though the number of schools being supported nearly tripled, reducing the amount of support per school by PRIMR staff. This suggests that the strong impact was due not to an increased amount of time spent by the PRIMR staff, but to the program itself, as implemented by the TSC TAC tutors and MoEST education staff, which was successful at improving English outcomes. Effect sizes were moderate to large across subtasks, with an average overall effect size of 0.47 SD in Class 1 and 0.49 SD in Class 2 SD. This shows a remarkable impact on learning outcomes using a cost-effective and sustainable set of strategies.

Table 12. Endline impact of PRIMR on English outcomes (all measures)

	Overall			Class 1			Class 2		
English EGRA subtasks	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*
Letter-sound fluency (correct letters per min.)	**47.0**	25.7	*0.73*	**43.5**	24.6	*0.68*	**50.8**	26.8	*0.78*
Decoding fluency (correct nonwords per min.)	**28.4**	20.7	*0.41*	**23.1**	15.3	*0.46*	**34.0**	26.1	*0.42*
Segmenting (% correct out of 10 items)	**55.6**	31.3	*0.62*	**54.4**	29.9	*0.64*	**56.8**	32.7	*0.61*
Oral reading fluency (correct words per min.)	**45.1**	31.4	*0.40*	**32.2**	20.1	*0.44*	**58.9**	42.8	*0.45*

English EGRA subtasks	Overall PRIMR	Overall Control	Overall *Effect size*	Class 1 PRIMR	Class 1 Control	Class 1 *Effect size*	Class 2 PRIMR	Class 2 Control	Class 2 *Effect size*
Vocabulary (% correct out of 20 items)	**66.8**	60.7	*0.30*	**62.6**	54.6	*0.40*	**71.4**	66.7	*0.24*
Reading comprehension (# correct out of 5 questions)	**34.3**	19.4	*0.38*	**21.1**	9.8	*0.38*	**48.4**	29.1	*0.44*
Reading at benchmark (% of pupils reading 65 wpm+)	**28.3**	12.6	*0.36*	**14.0**	4.0	*0.32*	**43.7**	21.3	*0.45*
Average effect size			***0.46***			***0.47***			***0.49***

The PRIMR approach is organized to support pupils' acquisition of Kiswahili as a primary language of literacy engagement. Pupils are systematically taught two letters per week from the beginning of Class 1, and by the end of Class 1 are able to read any sound and decode the key words in Kiswahili. The Kiswahili program emphasizes vocabulary development and comprehension strategies and engages the learner across the spectrum of learning outcomes expected by the KICD syllabus.

This section and Table 13 below compare outcomes for pupils who used the PRIMR Kiswahili materials and those who did not, as randomly selected and assigned. The results for Kiswahili letter-sound fluency show that the PRIMR effect was 15.6 clpm for Class 1 and 22.1 clpm for Class 2. The overall effect size for letter-sound fluency was 0.63 SD. Surprisingly, while the average classroom in Kenya spends a great deal of time on syllable fluency, the PRIMR program still showed a 0.41 SD effect on syllable fluency. In Class 2, that equates to 11.9 more syllables correctly identified. PRIMR effects on oral reading fluency were 7.0 cwpm (0.41 SD) in Class 1 and 6.7 cwpm (0.35 SD) in Class 2. While these gains were somewhat more modest than expected, the reading comprehension score gains were larger, with a 0.45 SD effect in Class 1 and a 0.32 SD effect in Class 2. Listening comprehension and maze effects were relatively small, in part because the maze scores were so low overall and because most pupils did relatively well on the Kiswahili listening comprehension task.

Critically for the proportion of pupils reading at the benchmark, scores in PRIMR were 9 times larger in Class 1 (0.28 SD) and two times larger in Class 2 (0.30 SD). Overall, the effect of PRIMR in Kiswahili was 0.39 SD in Class 1 and 0.36 SD in Class 2. Outcomes were much higher than at the midterm assessment, and much higher than expected given the challenges of implementation in 2013. Overall effect sizes for Kiswahili were 0.39 for Class 1 and .36 for Class 2, very large gains for a program of this size.[6]

[6] Cohen (1988) cautioned researchers about the overuse of terms such as large, medium and small in explaining the magnitude of effect sizes. Glass, McGawnd and Smith (1981) acknowledged the danger of using such terms out of context. They argued that the effectiveness of a particular intervention can only be interpreted in relation to other interventions that seek to produce the same effect. In addition, the practical importance of an effect depends entirely on its relative costs and benefits. (Hill et al, 2007) argue that effect size should however, be

Table 13. Endline impact of PRIMR on Kiswahili outcomes (all measures)

	Overall			Class 1			Class 2		
Kiswahili EGRA subtasks	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*
Letter-sound fluency (correct letters per min.)	**47.5**	28.8	*0.63*	**42.4**	26.8	*0.57*	**52.9**	30.8	*0.70*
Syllable fluency (correct syllables per min.)	**45.7**	34.6	*0.41*	**38.4**	27.6	*0.42*	**53.3**	41.4	*0.45*
Decoding fluency (correct nonwords per min.)	**21.9**	16.5	*0.33*	**17.1**	11.6	*0.38*	**26.8**	21.3	*0.33*
Oral reading fluency (correct words per min.)	**27.4**	20.6	*0.35*	**20.9**	13.9	*0.41*	**34.0**	27.3	*0.35*
Reading comprehension (% correct out of 5 questions)	**35.9**	25.8	*0.34*	**25.6**	14.9	*0.45*	**46.6**	36.5	*0.32*
Listening comprehension (% correct out of 3 questions)	**63.5**	56.7	*0.25*	**57.5**	49.2	*0.31*	**69.8**	64.1	*0.22*
Maze (% correct out of 17 items)	**20.1**	16.8	*0.23*	**15.7**	12.2	*0.31*	**24.7**	21.4	*0.22*
Reading at benchmark (% of pupils reading 65 wpm+)	**15.9**	6.7	*0.27*	**7.2**	0.8	*0.28*	**24.9**	12.5	*0.30*
Average effect size			***0.35***			***0.39***			***0.36***

The PRIMR mathematics program showed modest positive results at midterm. However, given the very limited amount of time that the books and teachers' guides were in classrooms, the researchers were not convinced that any positive effect for math was due to the program, so they did not claim a significant impact on outcomes. In 2013, however, the math materials were in schools on time and the training between literacy and numeracy was done concurrently.

Table 14 below presents the impact of PRIMR on mathematics outcomes at the October 2013 endline. It shows a modest effect of PRIMR on math, of 0.16 SD for Class 1 and 0.26 SD for Class 2. PRIMR seemed to improve outcomes on number identification (0.27 SD), and missing number (0.29 SD), but had no effect on quantity discrimination (0.03 SD). The computational measures showed some effect, with higher outcomes in addition fluency, addition level 2, subtraction fluency, and subtraction level 2. For the most part, the impact was larger in Class 2 than it was in Class 1, and sometimes the impact was statistically

interpreted with respect to empirical benchmark that are relevant to the intervention, target population and outcome measure being considered.

insignificant in Class 1.Word problems showed a small impact (0.13 SD), once again larger in Class 2 than in Class 1.[7] This is evidence of a moderate positive impact of the PRIMR math program on pupil achievement in math.

Table 14. Endline impact of PRIMR on mathematics outcomes (all measures)

	Overall			Class 1			Class 2		
	PRIMR	Control	*Effect size*	**PRIMR**	Control	*Effect size*	**PRIMR**	Control	*Effect size*
Number identification (correct numbers per min.)	**24.5**	21.3	*0.27*	**19.6**	16.7	*0.31*	**29.6**	25.7	*0.33*
Quantity discrimination (% correct comparisons)	**59.9**	59.2	*0.03*	**48.4**	44.6	*0.16*	**72.0**	73.0	*-0.04*
Missing number (% correct)	**43.5**	36.8	*0.29*	**32.8**	28.6	*0.23*	**54.7**	44.6	*0.45*
Addition fluency (correct items per min.)	**10.1**	9.3	*0.17*	**7.9**	7.5	*0.10*	**12.4**	10.9	*0.33*
Addition level 2 (% correct)	**34.2**	26.8	*0.23*	**20.7**	17.7	*0.12*	**38.3**	35.4	*0.08*
Subtraction fluency (correct items per min.)	**7.1**	6.2	*0.21*	**5.4**	4.7	*0.18*	**8.9**	7.5	*0.34*
Subtraction level 2 (% correct)	**21.9**	15.4	*0.24*	**13.1**	11.3	*0.09*	**31.1**	19.2	*0.38*
Word problems (% of 5 items correct)	**40.7**	37.4	*0.13*	**33.9**	31.6	*0.10*	**47.8**	42.9	*0.18*
Average effect size			***0.20***			***0.16***			***0.26***

Recall that 2012 saw very little time for PRIMR mathematics to be implemented prior to the midterm analysis. The 2013 academic year was, for all intents and purposes, the first year that the PRIMR mathematics intervention was fully implemented. The gains indicated in Table 14 and echoed (but more modestly) in a set of differences-in-differences (DID) analyses in ***Annex A*** suggest that the PRIMR mathematics instructional program has real promise for improving the quality of mathematics outcomes in Kenya. Further analysis of PRIMR mathematics from the DFID-funded Rural Expansion schools at the end of 2014 will reveal the impact of PRIMR mathematics after two years of implementation.

In ***Annex B*** we present disaggregated analyses of the PRIMR effect on English, Kiswahili, and math. Annex B presents PRIMR program effects and effect sizes disaggregated by type of school for Class 1 (Table B1) and Class 2 (Table B2). Table B3 in the same annex shows

[7] Note that some of the impacts identified in differences-in-differences analyses (see ***Annex A***) were smaller than the ones presented in this section of the report.

program effect and effect sizes for Kiswahili and English disaggregated by gender, while the math program effects and effect sizes disaggregated by gender are shown in Table B4. Annex B shows that the PRIMR effect was sustained for both types of schools and both genders.

As indicated, Annex A presents our analysis of the comparison of the results using (1) simple ordinary least squares (OLS) regression comparisons between the treatment and control at the endline, and (2) differences-in-differences estimates. Figure A1 in this annex shows a comparison between the DID estimate of PRIMR effect and the simple comparison at the endline on the proportion of pupils reading fluently using the two measures. Figure A2 shows the comparison between the DID estimates and simple differences at endline on addition fluency. In summary, Annex A shows why we chose to primarily use the simple differences rather than DID estimates.

4.2 Analytic Strategy

The multiple assessments and the random selection and assignment strategy allowed the researchers to estimate whether there were any differences in outcomes at the endline. The design also allowed for causal inferences of the impact of the PRIMR intervention.

The figures below primarily present the causal impact of PRIMR as calculated from simple regression analyses at the endline. In some parts of the report we present the differences-in-differences estimate (the coefficient on the endline multiplied by the PRIMR interaction, ignoring the midterm results). We were interested in determining whether regression analyses would reveal significant differences in the outcomes, and what the implications of including a set of control variables would be. For the endline analysis, we considered it essential to include control variables in our regression models, given the growing evidence that several key variables were strongly associated with pupil outcomes. In order to create a parsimonious model, we began with a set of control variables (gender, age, parental literacy, wealth factors, a composite of pupil wealth, reading material availability, textbook ratio), and determined whether they were correlated with the outcomes of interest as well as with each other. Our final models presented in this section include controls for gender and a composite of the number of household wealth indicators that the pupils reported to have. Both of these covariates were frequently statistically significantly predictive of the outcome variables, even in models that measured PRIMR impact. Given the previous research that PRIMR project team members have done to show that poverty and other factors matter quite a bit (Piper, Jepkemei & Kibukho, in press), we were interested in whether we should include both controls in the analysis of the impact of PRIMR.

In this section we present only the gains made over the baseline scores. This is different from the analyses above, which looked only at the outcomes of PRIMR and control schools at the endline, ignoring any differences at the baseline. The graphs below were derived from fitted differences-in-differences multiple regression models, with control variables, which removes the baseline scores. The gray bars represent the gains of PRIMR treatment pupils on this task over the baseline scores, while the white bars represent the gains of the control pupils on a given subtask over the baseline. These graphs are included to ground-truth the notion that PRIMR increased outcomes over the baseline.

Figure 1 shows the changes since the baseline for Class 1 pupils. For English letter-sound fluency, the gain for treatment pupils was 21.3 letters, compared to 7.6 letters for control pupils. This shows a causal effect of 13.7 clpm. Stated another way, this also shows that, for letter-sound fluency, treatment pupils gained nearly three times as much as did the control pupils. Figure 1 shows that gains since the baseline were larger for PRIMR than for control on all tasks, both English and Kiswahili. Moreover, the gains were much larger for PRIMR than for control for the key measures of the percentage of pupils reading at benchmark and the percentage of pupils comprehending 80% or more. For these outcome variables, PRIMR's treatment results were three or four times higher in Class 1.

Figure 1. Gains since baseline for PRIMR and control in Class 1, English and Kiswahili

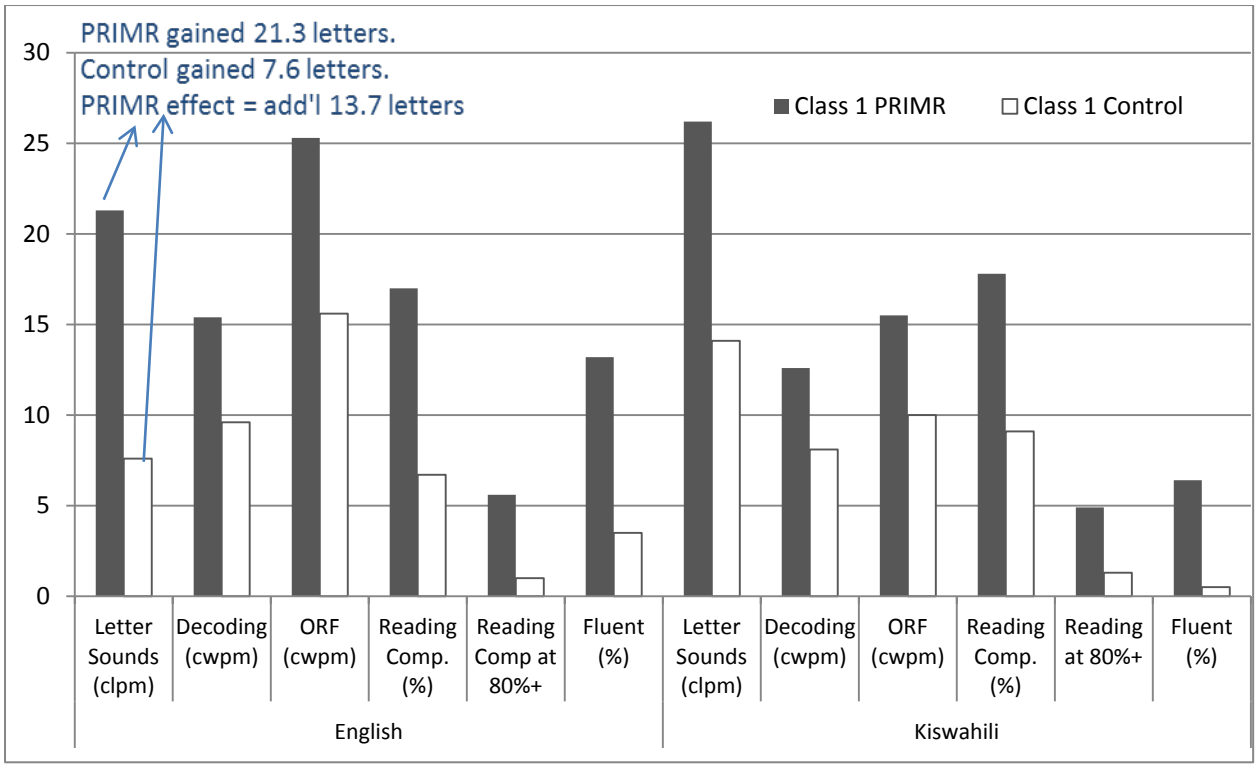
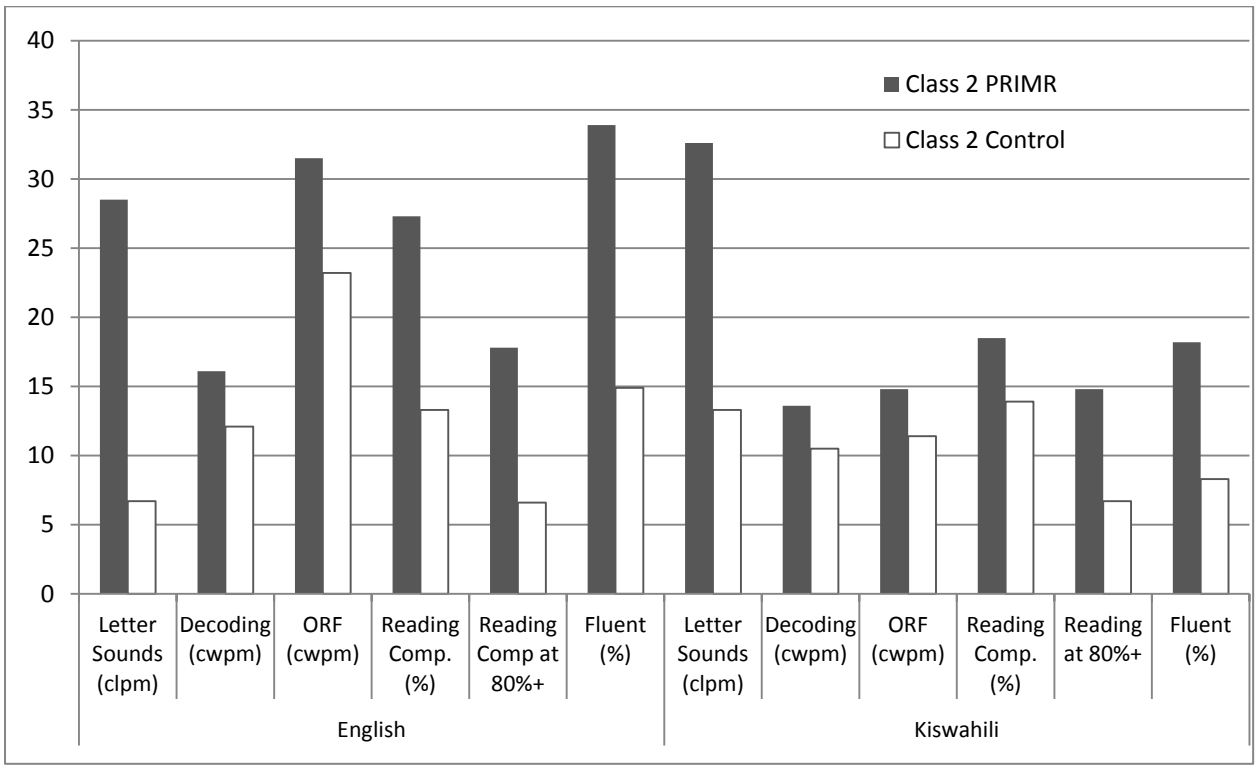


Figure 2 presents the gains since the baseline for PRIMR and control pupils in Class 2 across all subtasks measured at both baseline and endline, similar to the figure above for Class 1. Results from the differences-in-differences models for Class 2 showed statistically significant improvements by PRIMR for all subtasks in both English and Kiswahili. The magnitude of the PRIMR impacts differed by subtask, with relatively modest impacts on decoding fluency, but much larger impacts on the percentage of pupils able to read at the KNEC benchmark and the percentages of pupils comprehending 80% or better. This figure shows the magnitude of the PRIMR effect, which can be understated in a simple examination of the endline results. Given that Kenya's education system saw some achievement from control pupils even in Class 1, it is important to compare how much additional learning occurred both within and beyond the program since its inception. In Kenya, particularly in Class 2, these pupils were learning. Oral reading fluency gains were more than 20 and 10 cwpm in English and

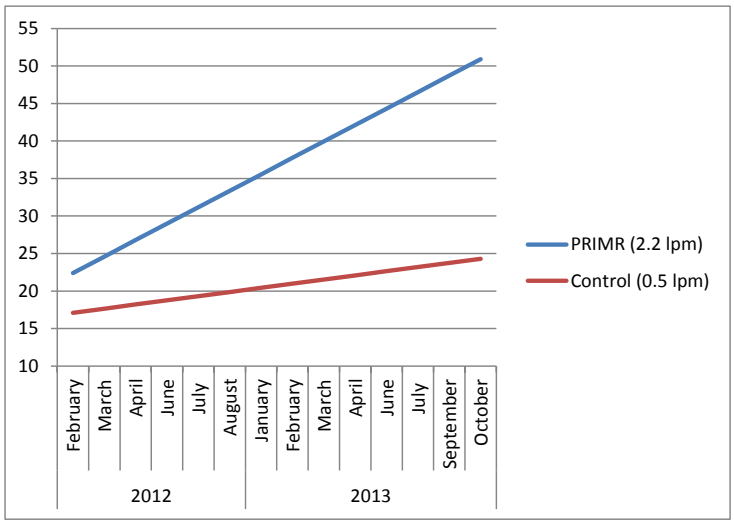
Kiswahili, respectively. So to be statistically significant, PRIMR's results would have to be quite large. Figure 2 shows that the results were, in fact, quite substantial for treatment pupils in Class 2.

Figure 2. Gains since baseline for PRIMR and control in Class 2



4.3 Rate of Improvement

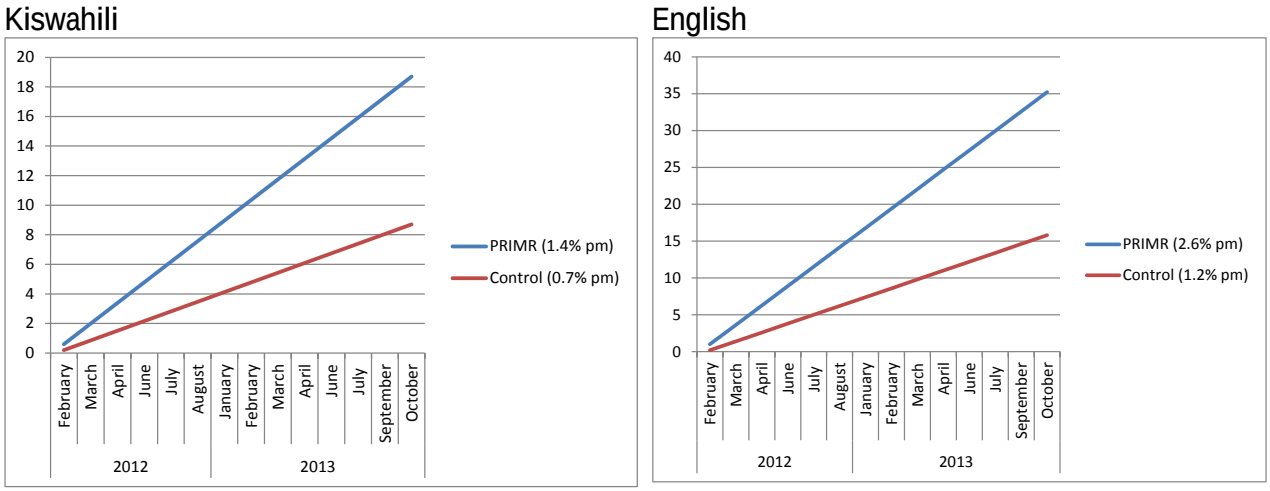
Figure 3. Rate of increase in English letter-sound fluency from baseline to endline, PRIMR and control schools, Class 2



This section presents graphically the increased rate of learning that pupils in PRIMR schools had compared with control schools. ***Figure 3*** presents the rate of learning English letter sounds from the baseline to the endline for Class 2 schools. The rate of increase for PRIMR schools was more than four times faster than for the control schools. This shows two important things. First, the increase in learning due to PRIMR was significant. Second, the learning rate in typical classrooms in Kenya was not sufficient for pupils to learn their English letters in order to decode English.

To understand how the PRIMR Initiative improved outcomes in a more systematic way, see ***Figure 4.*** It shows, for both English and Kiswahili, the inferred rate of monthly increase in the proportion of pupils able to read at the MoEST benchmark. In Kiswahili, the rate of increase over the life of PRIMR (the 2012 and 2013 academic years) was 1.4% per month for PRIMR and 0.7% for control. Given that the percentage of pupils reading at benchmark was very similar (in Class 1) for these Class 2 pupils, this means a very dramatic shift in the proportion of readers in Kiswahili. The results were similar in English, though slightly more stark, as the percentage gain per month was 2.6% for PRIMR and 1.2% for control. Under the control condition, the rate of increase was simply much too slow to ensure that these Kenyan learners could become significantly more literate.

Figure 4. Rate of increase in the proportion of pupils reading at KNEC benchmark: Kiswahili and English, Class 2



Similar to the figures above, ***Figure 5*** presents the increase in the proportion of Class 2 pupils able to comprehend in English at 80% or above for PRIMR and control schools in the PRIMR sample. For this measure, the rate of increase in learning was nearly three times higher in PRIMR schools as in control schools. These pupils were learning how to read and how to comprehend what they read in significantly more efficient ways under PRIMR.

Figure 5. Rate of increase in the proportion of pupils reading with 80% or higher comprehension: English, Class 2

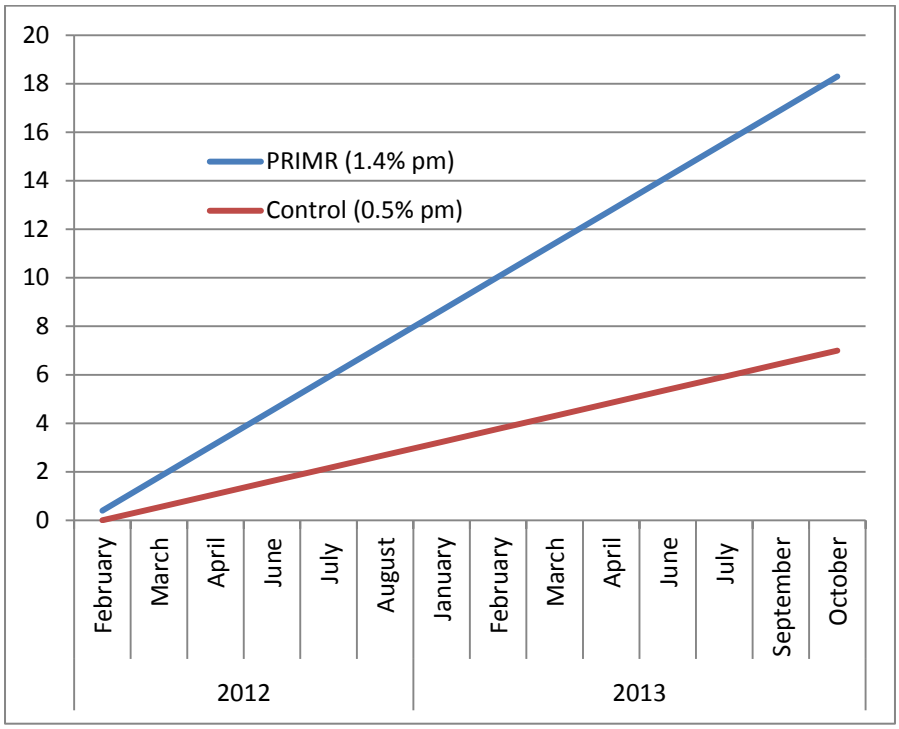
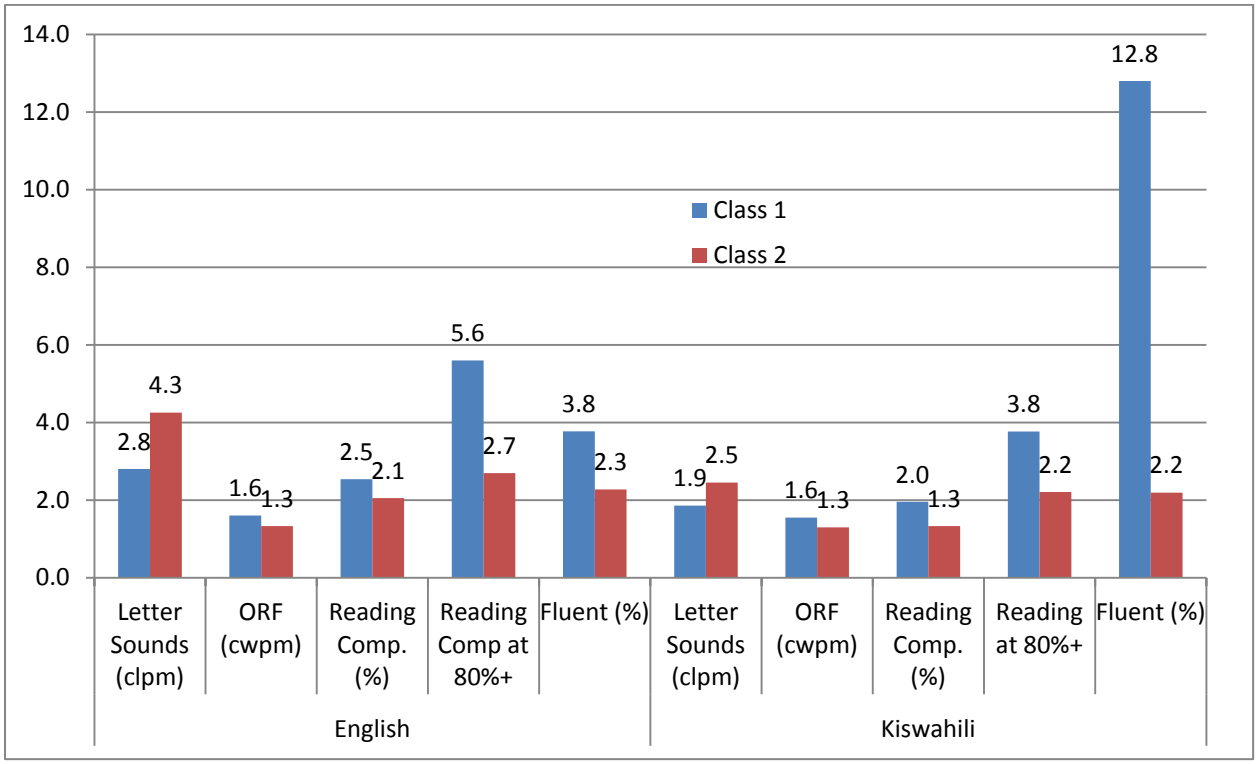


Figure 6 represents our attempt to understand how PRIMR returns related to the typical classroom in Kenya. This figure presents the PRIMR gains over baseline against the control gains over baseline as a proportion. For example, when Figure 6 shows that letter sounds in English is 2.8, it means that the gains in English letter-sound fluency were 2.8 times more in treatment schools than in control schools.[8] The lowest gain was 1.3 times more in treatment than in control schools. The highest gain was 12.8 times more in treatment than in control schools. Figure 6 shows the majority of subtasks as being around two times more in treatment than in control schools, with an average slightly above that. PRIMR was dramatically improving the rate of learning for these pupils, at between two and three times on average.

[8] These results were derived from the differences-in-differences regression models.

Figure 6. PRIMR gains as a proportion of control over baseline



4.4 Progress Toward KNEC Benchmarks

In this section we present the outcomes of PRIMR in comparison to the benchmarks set by KNEC and the MoEST in August 2012. The benchmarks for *fluent reader* were set at the levels of fluency needed to comprehend at 80% or higher, which were 45 cwpm for Kiswahili and 65 cwpm for English. The benchmarks for *emergent reader* were set at the levels of fluency needed to comprehend at 20% or higher, which were 17 cwpm for Kiswahili and 30 cwpm for English. These measures are the primary method of determining program success in both PRIMR and the upcoming USAID-funded Tusome program at the national level.

Sadly, as of the writing of this report, PRIMR had not been able to hold a mathematics benchmarking exercise with the MoEST to determine measures for quality that would provide meaning for pupils in Kenya. When this task has been completed, it will be much easier to determine whether and how PRIMR-type instruction impact mathematics outcomes in meaningful ways.

Figure 7 is a graphical representation of the impact of PRIMR on the percentage of pupils reading at the KNEC English benchmark. It shows that significant gains were made by pupils in Class 1 and 2 and in public and LCPS. The rates of increase between treatment and control schools were dramatically different. In short, PRIMR was helping pupils become literate much more easily than the control public or LCPS's approaches were able to. For this figure, Class 1 was being measured against a Class 2 benchmark, so gains were expected to be modest.

Figure 7. Proportion of PRIMR and control pupils reading at English benchmark

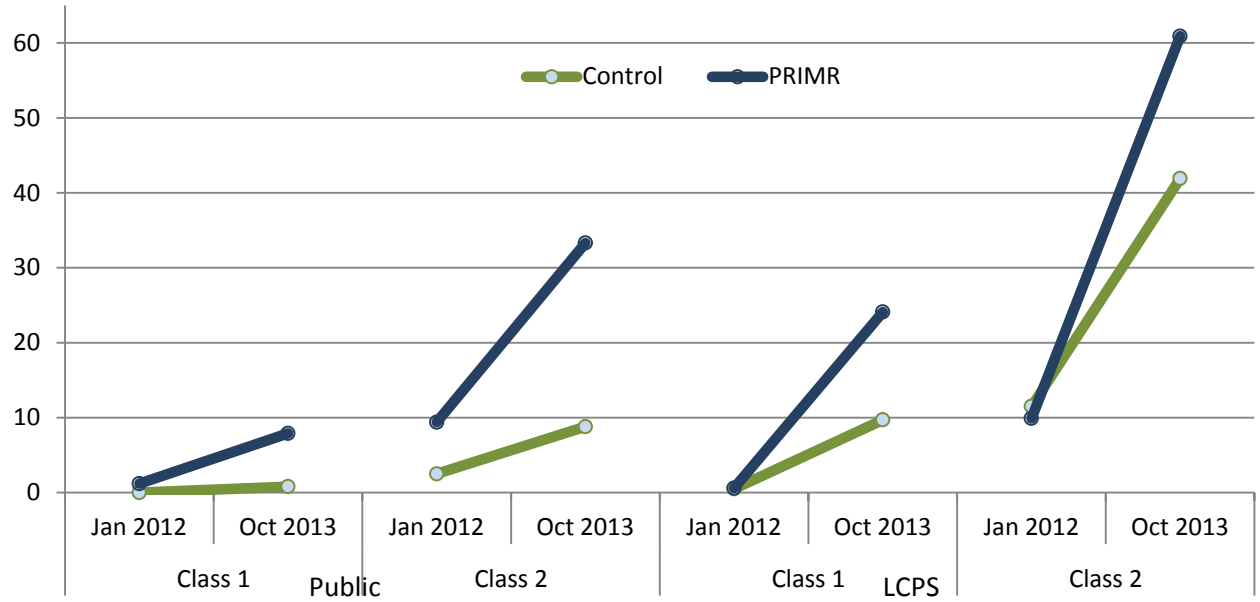
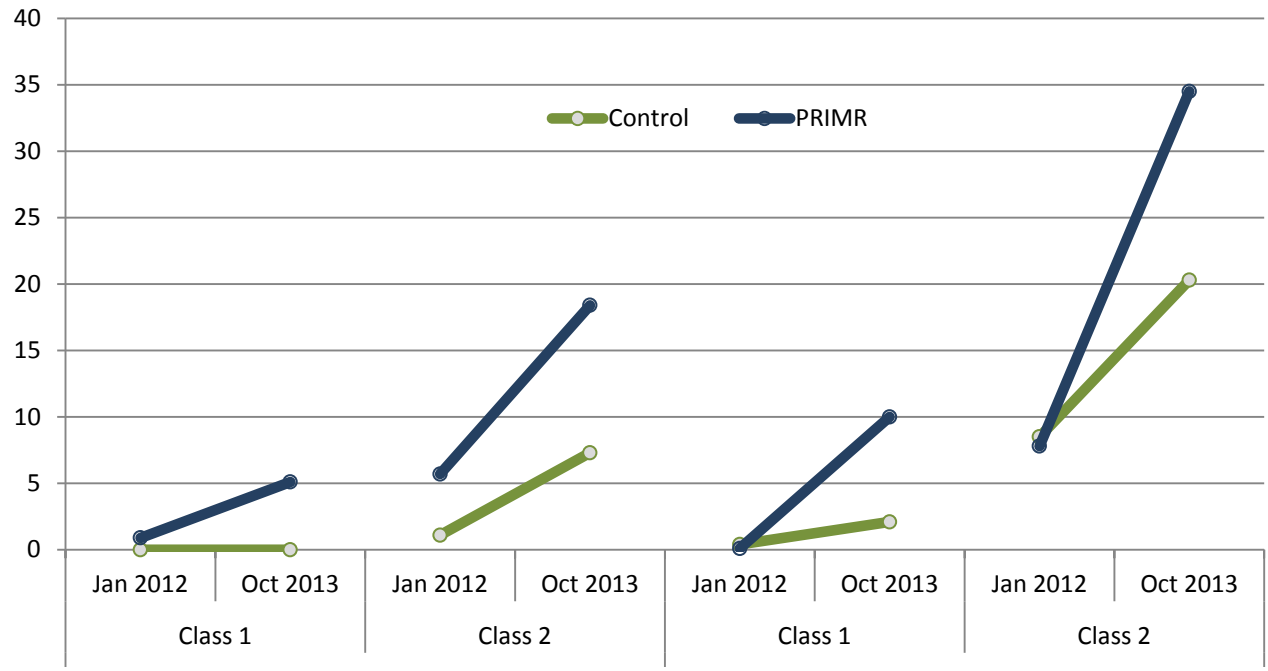


Figure 8 compares the ability of PRIMR and control schools to **increase** the proportion of pupils reading at the MoEST and KNEC benchmark in Kiswahili. The rates of increase were nearly the same between the two languages. This shows that PRIMR was able to fundamentally change outcomes for pupils in both languages, grades, and settings.

Figure 8. Proportion of PRIMR and control pupils reading at Kiswahili benchmark



Looking at the benchmark data another way, in ***Figure 9*** we present proportions of pupils at four particular levels of literacy: zero words per minute, below the emergent level, above emergent but below fluent, and truly fluent. The figure shows that the endline results were significantly different from the baseline results, both for treatment and control schools. Most notably, ***Figure 9*** shows that zero scores in treatment schools were nearly half of those in control schools (6.9% compared to 13.1%). On the other hand, treatment school fluent scores were more than double those of the control schools (43.7% compared to 21.3%). Figure 9 shows that the impact of PRIMR was being felt across the distribution of scores, and PRIMR was helping pupils who were struggling with the basics, as well as pupils who were on the verge of breakthroughs in decoding and comprehension.

Figure 9. Progress against English KNEC benchmarks, January 2012 to October 2013

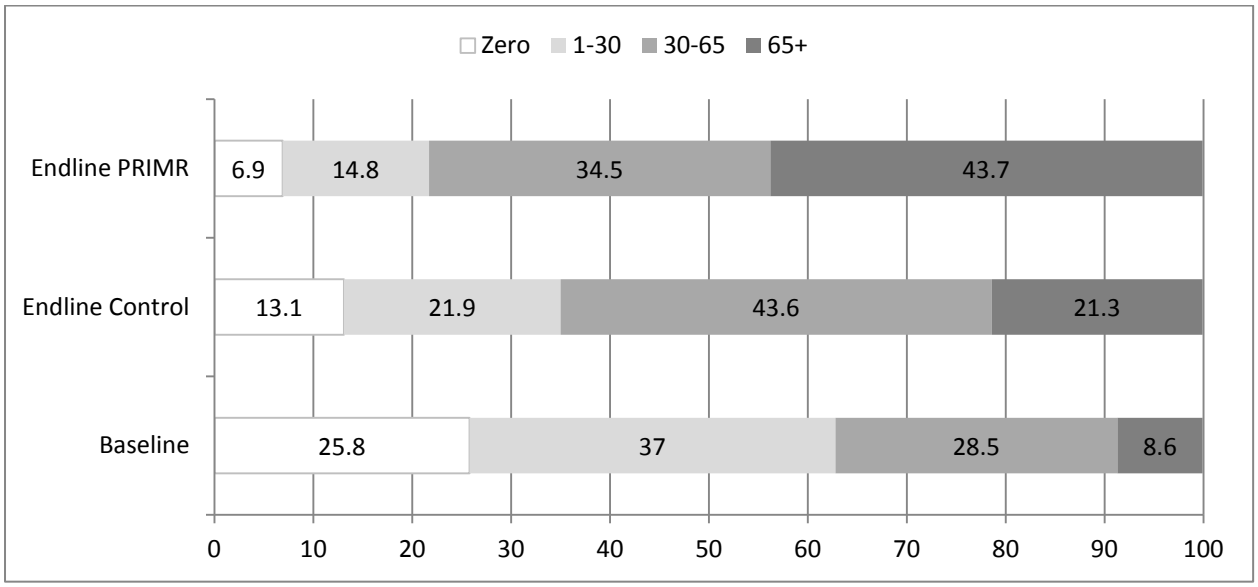
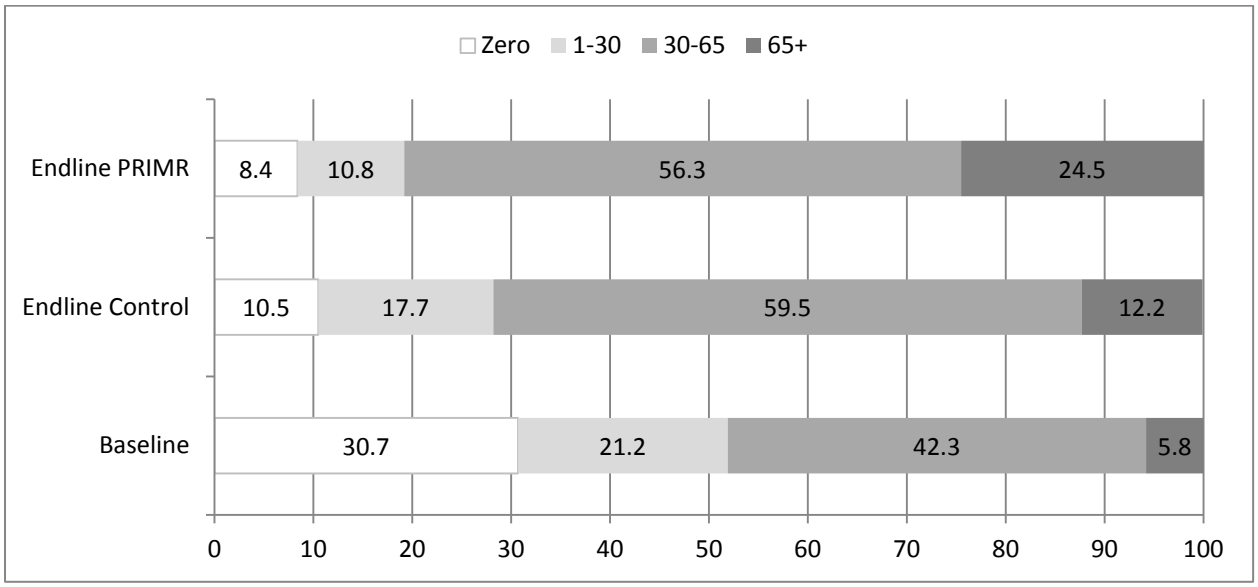


Figure 10 presents the impact of PRIMR on Kiswahili outcomes across the distribution. There was a modest impact of PRIMR on the percentage of pupils who read zero words a minute (8.4% compared to 10.5%). The treatment schools had many fewer pupils who read below benchmark (1–30 cwpm) than control schools, with 10.8% compared to 17.7%. The percentages of pupils who read at the emergent level were very similar for PRIMR (56.3%) and control (59.5%). Most notably, the PRIMR intervention doubled the proportion of pupils reading at the KNEC benchmark (24.5% compared to 12.2%). In short, PRIMR improved literacy across the distribution of Kiswahili outcomes.

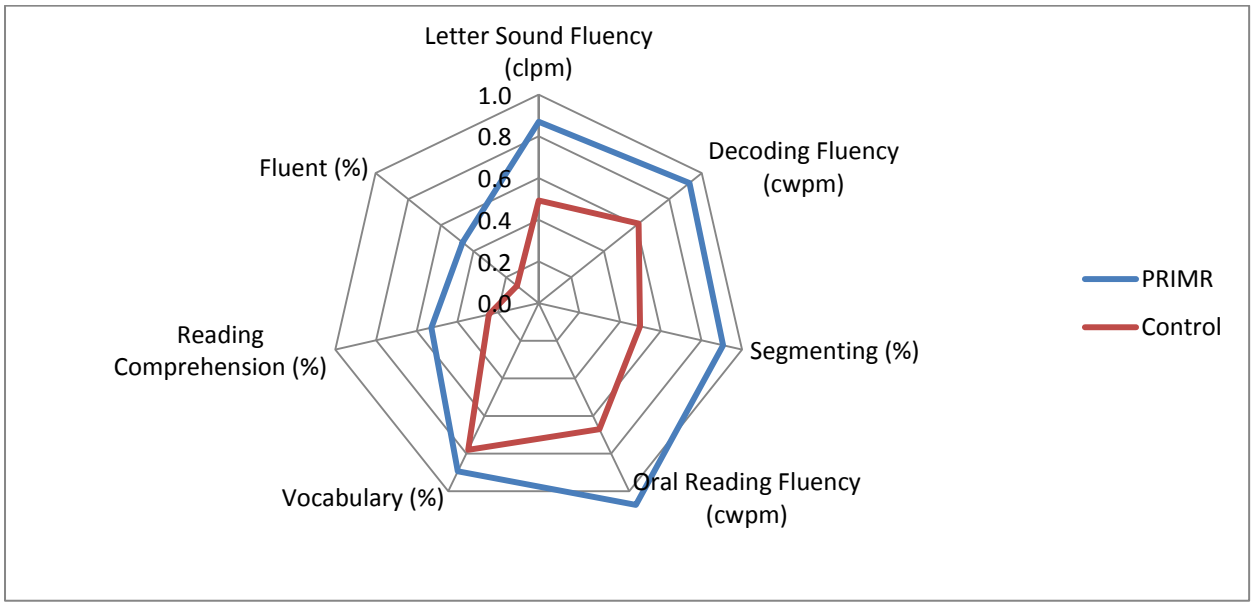
Figure 10. Progress against Kiswahili KNEC benchmarks, January 2012 to October 2013



It was also important, if possible, to investigate a dual comparison: (1) the progress of pupils receiving PRIMR treatment against the expected outcomes for pupils in each class against (2) how treatment pupils' progress in those areas compared with that of the control group. The radial plots in ***Figure 11*** through ***Figure 13*** present the treatment and control group mean scores for Class 1 for each subtask against the expected outcomes (benchmark) on that subtask (that is, the outer ring of each plot represents the benchmark value).[9] The expected outcomes were generated by the PRIMR monitoring and evaluation (M&E) team, and allowed for all subtasks to be put on one scale. For English, this shows that treatment pupils were outperforming their control colleagues by significant margins, except for the vocabulary subtask, where the results were much closer. The figure also shows that treatment pupils were performing at the expected level in oral reading fluency, and were quite close to the expected level in decoding fluency, segmenting and vocabulary. There remains significant work to be done in reading comprehension and the percentage of pupils reading at the KNEC benchmark.

[9] This document provides expert estimates as to the appropriate benchmark levels. This must be validated and updated based on the views and expectations of the MoEST and KNEC. This could occur during a benchmarking exercise taking into account the curriculum, achievement, and expected gains year by year.

Figure 11. English, Class 1: PRIMR and control groups, mean scores and benchmarks, for each subtask



For Kiswahili, the results are presented in ***Figure 12.*** PRIMR outcomes were better than those of the control schools on all measures, but the largest gaps were for oral reading fluency and reading comprehension. Relatively smaller gaps were found for listening comprehension and the proportion of pupils reading at the KNEC benchmark. This might suggest that the Kiswahili KNEC benchmark is set somewhat too high, but more analysis will be necessary to determine whether that is the case. It is worth noting that syllable fluency results were better in PRIMR than in control schools, even though syllable practice is something that is present in all Kenyan classrooms in abundance. The instructional strategies in PRIMR made pupils more able to improve their syllable skills, apparently.

Figure 12. Kiswahili, Class 1: PRIMR and control groups, mean scores and benchmarks, for each subtask

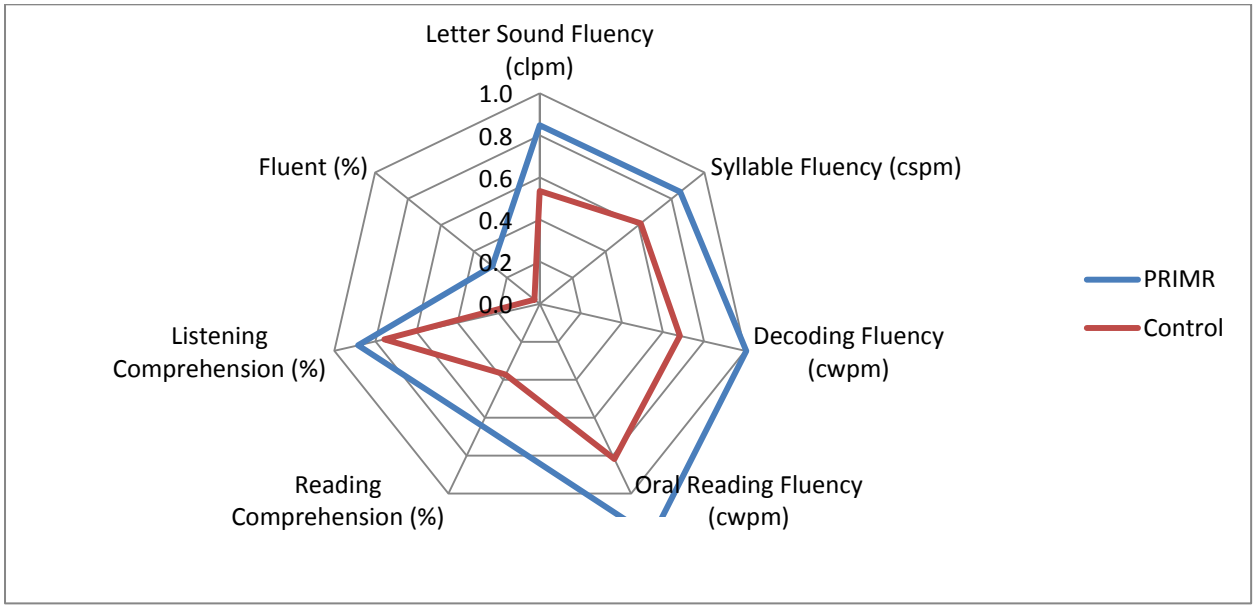
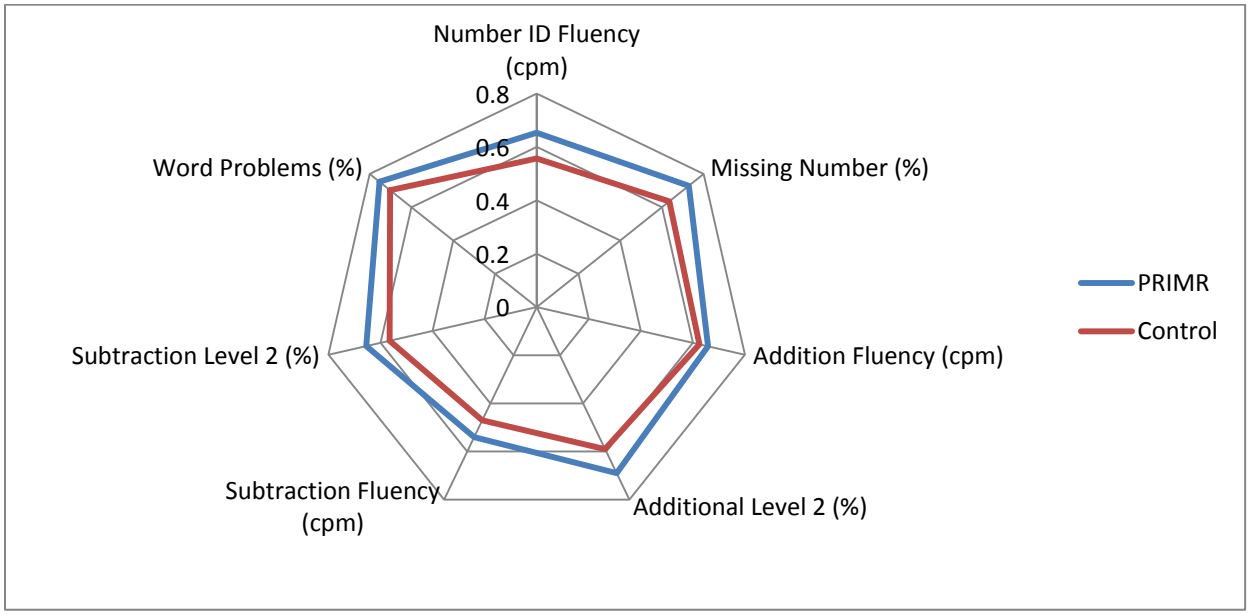


Figure 13 shows that treatment pupils in PRIMR outperformed their control counterparts in all areas of mathematics in Class 1. However, the magnitude of the difference was relatively small. For example, the gaps in addition fluency between PRIMR and control were statistically significant but substantively small. Earlier parts of this report showed that the PRIMR effect size in math was smaller in Class 1 than in Class 2. In both PRIMR and control schools, the results were such that the outcomes remained somewhat far from the benchmark at the end of Class 1.

Figure 13. Math, Class 1: PRIMR and control groups, mean scores and benchmarks, for each subtask



Turning now to Class 2 comparisons, the results showed a similar dramatic difference between PRIMR and control on the majority of subtasks.

For example, in English (see ***Figure 14***), average scores for Class 2 pupils who received the PRIMR treatment approached the expected outcomes for vocabulary, oral reading fluency, and the percentage of pupils reading at the KNEC benchmark. There was a much smaller difference in the vocabulary measure between PRIMR and control. This was partially because the vocabulary subtask was somewhat easy, and therefore resulted in less variation, but also because vocabulary received heavy emphasis in the control schools.

Figure 14. English, Class 2: PRIMR and control groups, mean scores and benchmarks, for each subtask

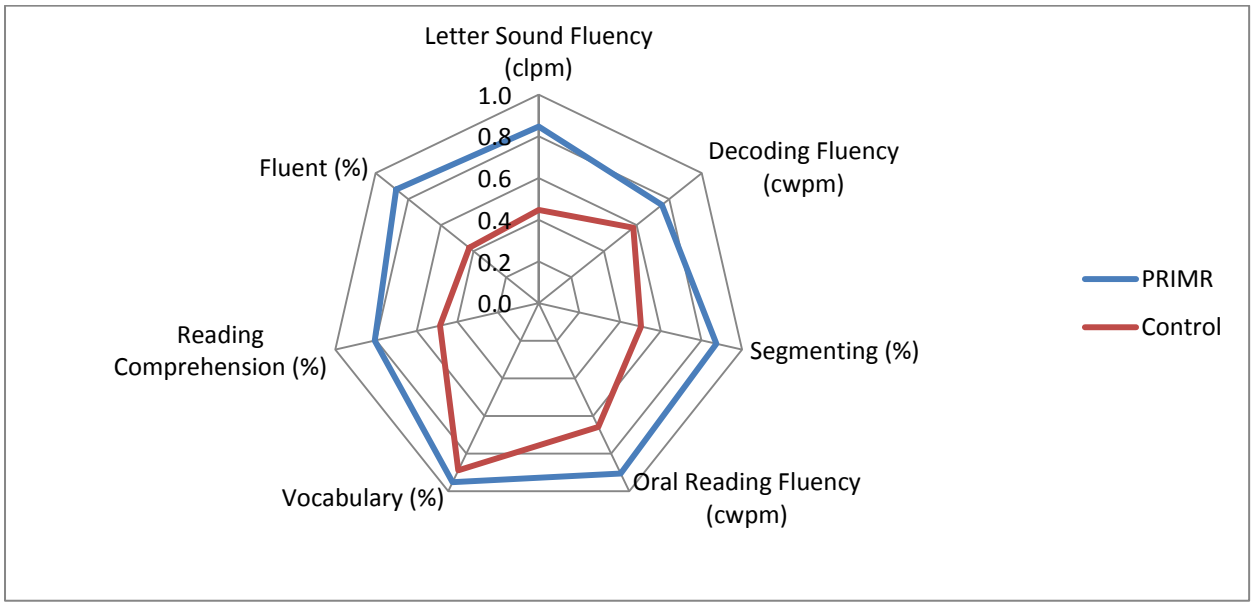
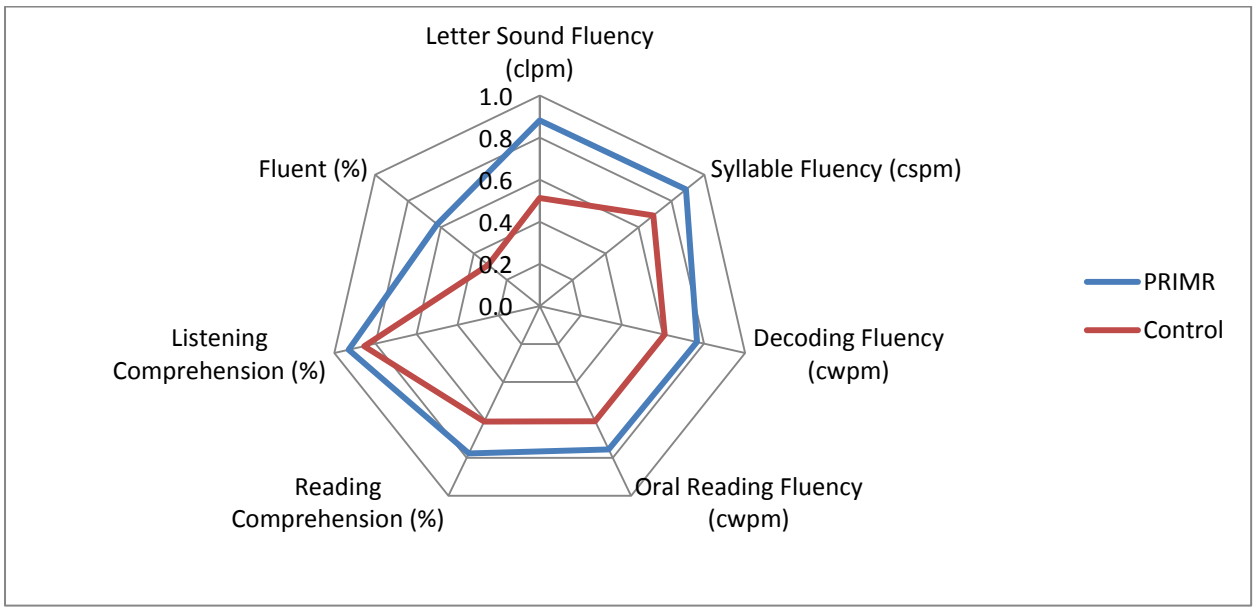


Figure 15 presents the progress against the expected outcomes for Kiswahili in Class 2. The stark difference in achievement between PRIMR and control is evident here. The radial plot shows that pupils in PRIMR did much better in letter-sound fluency and in the percentage of pupils reaching the benchmark. It also shows a small gap for listening comprehension, which means that the actual achievement in that area in the control schools was relatively high, which was not unexpected.

Figure 15. Kiswahili, Class 2: PRIMR and control groups, mean scores and benchmarks, for each subtask



Similar to Class 1, the impact of PRIMR on mathematics performance was significant but smaller than in literacy (see ***Figure 16***). The difference between treatment and control was larger in Class 2 than in Class 1, and was most notable in the gaps associated with missing

number and subtraction level 2. The achievement levels for both groups were relatively consistent in math, with treatment pupils' results approaching 80% of the achievement level needed for the benchmark on most tasks, except for subtraction fluency and subtraction level 2. This shows that the most difficult skill was also the one with the most significant difference between control and PRIMR.

Figure 16. Math, Class 2: PRIMR and control groups, mean scores and benchmarks, for each subtask

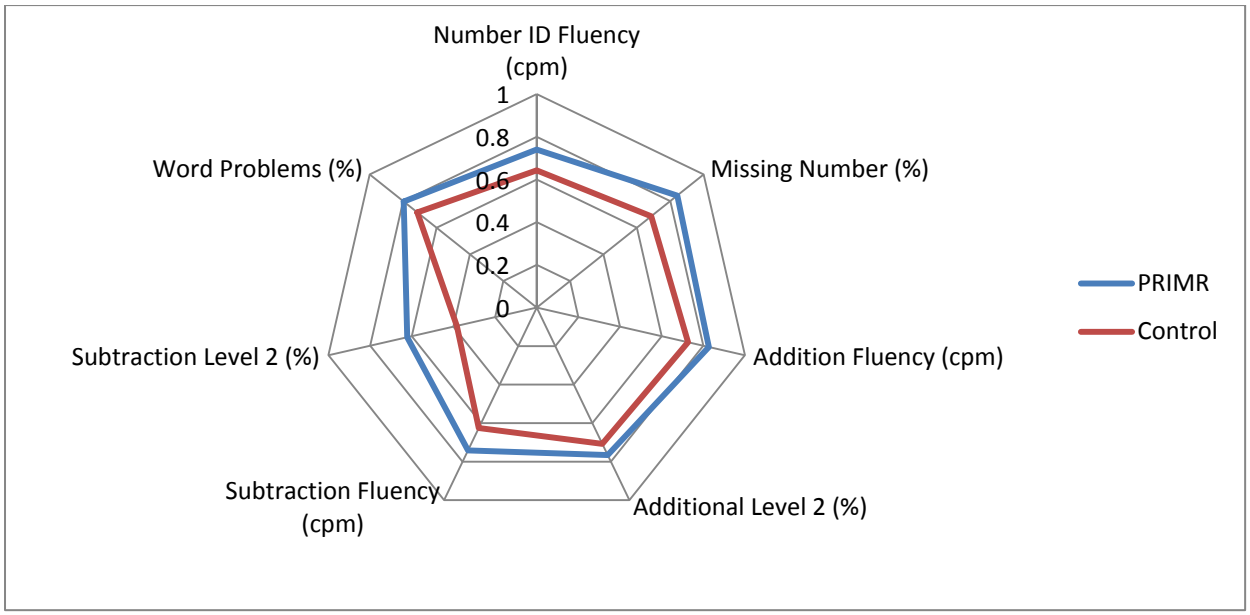
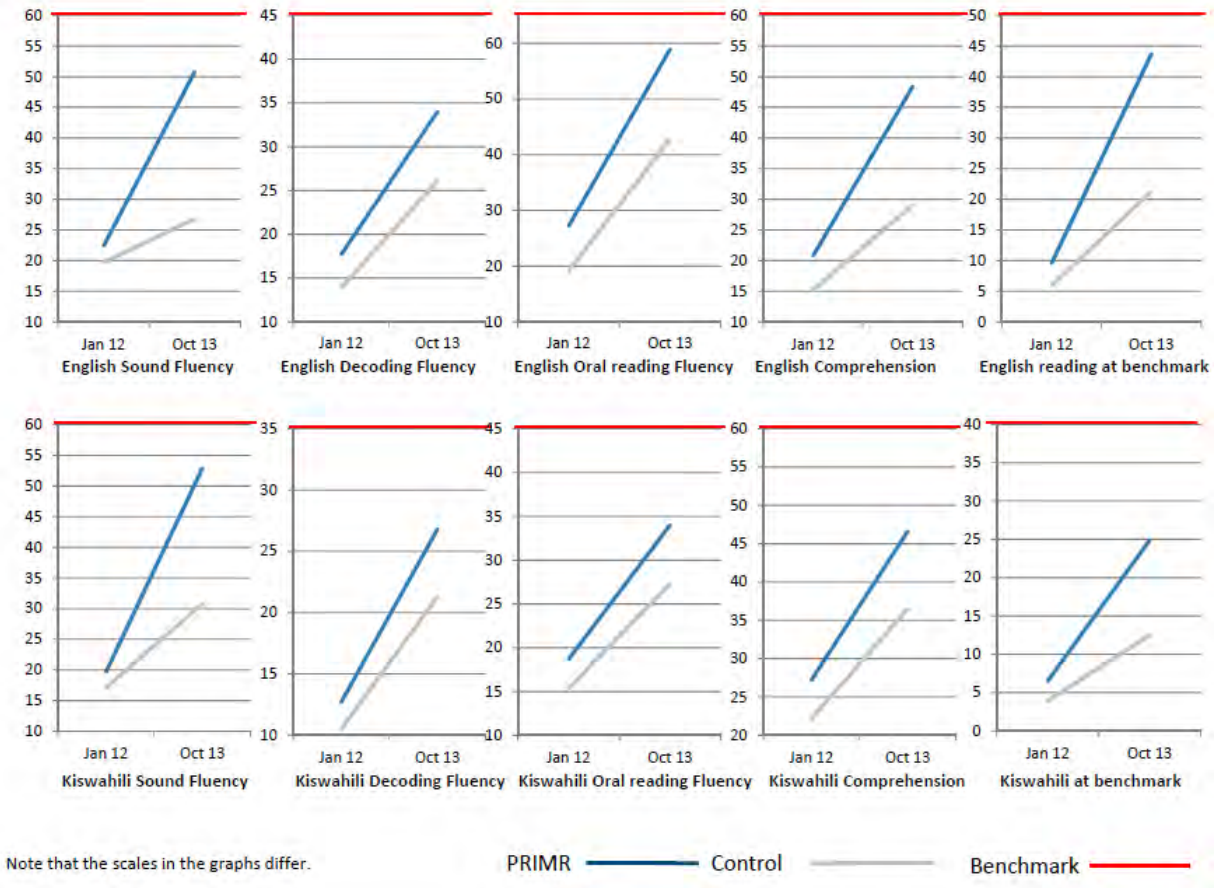


Figure 17 presents an alternate route to understanding how PRIMR impacted learning. It shows 10 measures side by side (5 for English and 5 for Kiswahili) and the outcomes on each of those measures for PRIMR and control schools, compared with the benchmarks for Class 2. The results are from the baseline (January 2012) and the endline (October 2013).

In most of the graphs, one can see a small (and sometimes statistically significant) difference between treatment and control outcomes at baseline. Therefore what is most important in determining the relative effect of PRIMR is investigating the trend in achievement between the baseline and endline. The slopes of the curves should be different if PRIMR is successful. This is easiest to see in the English and Kiswahili subtasks for letter-sound fluency, English and Kiswahili reading comprehension, and English and Kiswahili reading at benchmark. The slopes are different for these subtasks, and less so for decoding fluency and oral reading fluency. This suggests two things. First, pupils *are* learning in Kenya without PRIMR, albeit at a lower rate. PRIMR's responsibility was enormous, then, to increase learning over the current structure and to show consistent improvement in achievement.

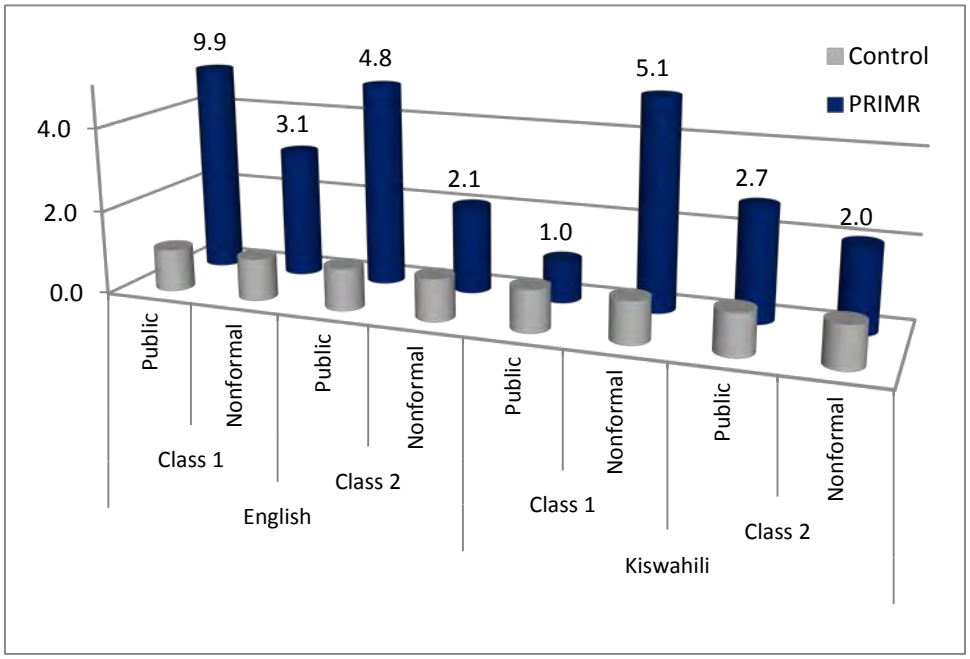
Figure 17. Side-by-side comparison of progress toward benchmarks, from baseline to endline, English and Kiswahili subtasks



The majority of the analyses presented above used OLS regression to determine the impact of PRIMR in progressing toward the KNEC benchmarks. To produce ***Figure 18,*** the PRIMR researchers ran logistic regression models with control variables, holding constant the effect of wealth and gender. Interestingly, the models clearly showed the odds that pupils in PRIMR treatment schools would achieve at the benchmark level in comparison to the pupils in control schools.

Each of the gray bars shows "1," for one pupil able to read at the KNEC benchmark. The blue bars show the increased likelihood of being able to read at the KNEC benchmark if a pupil was in a treatment school. Thus, the interpretation of 9.9 in public Class 1 for English is that pupils in treatment schools were 9.9 times as likely to be able to read at benchmark as pupils in control schools. Stated another way, all things constant, there would be 9.9 pupils reading at benchmark in treatment schools compared with 1 pupil in the control schools.

Figure 18. Logistic regression results on the likelihood of reading at KNEC benchmark in PRIMR and control schools (wealth and gender held constant)



4.5 Cost Analysis

The discussion above focused on answering the question of whether PRIMR was effective at improving literacy and numeracy outcomes in Kenya. The results showed that, yes, PRIMR was effective. The next portion of the analysis examines whether PRIMR was cost-effective. This analysis requires that we incorporate cost figures.

To illustrate how the cost of books influenced overall costs, ***Figure 19*** shows the average per book costs for PRIMR and current per capita costs. Analysis of book costs at bookstores revealed that, on average, pupil books for lower primary cost 375 shillings, or US$4.41.[10] The PRIMR learner books completed for distribution in January 2014—with their full-color pages, full-color covers, high-quality paper, and high-quality cover stock—cost 70.6 shillings apiece, or less than US$0.84. The books on the market, which typically had fewer pages and lower-quality content and layout, cost more than five times as much as the PRIMR books. This does not include the cost of distribution and logistics. The next pair of bars in Figure 19 shows that the total per capita allocation under the KESSP plans dedicated to learning materials, including books, was 400 shillings, or US$4.71. Purchasing three PRIMR books—i.e., one each for English, Kiswahili, and math—cost US$2.50. This comparison suggests that the per capita allocation would be more than sufficient to have 1:1 ratios in those three key subjects every year, except that the cost of books in the current Kenyan market is much too high. As a result, the Simba fund[11] can purchase only one market-priced book per pupil per

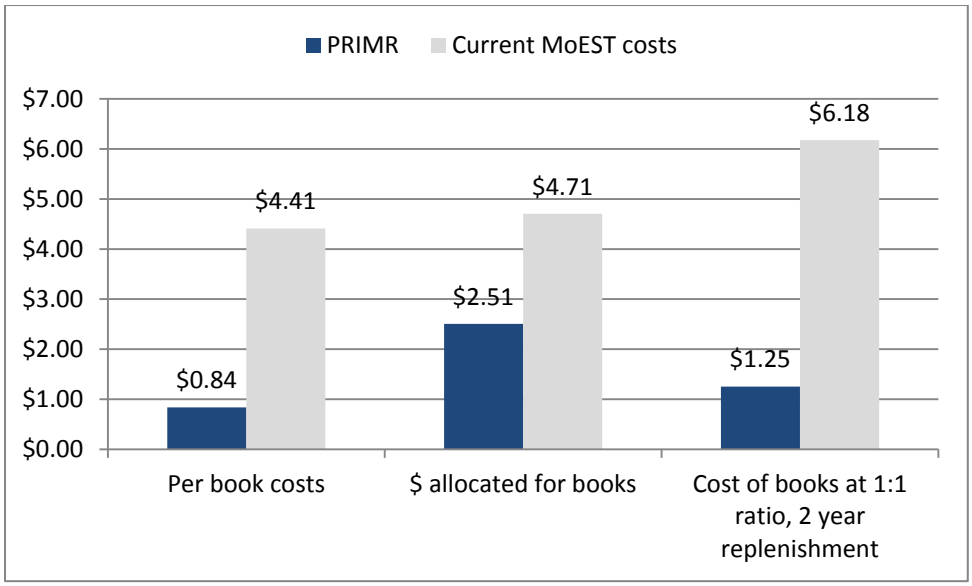
[10] A recent review of textbooks at Textbook Centre revealed a much higher price of 505 shillings per book. This was during the busy January season, so it is possible that textbooks were being marked up. We have continued to use the 375 shilling cost to make this analysis conservative.

[11] The Simba account allocates funds to schools at a rate of 350 shillings (US$4.71) per child per year to support school repairs, transportation, and worker pay. A second education account—the General Purpose account—

year, whereas a little over half that amount would purchase one PRIMR book per subject per pupil per year.

The final pair of bars shows the estimated cost of achieving a 1:1 ratio of books to pupils for all three subjects with a very liberal two-year replenishment cycle. For PRIMR, at current rates, the cost would be US$1.25 per pupil. For the current per capita expenditures, the estimated amount would be US$6.18 per pupil per year. This analysis suggests that the problem is not necessarily a lack of funds in the system: The 400 shillings currently available would be enough for a 1:1 textbook ratio for all pupils in Kenya without any additional funds required from Treasury, if the cost of the books were less expensive.

Figure 19. Average per book costs for PRIMR and current per capita spending (estimated) (US$)

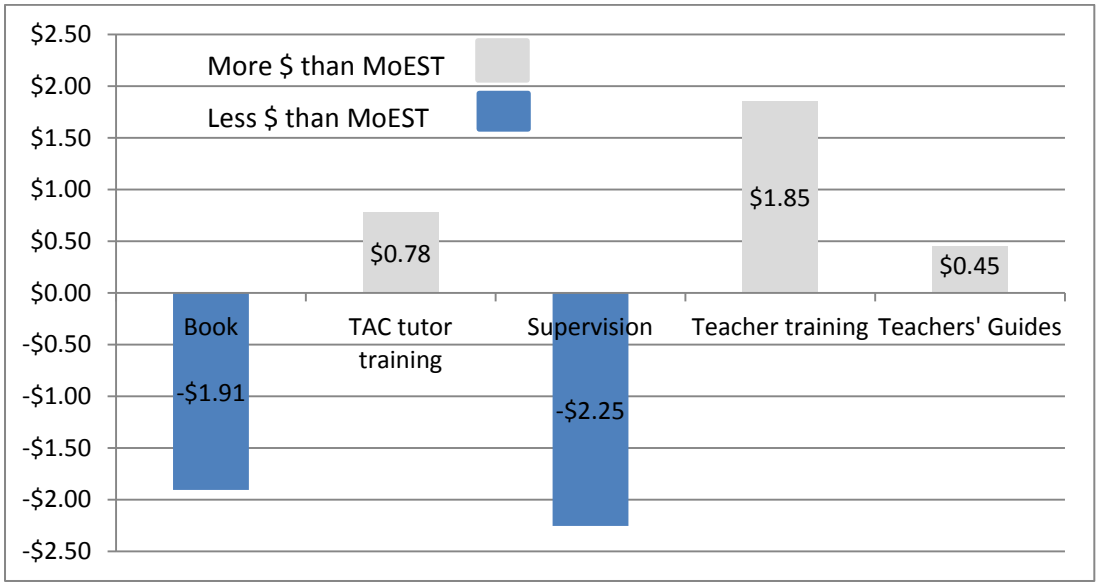


Given the complexity of the cost structures in the Kenyan education system, it was important to clearly analyze how PRIMR costs differed from what the MoEST was budgeting under KESSP. ***Figure 20*** presents the costs of PRIMR compared with the costs in the KESSP plans, in five categories.[12] The gray bars indicate the costs in areas where PRIMR spent more money than the per capita spending from the KESSP plans in 2013, while the blue bars indicate the costs in areas where PRIMR spent less money than the expenditure expected from KESSP. This analysis included the budgets for KESSP at the pupil level as well as the per pupil expenditures in PRIMR.

valued at 670 shillings per pupil annually, is supposed to pay for textbooks and instructional materials, but recently has been insufficient to cover the full cost of pupil books.

[12] This analysis depended on many assumptions. First, the PRIMR costs include only the allowances and per diem rates provided for training TAC tutors and teachers, the TAC tutor transportation costs, the cost of the books, and the cost of the teachers' guides (which subsumes other instructional materials). The PRIMR figures do not include the PRIMR staff salaries or associated costs. The assumption is that at scale, these costs will have already been spent to create the system, and the bulk of the scale-up costs will be focused on the training, support, and material provision. For the MoEST costs, the estimates are from allocated and budgeted costs, rather than actual expenditures. This is an important distinction because for both the books and the supervision line items, the expenditures are potentially lower than the budgeted costs. Further analyses could be done to determine actuals and this could inform a revised cost-effectiveness analysis for PRIMR.

Figure 20. PRIMR costs compared to current per capita costs (estimated)



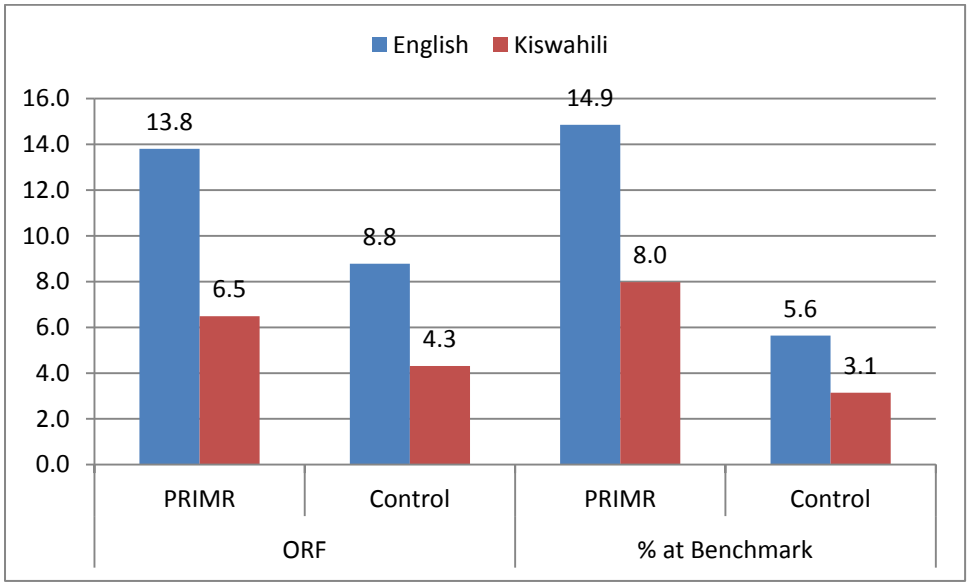
This comparison reveals that, as discussed above, the estimated current per capita spending was much larger per pupil on books, and at the same time, getting a much higher book-to-pupil ratio than in PRIMR. This was due almost entirely to the high costs of the books in the open market, but the challenges of distributing books through the booksellers were also problematic. Similarly, the budget lines allocated to instructional supervision at the national, county, and sub-county levels were much higher than PRIMR was spending to support TAC tutors to visit schools. This was partially because the KESSP allocations for observations were not always spent in the way they were budgeted, but also because PRIMR gave only local daily transport reimbursements rather than per diems and meal allowances.

On the other hand, Figure 20 shows three areas where PRIMR outspent the estimated per capita expenditures. First, the TAC tutors under PRIMR received 15 days of training in 2013, much more than the current system currently provides. However, given that each TAC tutor oversaw many pupils, the actual per pupil cost of the TAC tutor training was modest. The teacher training costs in PRIMR represented the biggest difference from the current system. PRIMR spent approximately US$1.85 per pupil on transport and meals for teachers trained over a 10-day period. These payments were essential to ensure attendance by teachers, and attendance rates were in fact very high in comparison to those obtained by the MoEST's Centre for Mathematics, Science and Technology Education in Africa (CEMASTEA) training program, which does not cover teacher transport costs. But the PRIMR rate is also somewhat too high to be sustainable, so the researchers suggest that under the new national literacy program, the implementer and the MoEST agree on a slightly lower transport reimbursement, and potentially use fewer days for training teachers, particularly if Tusome is focused only on literacy (and not numeracy). Finally, for its basic experimental intervention, PRIMR spent more resources on instructional materials, which included teachers' guides for every subject, a pocket chart, and flashcards for literacy and numeracy. These materials were essential to the success of PRIMR and we believe they are absolutely worth the cost.

4.6 Cost-Effectiveness Analysis

Determining whether costs are viable should be done in comparison to the impact associated with that cost. ***Figure 21*** takes the entire cost of the PRIMR program, as well as the entire estimated per pupil cost of the current system, and for the 2013 academic year, estimates the gains in outcomes achieved per U.S. dollar. In order to understand how this works, recall that we presented the gains from PRIMR over the baseline, in comparison to the control. To create the cost-effectiveness analysis in Figure 21, these same gains, specifically for oral reading fluency and the proportion of pupils reaching the KNEC benchmark, were divided by the per pupil costs. The graph shows that PRIMR was much more cost-effective than the current MoEST system. For example, PRIMR increased oral reading fluency for English by 13.8 cwpm per U.S. dollar, while the MoEST system increased oral reading fluency by 8.8 cwpm per U.S. dollar. The cost-effectiveness benefit of PRIMR is even more evident in the measure looking at the percentage of pupils reading at the KNEC benchmark. This shows that for both English and Kiswahili, PRIMR was more than two and nearly three times as cost-effective as the current system. Note that this is largely due to the differences in the cost of books.

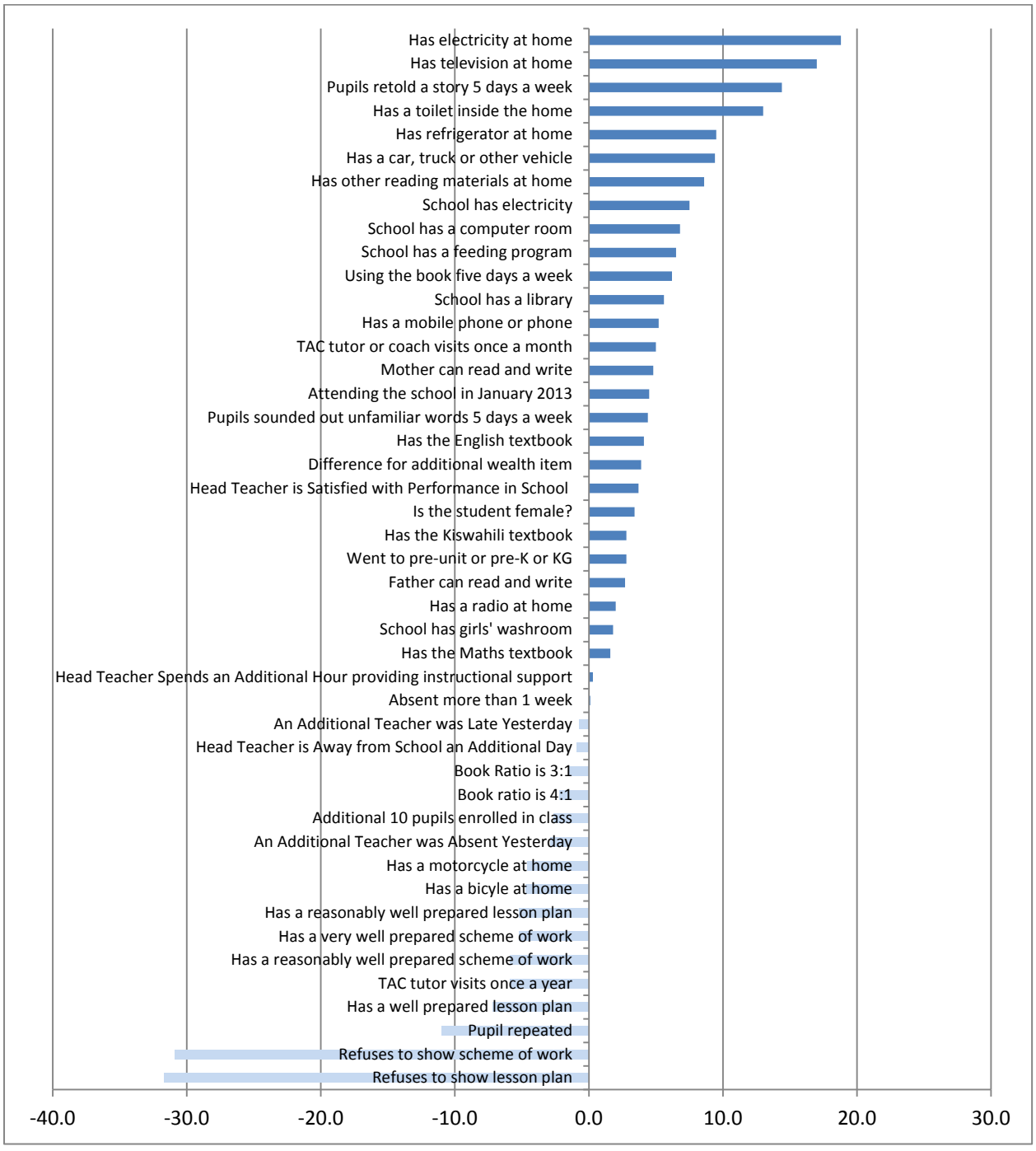
Figure 21. Cost of PRIMR vs. the current system, 2013: Comparisons based on gains in oral reading fluency and pupils reading at benchmark



4.7 Impact of Other Factors

The main focus of this report has been on whether PRIMR was successful, which groups it best succeeded with, and what the cost of PRIMR was relative to its impact. In this section, we investigate what other factors make it more or less likely for pupils to improve in literacy and numeracy in Kenya. Methodologically this means that we are presenting the coefficients on regression variables included in the DID estimates for the impact of PRIMR. In other words, these factors still mattered for achievement after PRIMR's impact was accounted for. ***Figure 22*** presents the relationship with English oral reading fluency associated with key factors of interest in Kenya. (These data were collected via the SSME interviews and checklists as well as brief pupil questionnaires.)

Figure 22. Effect of social and demographic factors on oral reading fluency scores



Our findings showed that pupils who had electricity at home read 18.8 cwpm more fluently than those who did not, and those with a television at home read 17.0 cwpm more fluently than those who did not. Given that these variables were a proxy for urbanicity, there seems to be no obvious policy implication of this finding. On the other hand, the next bar shows that pupils who retold stories 5 days a week read 14.4 words more fluently than those who did not. This is an instructional activity that Kenyan teachers could be encouraged to do consistently, as they do in PRIMR. Similarly, the findings showed that pupils who had other reading materials at home read 8.6 cwpm faster than those who did not. Pupils who used a

book in class 5 days a week read 6.2 cwpm faster, suggesting that consistency and book usage are essential. If a TAC tutor or coach visited at least once a month, these pupils read 5.0 cwpm more than those whose TAC tutor visited less frequently. This is a very large finding for this outcome and suggests that TAC tutor visits should be stressed. Sounding out unfamiliar words was worth 4.4 cwpm, and having the English textbook was worth 4.1 cwpm. Interestingly, having attended pre-unit or pre-kindergarten or kindergarten was worth only 2.8 cwpm. This suggests that more work should be done to improve the quality of instructional content provided in early childhood development programs.

Several variables were related to poor performance. For teachers reporting late on the day of the assessment, pupils read 0.7 cwpm less, and for every day the head teacher was away per month, pupils read 0.9 cwpm less. Focus on teaching and time in class is therefore essential. Pupils with worse book ratios also did worse on the assessment, as those with 3:1 ratios read 1.5 cwpm lower, and those with 4:1 ratios read 2.2 cwpm lower. Class size had a modest negative effect, with every additional 10 pupils associated with 2.6 cwpm less. Teacher absenteeism was a significant problem, as pupils read 2.9 cwpm less for every additional teacher absent. Interestingly, having a reasonably well prepared or a well-prepared lesson plan or scheme of work showed a negative relationship with English oral reading fluency of between 5.2 cwpm to 7.2 cwpm.[13] This shows, controversially, that spending time preparing lesson plans, rather than using pre-prepared plans such as the ones that PRIMR offers, was associated with lower achievement.

The variable that had the largest negative relationship with English oral reading fluency was pupil repetition. Pupils who said they participated in this type of instruction identified 11.0 fewer cwpm than those who did not, a very large impact. Repeating seems to be an ineffective solution to improve outcomes. The pupils in classrooms where teachers refused to show the scheme of work or the teachers' guide suffered the most, reading 31.7 and 30.9 cwpm less than those who were willing to do so.[14] This suggests, quite powerfully, that the current system of schemes of work and lesson planning is not working very efficiently and alternative solutions should be suggested.

4.8 Classroom Visit Frequency

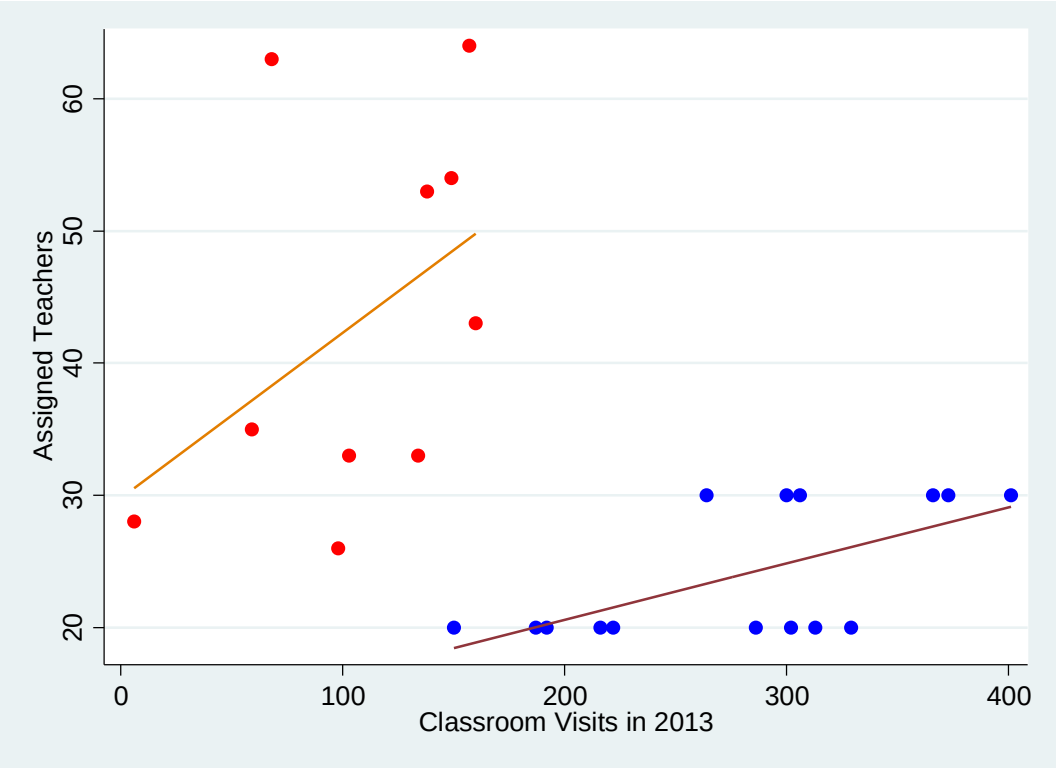
We examined whether the number of schools assigned to the TAC tutor or instructional coach had an impact on the frequency of visits by the TAC tutor and coach. To obtain this estimate, we created scatterplots comparing the number of classroom visits each TAC tutor and coach made with the number of teachers assigned to the tutor or coach within the zone or cluster. ***Figure 23*** presents the findings. The most obvious point is that instructional coaches were much more likely to visit classrooms than TAC tutors. The reasons for this are myriad; for example, the instructional coaches had far lower transport costs for individual lesson observations; and the PRIMR program was one of many activities that tutors were involved

[13] In 2014, PRIMR moved away from lesson plans to teachers' guides. At the time when this data was collected, October 2013, the teachers' guides were still in use. Therefore, this figure and the associated text discusses the PRIMR lesson plans.

[14] This research question was of interest to PRIMR given the difference of opinion within leadership in the MoEST about how frequently lesson plans and schemes of work were completed. In addition, some leaders in the education system believed that considering other ways to ensure that teachers prepare for lessons could be examined.

in, while it was the only work that the coach was engaged in. In any case, while the PRIMR program dramatically increased the number of classroom visits that TAC tutors undertook, it was still 2.6 times fewer than the coaches focused entirely on classroom observation in PRIMR.

Figure 23. Assigned teachers and classroom visits for TAC tutors serving formal schools (in red) and instructional coaches serving LCPS schools (in blue), with fitted trend lines



Annex C contains additional detailed discussion around the ratio of tutors/coaches to schools and estimated program effects, as derived from a series of regression analyses.

Taking all of these endline results analyses into account, the next two sections summarize key lessons learned and recommendations for moving forward.

5. Lessons Learned

This section presents key lessons learned from PRIMR in a variety of key areas focused on quality improvement in Kenya's primary schools.

1. **Training for TAC tutors:** As the results show, TAC tutors' visits to schools are critical for supporting teachers and improving pupils outcomes. Proper training of TAC tutors is essential so that they can effectively support teachers. The results also indicated that schools that are visited frequently are likely to have stronger pupil

performance; hence, TAC tutors should focus on making frequent and consistent classroom observations, even in the face of their heavy workload.

2. **Travel reimbursement structures:** PRIMR successfully facilitated TAC tutors to visit classrooms. This utilized a modest reimbursement that incentivized TAC tutors to visit classrooms consistently.
3. **Teacher training:** Training of teachers is a complex task that must assume that teachers are adult learners who learn best by doing and interacting with other professionals. This implies that teacher training should be organized around modeling and practice, and that having brief trainings with follow-up and refresher meetings is more effective than longer trainings.
4. **Distribution of classroom materials:** Distribution of materials to schools is a complex task. It requires accurate school enrollment data, prior planning, and a sophisticated distribution network. Ensuring that materials reach the schools on time was an essential PRIMR task.
5. **Priorities in the school calendar:** During the implementation of PRIMR it became apparent that certain times of the academic year required that the TAC tutor spend significant time away from the classroom. This occurred primarily during the extracurricular activity periods. These are clearly important for a balanced learning experience for pupils, but better understanding how these extracurricular activities could be organized so that they do not impede the TAC tutors' ability to support instruction is important.
6. **In-service training:** During PRIMR assessments and implementation, the evidence suggested that most of the teachers supported by PRIMR had not attended professional development courses or in-service courses for several years since leaving college or becoming teachers. The PRIMR Initiative's regular professional development through training and other activities filled a demand for increased instructional practice and support.
7. **Changes in instructional approaches:** Old habits take time to change, and the shift from traditional teaching to more active, sequenced, pupil-focused approaches was the central focus of PRIMR. Some teachers continued to use the two approaches concurrently at the beginning of PRIMR, in part because of concern about whether the lessons properly covered the material that would appear in the national end-of-year examinations. Advocacy was needed to change the mindset of some teachers.

Recommendations

Some recommendations from the endline assessment have already informed the program's final year of implementation. Others should be considered prior to a scale-up of PRIMR.

1. **Results and scale-up:** PRIMR's results to date have shown remarkable improvements in pupils' literacy and numeracy abilities, especially for pupils starting at the lowest levels of literacy and numeracy. The MoEST should therefore consider scaling up PRIMR activities to improve the quality of instruction in Class 1 and 2.
2. **Girls' performance:** The results indicated that, overall, girls were performing at the same level as—if not better than—boys, especially in literacy. Teachers should be trained in strategies for motivating girls so that they remain competitive as they move to upper primary.

3. **Zonal size:** The results showed that TAC tutors in large zones were less likely to have a significant an impact on pupil outcomes than those in smaller zones. Considerations should be made to limit the number of schools that the TAC tutors are responsible for. This would make TAC tutors more effective in supporting teachers frequently.
4. **Textbook ratio:** Provision of books to pupils at a 1:1 ratio is paramount in improving pupils' literacy and numeracy. The PRIMR analysis suggested that the government's current allocation would be enough to have a 1:1 ratio of books for all pupils in Kenya at low cost, if the cost of the books was more competitive.
5. **Advocacy and uptake:** There should advocacy of PRIMR's success through sharing of research results with a wider circle of stakeholders, including the MoEST and SAGAs.
6. **Language of instruction:** The language of instruction remains a complex issue for the Kenyan education system. Any attempt to scale up PRIMR activities without resolving this issue is likely to increase complexity during the implementation. The DFID PRIMR study, which is funding instructional materials and support in two mother tongues, will provide evidence as to the effectiveness of mother tongue compared with a basic instructional support program.
7. **Textbook policies:** The findings on cost and impact suggest that there is a need to consider the guidelines regarding vetting and selection of textbooks for use in schools. The complexity of multilingual literacy and numeracy instruction requires vetting guidelines that are tailored to the instructional characteristics of Kenya's system.
8. **Daily literacy and numeracy instruction:** Lesson time could be revised to accommodate more literacy and numeracy instructional time during the week. This is true not only because Kenya's literacy and numeracy allocations are paltry compared to the rest of East Africa, but also because of the evidence that in control schools, pupils spent very little time actually reading texts.
9. **Teacher assignments:** The transfer of teachers trained in PRIMR should be minimized to avoid the need for repeated onboarding and introductory training on a rolling basis. The TSC worked tirelessly to ensure that transfers were kept to a minimum, and we hope that can continue in the future.

Implementing these recommendations would increase the likelihood of PRIMR and any successor program having high levels of uptake by teachers and head teachers, as well as enthusiasm for the program from the County Education offices and TSC offices.

Most critically, the objective ensuring that all pupils are literate and numerate by Class 2 would be realized.

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Hill, C.J., Bloom, H.S., Black, A.R. & Lipsy, M.W. (2007). *Empirical Benchmarks for Interpreting Effect Sizes in Research.* MDRC Working Papers on Research Methodology.

Glass, G.V., McGaw, B. and Smith, M.L. (1981) *Meta-Analysis in Social Research*. London: Sage.

Ministry of Education, Science and Technology (MoEST). (2008). *The development of education: National report of Kenya.* A report presented at the international conference on education, Geneva.

Piper, B. (2010). *Kenya Early Grade Reading Assessment findings report.* Prepared for the William and Flora Hewlett Foundation, under the Monitoring of Learning Outcomes in Sub-Saharan Africa project, Contract No. 2008-3367. Research Triangle Park, North Carolina, USA: RTI International. https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=275

Piper, B., Jepkemei, E., & Kibukho, K. (In press). Pro-Poor PRIMR: Improving early literacy skills for children from low-income families in Kenya. *Africa Education Review.*

Piper, B., & Kwayumba, D. (2014). *The Primary Math and Reading (PRIMR) Initiative: USAID ICT Kisumu endline report*. Prepared for USAID under the Education Data for Decision Making (EdData II) project, Task Order No. AID-623-M-11-00001. Research Triangle Park, NC, USA: RTI International.

Piper, B., & Mugenda, A. (2012). *The Primary Math and Reading Initiative: Baseline report.* Prepared for USAID/Kenya under the Education Data for Decision Making (EdData II) project, Task Order No. AID-623-M-11-00001. Research Triangle Park, North Carolina, USA: RTI International. https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=480

Piper, B., & Mugenda, A. (2013). *The Primary Math and Reading (PRIMR) Initiative: Midterm impact evaluation.* Prepared for USAID under the Education Data for Decision Making (EdData II) project, Task Order No. AID-623-M-11-00001. Research Triangle Park, NC, USA: RTI International. https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=486

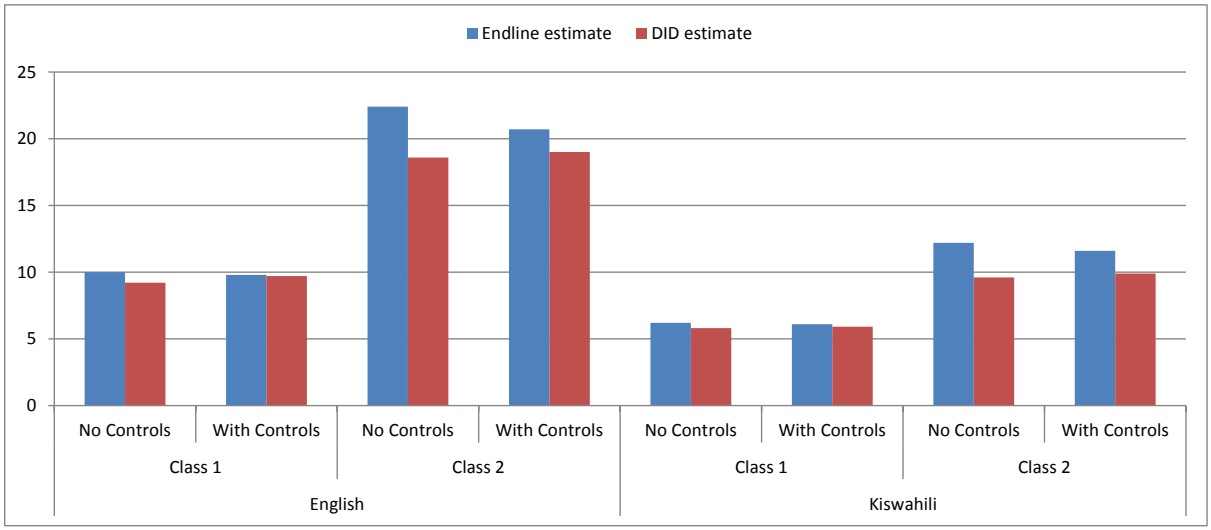
Uwezo. (2012). *Are our children learning? Annual learning assessment report: Kenya, 2012.* Nairobi: Uwezo.

Annex A. Methods: Endline Regression or Differences-In-Differences (DID) Estimates

The question of whether the endline data analyses should use a simple comparison of mean scores at the endline, depending on the randomized selection and assignment, or differences-in-differences analysis, is the central question in this annex. The PRIMR midterm analysis report (Piper & Mugenda, 2013) included both strategies. DID estimates allow for the removal of any differences at baseline between treatment groups. These estimates are straightforward to fit with two data points, but with the third data point provided by the October 2013 endline, the analyses and interpretations became significantly more complicated, particularly for any comparisons between groups. Given the similar but slightly different results using the simple regression analysis and DID estimates at the baseline, we fit both models to the endline results, to determine which ones to emphasize in the PRIMR endline report.

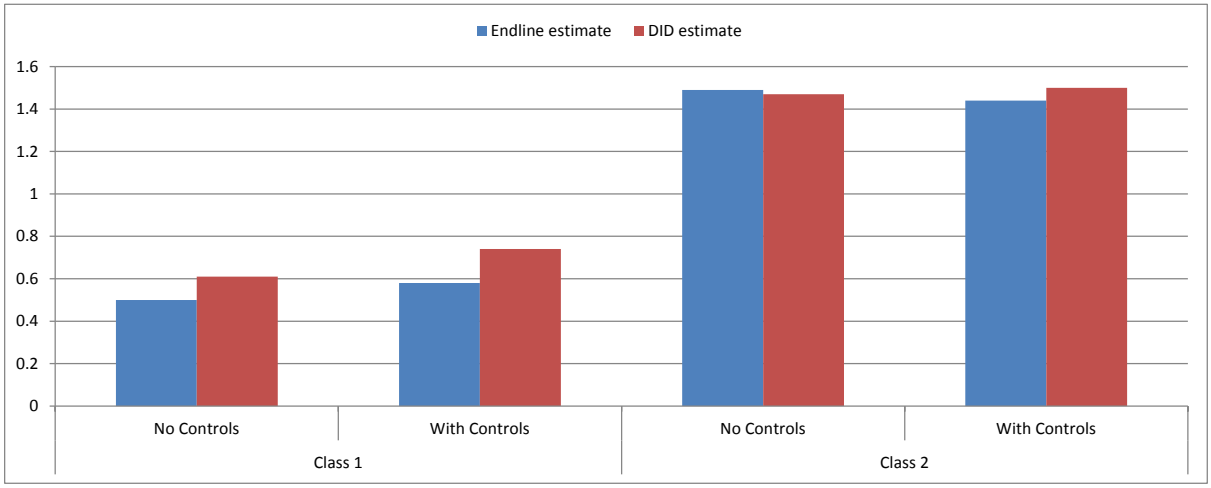
Figure A1 presents the impact of PRIMR for several models—some that include controls and others that do not, and models with DID estimators and others with basic regression. In Figure A1, the outcome variable was the percentage of pupils who reached the KNEC benchmarks for fluency and comprehension. The estimates were consistently somewhat higher for the endline regression estimate than for the DID estimate in Class 2. The difference was much smaller when controls were included. When controls were included, the estimate of PRIMR impact was somewhat smaller. Analyses using the outcome variable of the percentage of pupils reading at 80% or higher showed very similar results and are not presented here.

Figure A1. Comparison between differences-in-differences estimate of PRIMR effect and endline comparison estimate of PRIMR effect on the proportion of pupils reading fluently



Similarly, ***Figure A2*** compares the results from the basic regression models and the DID estimate. Unlike for literacy, where any small differences were in the favor of the endline estimate, these comparisons showed slightly higher outcomes for the DID estimate. The effect of the control variables was small, but in some cases significant.

Figure A2. Comparison between differences-in-differences estimate and endline estimate of PRIMR effect on addition fluency



Given the similarities between the two estimates and the simplicity of the endline estimate over the DID estimate, and the importance of the control variables, in this report we chose primarily to present the basic endline comparisons between treatment and control with the control variables of gender and pupil wealth included.

Annex B. Disaggregated Analyses of the PRIMR Effect on English, Kiswahili, and Math

Table B1. Program effect and effect sizes for EGRA (English and Kiswahili) and EGMA for Class 1, disaggregated by type of school

Class 1 PRIMR Effect											
		Formal					LCPS				
Subject	Subtask	PRIMR	Control	Effect	Standard deviation	Effect size	PRIMR	Control	Effect	Standard deviation	Effect size
English	Letter-sound fluency (clpm)	36.9	20.9	16.0	20.7	**0.773**	54.6	31.1	23.5	36.9	**0.637**
	Decoding fluency (cwpm)	18.3	12.4	5.9	12.8	**0.461**	31.3	20.5	10.8	20.6	**0.524**
	Segmenting (%)	49.5	29.3	20.2	32.3	**0.625**	62.8	30.8	32.0	45.7	**0.700**
	Oral reading fluency (cwpm)	23.3	14.7	8.6	19.8	**0.434**	47.1	29.8	17.3	35.8	**0.483**
	Vocabulary (%)	57.9	49.5	8.4	16.6	**0.506**	70.5	63.8	6.7	21.1	**0.318**
	Reading comprehension (%)	14.0	5.9	8.1	20.6	**0.393**	32.9	16.7	16.2	42.6	**0.380**
	Fluent (%)	7.9	0.8	7.1	20.1	**0.353**	24.1	9.7	14.4	48.9	**0.294**
Kiswahili	Letter-sound fluency (clpm)	36.2	25.0	11.2	20.9	**0.536**	52.3	29.9	22.4	35.5	**0.631**
	Syllable fluency (cspm)	32.6	25.4	7.2	21.6	**0.333**	47.7	31.6	16.1	28.5	**0.565**
	Decoding fluency (cwpm)	13.6	10.0	3.6	11.7	**0.308**	22.8	14.4	8.4	17.4	**0.483**
	Oral reading fluency (cwpm)	17.4	11.6	5.8	14.2	**0.408**	26.6	18.0	8.6	18.8	**0.457**
	Reading comprehension (%)	20.5	12.8	7.7	18.0	**0.428**	33.6	18.5	15.1	30.7	**0.492**
	Listening comprehension (%)	55.1	44.9	10.2	22.9	**0.445**	61.4	56.8	4.6	31.8	**0.145**
	Maze (%)	13.0	10.0	3.0	8.7	**0.345**	19.9	15.9	4.0	14.4	**0.278**
	Fluent (%)	5.4	0.0	5.4	16.4	**0.329**	10.1	2.2	7.9	33.3	**0.237**
Math	Number identification fluency (cpm)	17.6	15.7	1.9	8.1	**0.235**	23.2	18.8	4.4	10.3	**0.427**
	Quantity discrimination (%)	45.7	42.7	3.0	19.8	**0.152**	53.2	48.3	4.9	26.8	**0.183**
	Missing number (%)	29.7	26.9	2.8	15.1	**0.185**	38.2	32.1	6.1	21.3	**0.286**
	Addition fluency (cpm)	7.2	7.4	-0.2	3.5	**-0.057**	9.3	7.7	1.6	4.9	**0.327**
	Addition level 2 (%)	18.0	15.7	2.3	19.5	**0.118**	25.4	21.6	3.8	33.3	**0.114**
	Subtraction fluency (cpm)	4.7	4.5	0.2	3.1	**0.065**	6.7	5.1	1.6	4.8	**0.333**
	Subtraction level 2 (%)	11.1	8.6	2.5	15.1	**0.166**	16.5	16.9	-0.4	29.3	**-0.014**
	Word problems (%)	33.9	29.9	4.0	19.4	**0.206**	33.9	35.0	-1.1	30.1	**-0.037**

Table B2. Program effect and effect sizes for EGRA (English and Kiswahili) and EGMA for Class 2, disaggregated by type of school

Class 2 PRIMR Effect											
		Formal					LCPS				
Subject	Subtask	PRIMR	Control	Effect	Standard deviation	Effect size	PRIMR	Control	Effect	Standard deviation	Effect size
English	Letter-sound fluency (clpm)	45.5	22.5	23.0	23.7	**0.970**	59.6	33.8	25.8	39.7	**0.650**
	Decoding fluency (cwpm)	30.1	22.5	7.6	15.6	**0.487**	40.4	32.0	8.4	21.5	**0.391**
	Segmenting (%)	51.4	30.3	21.1	34.0	**0.621**	65.6	36.5	29.1	45.3	**0.642**
	Oral reading fluency (cwpm)	49.7	34.1	15.6	28.2	**0.553**	74.1	57.2	16.9	41.1	**0.411**
	Vocabulary (%)	66.3	61.7	4.6	17.0	**0.271**	79.8	75.1	4.7	18.2	**0.258**
	Reading comprehension (%)	37.6	19.8	17.8	34.4	**0.517**	66.4	44.4	22.0	53.2	**0.414**
	Fluent (%)	33.3	8.8	24.5	38.4	**0.638**	60.9	41.9	19.0	60.7	**0.313**
Kiswahili	Letter-sound fluency (clpm)	49.0	27.3	21.7	26.2	**0.828**	59.1	36.5	22.6	38.7	**0.584**
	Syllable fluency (cspm)	50.4	39.3	11.1	23.4	**0.474**	58.1	44.8	13.3	28.6	**0.465**
	Decoding fluency (cwpm)	24.1	19.3	4.8	14.1	**0.340**	31.2	24.7	6.5	19.6	**0.332**
	Oral reading fluency (cwpm)	30.6	24.5	6.1	16.9	**0.361**	39.5	31.9	7.6	20.9	**0.364**
	Reading comprehension (%)	40.5	33.0	7.5	25.4	**0.295**	56.4	42.3	14.1	37.4	**0.377**
	Listening comprehension (%)	67.1	60.7	6.4	22.1	**0.290**	74.1	69.7	4.4	29.6	**0.149**
	Maze (%)	21.7	18.6	3.1	12.1	**0.256**	29.4	26.0	3.4	18.8	**0.181**
	Fluent (%)	18.9	7.5	11.4	31.5	**0.362**	34.5	20.7	13.8	56.8	**0.243**
Math	Number identification fluency (cpm)	26.5	23.3	3.2	8.8	**0.364**	34.7	29.8	4.9	15.5	**0.316**
	Quantity discrimination (%)	67.6	69.4	-1.8	20.3	**-0.089**	79.3	79.1	0.2	27.5	**0.007**
	Missing number (%)	51.4	41.2	10.2	19.1	**0.534**	60.2	50.2	10.0	26.1	**0.383**
	Addition fluency (cpm)	11.8	10.3	1.5	3.7	**0.405**	13.4	11.9	1.5	5.4	**0.278**
	Addition level 2 (%)	44.0	28.8	15.2	28.9	**0.526**	55.5	46.4	9.1	41.3	**0.220**
	Subtraction fluency (cpm)	8.4	7.1	1.3	3.6	**0.361**	9.8	8.2	1.6	4.6	**0.348**
	Subtraction level 2 (%)	28.9	15.3	13.6	26.4	**0.515**	34.8	25.6	9.2	38.8	**0.237**
	Word problems (%)	46.9	41.5	5.4	23.1	**0.234**	49.2	45.3	3.9	32.0	**0.122**

Table B3. Program effects and effect sizes for EGRA (English and Kiswahili), disaggregated by gender

Subtask	Language	Gender	Treatment		Control		Program impact		
			Mean	Standard error	Mean	Standard error	Standard deviation	Program effect	Effect size
Letter sounds (clspm)	English	Boys	45.5	1.1	25.2	1.5	26.6	20.3	0.763
		Girls	48.5	1.1	26.1	1.5	26.5	22.4	0.845
	Kiswahili	Boys	46	1.1	26.1	1.2	26.6	19.9	0.748
		Girls	48.9	1.3	31.3	1.6	27.3	17.6	0.646
Syllable fluency (cspm)	English	Boys							
		Girls							
	Kiswahili	Boys	44.4	1.2	32.2	1.1	25.5	12.2	0.479
		Girls	47	1	36.8	1.6	24.8	10.2	0.411
Decoding fluency (cwpm)	English	Boys	27.9	0.7	20	0.9	18.4	7.9	0.429
		Girls	28.8	0.6	21.3	0.9	17.1	7.5	0.439
	Kiswahili	Boys	21.3	0.7	16	0.7	16.0	5.3	0.331
		Girls	22.4	0.6	17	0.8	15.3	5.4	0.354
Segmenting (%)	English	Boys	55.2	1.6	31	2.1	37.2	24.2	0.651
		Girls	56	1.5	31.5	1.9	35.5	24.5	0.690
	Kiswahili	Boys							
		Girls							
Oral reading fluency (cwpm)	English	Boys	43.9	1.3	30.6	1.6	33.3	13.3	0.399
		Girls	46.1	1.1	32.2	1.6	30.8	13.9	0.451
	Kiswahili	Boys	26.5	0.7	20.2	0.8	18.7	6.3	0.337
		Girls	28.1	0.7	21.1	0.9	17.6	7	0.399
Vocabulary (%)	English	Boys	67	0.8	59.8	1.1	20.5	7.2	0.351
		Girls	66.8	0.7	61.5	1.4	19.8	5.3	0.268
	Kiswahili	Boys							
		Girls							
Reading comprehension (%)	English	Boys	34.3	1.5	18.2	1.4	36.1	16.1	0.446
		Girls	34.1	1.4	20.5	1.8	36.5	13.6	0.373
	Kiswahili	Boys	36	1.2	25.8	1.3	29.2	10.2	0.349
		Girls	35.7	1	25.8	1.5	26.9	9.9	0.368
Listening comprehension (%)	English	Boys							
		Girls							
	Kiswahili	Boys	62.7	1	55.7	1.3	26.9	7	0.260
		Girls	64.3	0.9	57.7	1.7	27.1	6.6	0.244
Maze (%)	English	Boys							
		Girls							
	Kiswahili	Boys	19.9	0.6	16.9	0.7	14.5	3	0.208
		Girls	20.2	0.4	16.7	0.6	13.2	3.5	0.266
Fluent (%)	English	Boys	26.3	1.5	12.4	1.6	39.0	13.9	0.357
		Girls	30.2	1.5	12.9	1.6	39.5	17.3	0.438
	Kiswahili	Boys	15.8	1.1	7.7	1.3	32.0	8.1	0.253
		Girls	15.9	1.2	5.7	1.1	29.7	10.2	0.343

Table B4. Program effect and effect sizes for EGMA, disaggregated by gender

Subtask	Gender	Treatment		Control		Program impact		
		Mean	Standard error	Mean	Standard error	Standard deviation	Program effect	Effect size
Number identification fluency (cpm)	Boys	24.7	0.5	21	0.4	11.7	3.7	0.316
	Girls	24.3	0.4	21.7	0.5	10.3	2.6	0.252
Quantity discrimination (%)	Boys	61.1	1	59.4	1.4	27.4	1.7	0.062
	Girls	58.7	0.9	59	1.6	26.7	-0.3	-0.011
Missing number (%)	Boys	43.6	0.9	36.9	1.1	22.0	6.7	0.305
	Girls	43.4	0.9	36.8	1	20.1	6.6	0.329
Addition fluency (cpm)	Boys	10.2	0.2	9.2	0.2	4.7	1	0.213
	Girls	10.1	0.2	9.3	0.3	4.8	0.8	0.168
Addition level 2 (%)	Boys	35	1.3	27.1	1.9	31.7	7.9	0.250
	Girls	33.3	1.2	26.5	1.8	30.0	6.8	0.227
Subtraction fluency (cpm)	Boys	7.1	0.2	6.2	0.2	4.1	0.9	0.220
	Girls	7.1	0.2	6.2	0.2	4.3	0.9	0.212
Subtraction level 2 (%)	Boys	21.9	1.2	15.7	1.6	25.8	6.2	0.240
	Girls	21.8	1.3	15.1	1.5	24.8	6.7	0.270
Word problems (%)	Boys	40.9	1.1	38.2	1.7	25.9	2.7	0.104
	Girls	40.4	1.2	36.6	2	25.7	3.8	0.148

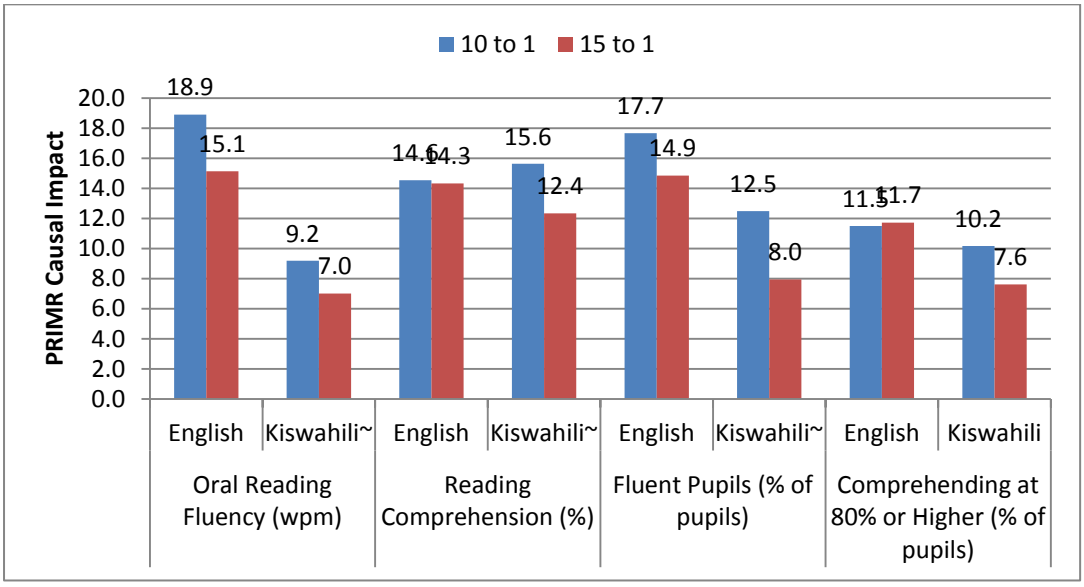
Annex C. Results of Linear Regression Analyses on Ratio of Tutors/Coaches to Schools

This annex elaborates on the researchers' data analyses around the ratio of coaches/tutors to schools and its program effects.

Linear regression models showed that the number of teachers to which a TAC tutor was assigned predicted 19.5% of the variation in the number of visits a TAC tutor made, while the same figure was 38.7% for instructional coaches. The models also highlighted that for every additional teacher a TAC tutor was assigned to, they made an additional 0.13 visits. The corresponding figure for instructional coaches was 0.04. In other words, the PRIMR system created incentives for TAC tutors and coaches who otherwise made very infrequent visits to increase those visits, and even PRIMR TAC tutors with large zones were likely to visit more schools.

The PRIMR research design allowed us to compare whether there were statistically significant differences in the magnitude of the PRIMR causal effect for key variables in PRIMR between pupils in clusters with 10 schools and those in clusters of 15 schools. ***Figure C1*** shows the causal gains (not the mean scores). It also indicates where the differences between the effect for 10:1 and the effect for 15:1 were statistically significantly different. None of the assessments showed statistical significance at the .05 level, although three of the assessments were significant at the .10 level (signified by a ~). All three of the areas that had a differential impact at the .10 level were in Kiswahili, namely oral reading fluency (effect size 0.102 SD), reading comprehension percentage score (effect size 0.095 SD), and the proportion of pupils reading at the KNEC benchmark (effect size (0.095 SD). There were no statistically significant differences in the impact of 10:1 and 15:1 clusters for the English outcome variables.

Figure C1. PRIMR causal impacts over control for 10:1 clusters and 15:1 clusters



"~" denotes statistically significant difference between 10:1 and 15:1 clusters.

Using the same set of outcome variables, we fit additional regression models within the sample of public schools using the number of schools in the zone as the predictor variable against the key variables of interest for the PRIMR initiative. While the size of the zone was not randomly assigned, this analysis allowed us to determine whether the directionality was similar to what we expected and what the data from the randomized controlled trial suggested. ***Table C1*** presents our findings.

We found that the zone size was a statistically significant predictor of English oral reading fluency in Class 1 (*p*-value .07) and Class 2 (*p*-value .07), of Kiswahili oral reading fluency in Class 2 (*p*-value .03), of English comprehension in Class 2 (*p*-value .06), of Kiswahili comprehension in Class 2 (*p*-value .04), of the proportion of pupils reading at benchmark in English in Class 1 (*p*-value .02), of the proportion reading at benchmark in Kiswahili in Class 2 (*p*-value .02), and of the proportion of pupils reading at benchmark in Kiswahili in Class 2 (*p*-value .04). The magnitude of the relationship was nontrivial. For example, in English Class 2, the difference in oral reading fluency was 0.9 cwpm per additional school. For situations like the LCPS, where the difference was 5 schools between the 10:1 and 15:1 schools, this suggests a 4.5 cwpm gap as a result of zone size. For comprehension, the magnitude of the effect was similar. For Kiswahili Class 2 comprehension percent score, the difference in comprehension rates associated with 5 more schools in a zone was 4.6%. It is worth noting that several of the models showed no statistically significant difference.

Table C1. Number of schools in zone as predictor variable

Item	Language	Grade	Coefficient	*T*	*p*-value	R^2
Oral reading fluency (cwpm)	English	1	-0.66 (0.36)	-1.82	.074~	.009
		2	-0.91 (0.50)	0.50	.073~	.008
	Kiswahili	1	0.10 (0.29)	0.34	.734	.000
		2	-0.73 (0.32)	-2.26	.028*	.014
Reading comprehension (% correct)	English	1	-0.07 (0.30)	-0.24	.812	.000
		2	-0.85 (0.44)	-1.92	.060~	.008
	Kiswahili	1	-0.30 (0.37)	-0.80	.425	.003
		2	-0.92 (0.44)	-2.07	.042*	.012
Fluent reader (% of population)	English	1	-0.81 (0.34)	-2.36	.022*	.012
		2	-0.40 (0.56)	-0.70	.484	.001
	Kiswahili	1	0.39 (0.30)	1.30	.198	.004
		2	-1.17 (0.50)	-2.34	.022*	.011

Item	Language	Grade	Coefficient	*T*	*p*-value	R^2
Comprehending 80% or higher (% of population)	English	1	-0.10 (0.18)	-0.48	.630	.000
		2	-0.54 (0.47)	-1.15	.257	.002
	Kiswahili	1	0.20 (0.25)	0.80	.429	.002
		2	-1.06 (0.51)	-2.08	.041*	.009

"~" denotes statistically significant difference between 10:1 and 15:1 clusters.
*$p < 0.05$.

We were also interested in whether the number of classroom visits in a zone or a cluster had an impact on pupil achievement. In order to answer the question, we included a variable at the pupil level that noted the number of visits that each TAC tutor or coach made during the year, and the number of visits that the TAC tutor or coach averaged per teacher in the zone. We fit regression models with those two variables separately as predictors, and English and Kiswahili oral reading fluency as outcome variables. The models were fit separately for the formal and LCPS. on a subsample of the PRIMR data set that included only treatment schools.[15] This was because we did not have externally collected reliable data on the number of observations that control TAC tutors undertook, if any. Our findings are presented in ***Table C2.***

In order to determine whether the number of visits had a relationship with the PRIMR effect, the reader should examine the *p*-values associated with the coefficients of interest. Table C2 includes models that regressed the total teacher visits or the average visits per teacher on English and Kiswahili fluency, for both public and LCPS. For English fluency, the number of coach visits per teacher in LCPS had a statistically significant relationship at the .10 level, whereby for every additional visit per year, pupils would read 0.85 words per minute higher (*p*-value .07). For Kiswahili fluency, the total number of visits in a zone had a positive statistically significant relationship with Kiswahili fluency in public schools (*p*-value .02). For Kiswahili fluency, the average number of visits per teacher in a LCPS cluster had a statistically significant relationship with fluency, such that for every additional visit per year, pupils read 0.37 wpm more. The R^2 for each of these models was extremely low (less than 1% of variation), so while these results are significant, they have a very minor relationship with the outcome of interest.

Table C2. OLS regression results for models fit for public and LCPS samples estimating the relationship between numbers of schools visited and outcomes

Item	Model	Measure	Coefficient	*T*	*p*-value	R^2
English fluency	Total teacher visits Public	Estimate	.024 (.042)	0.59	.56	.001
		Constant	32.17 (5.50)	5.85	<.001	

[15] One zone that had only six observations for the entire year was excluded from this analysis.

Item	Model	Measure	Coefficient	*T*	*p*-value	R^2
	Total teacher visits LCPS	Estimate	-.004 (.016)	-0.23	.82	.000
		Constant	61.33	4.86	<.001	
	Average visits per teacher Public	Estimate	-1.69 (1.68)	-1.01	.32	.003
		Constant	39.20 (4.86)	8.07	<.001	
	Average visits per teacher LCPS	Estimate	.853 (.466)	1.83	.07	.004
		Constant	50.17 (5.84)	8.59	<.001	
Kiswahili fluency	Total teacher visits Public	Estimate	.052 (.022)	2.38	.02	.009
		Constant	16.55 (2.57)	6.44	<.001	
	Total teacher visits LCPS	Estimate	-.006 (.008)	-0.74	.46	.001
		Constant	34.41 (2.44)	14.08	<.001	
	Average visits per teacher Public	Estimate	1.35 (.89)	1.51	.14	.004
		Constant	19.21 (2.37)	8.10	<.001	
	Average visits per teacher LCPS	Estimate	.372 (.214)	1.74	.09	.003
		Constant	28.25 (2.75)	10.26	<.001	

The findings in this annex are organized around three research questions. First, we investigated **whether the size of the zone or cluster had an impact on the number of visits.** The results showed that for both public school TAC tutors and LCPS instructional coaches, those with more teachers did observe more classrooms. This is, at least in part, a response to the incentives within the PRIMR program, which reimbursed TAC tutors based on the proportion of the teachers in their zone that they observed on a monthly basis. This shows, therefore, that the PRIMR incentive program was relatively effective in increasing classroom visits. The TAC tutors in the PRIMR initiative on average observed the teachers only 2.5 times each in 2013. On the other hand, the coaches in the 10:1 zones observed their teachers 12.2 times in a year, compared with those in 15:1 zones that observed 11.2 times.

Second, the randomized controlled trial research design of PRIMR made it possible to estimate **the relative effectiveness of organizing schools into clusters of 10 or clusters of 15.** Those options were chosen by PRIMR because they were the range of the majority of the other known supervisor, tutor, or coach instructional support programs in East Africa. There is a dearth, however, of literature driving the decision of what ratio is the most appropriate and cost-effective. To answer this research question, we fit eight OLS regression models that estimated the causal impact of PRIMR for both 10:1 and 15:1 school ratios, and then post-hoc

tests that allowed us to determine whether any apparent differences in the impact were significant. Of the eight models, three showed statistically significant differences at the .10 significance level. The effect size of these three impacts was in the range of 0.1 SD. Given that changing the ratio from 15:1 to 10:1 would require just below a 50% increase in salary, our cost-effectiveness analyses showed that even if 0.1 SD were the impact of 10:1 over 15:1 across all the measures (not just 3 out of 8), it would not be a cost-effective investment.

Third, using cross-sectional data in a non-causal design, we fit OLS regression models to **determine whether the size of the zone in the formal sector had a relationship with literacy outcomes.** Several of the variables showed statistically significant differences, as shown in Table C2. Below ***Figure C2*** portrays the size of the relationship for these variables in formal schools. This figure shows which variables were statistically significantly related to the size of the zone (in blue), and the magnitude of that relationship if the hypothetical zone had had five more schools in it. Additional analysis showed that the magnitude of these effects was relatively large, with some of the effect sizes as large as 0.35 SD. The directionality of the statistically significant relationships lent credence to the view that TAC tutors in Kenya with large zones were less likely to make a significant impact on pupil outcomes than those with smaller zones. Given the flexibility and importance of the TAC tutor program in Kenya, and the relatively large effect size of the differences in the public school sector compared to the LCPS sector, it appears that increasing the TAC tutor workforce to reduce the ratio of schools by five might be a cost-effective investment in Kenya. Currently, the TSC is currently considering this sort of change.

Figure C2. Difference in outcomes for pupils in schools in zones with five more (hypothetical) schools added