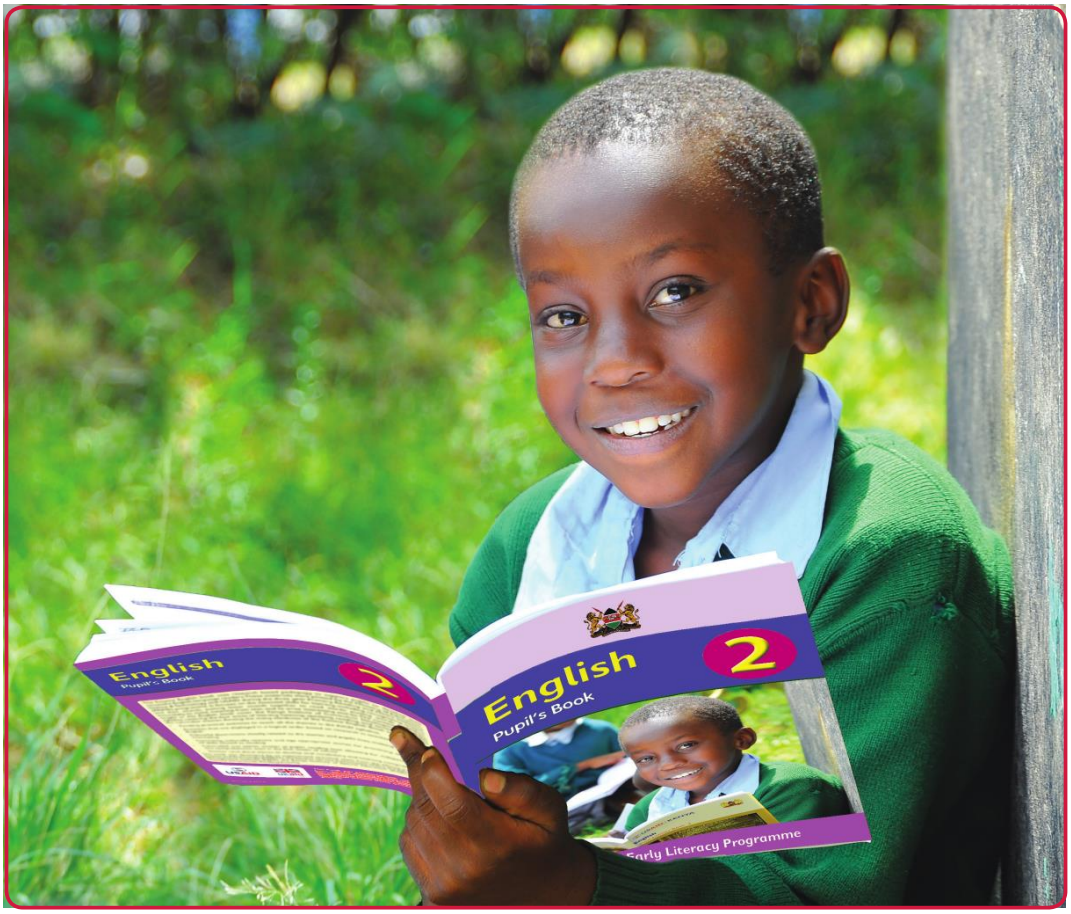
REPUBLIC OF KENYA
Ministry of Education

Tusome External Evaluation Midline Report



USAID
FROM THE AMERICAN PEOPLE

UKaid
from the British people

TUSOME EXTERNAL EVALUATION – MIDLINE REPORT

April 14, 2017

Contracted under AID-615-TO-16-00012
Midline Performance Evaluation of the Tusome Activity in Kenya

This midline report is a follow-up to the baseline report from January 25, 2016. The baseline report was also prepared independently by MSI.

April 2017

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by Elizabeth Freudenberger and Jeff Davis on behalf of Management Systems International, a Tetra Tech Company.

CONTENTS

TABLES

FIGURES

ACRONYMS

APBET	Alternative Provision of Basic Education and Training
BOM	Board of Management
CSO	Curriculum Support Officer
CWPM	Correct Words Per Minute
DFID	U.K. Department for International Development
EGRA	Early Grade Reading Assessment
ESQAC	Education Standards and Quality Assurance
GOK	Government of Kenya
IT	Information Technology
KEA	Kenya and East Africa
KICD	Kenya Institute for Curriculum Development
KNEC	Kenya National Examinations Council
MOE	Ministry of Education
MSI	Management Systems International
ORF	Oral Reading Fluency
PRIMR	Primary Math and Reading Initiative
QCO	Quality Control Officer
RTI	RTI International
SART	Secondary Analysis for Results Tracking
SOW	Statement of Work
TSC	Teachers Service Commission
USAID	U.S. Agency for International Development

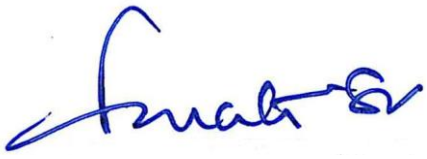
FOREWARD

The Government of Kenya realizes that to be internationally competitive and economically viable, the country requires an education system that will produce citizens who are innovative and are able to perform complex tasks and engage in lifelong learning. The education system should also produce individuals capable of problem solving, taking decisions, require minimum supervision, assume responsibility and have better reading, quantitative reasoning and expository skills.

The Constitution of Kenya (2010) Article 43 recognizes that every person has a right to education, and article 53(b) states that every child has a right to free and compulsory basic education. The Basic Education Act (2013) and the National Education Sector Plan (NESP) 2013 – 2018 emphasize the need to provide quality basic education. This will provide the essential foundation for successful future learning and contribution to Kenya's social and economic aspirations as set out in Kenya Vision 2030.

In line with the above, the Ministry of Education has embarked on key interventions towards improving the quality of education. This is reflected in the number of programmes such as Tusome, designed to address quality issues especially in the area of literacy in lower primary.

The Ministry shall continue providing guidelines to facilitate the implementation of Tusome and other education programmes. The report findings of the Midline evaluation study are a major milestone in literacy provision for future posterity and competency development of learners. I thereby urge all Stakeholders, Partners and Civil Society to continue supporting the Education Sector in the implementation of strategies to promote the quality of education.



Fred Matiang'i, PhD, EGH
CABINET SECRETARY
MINISTRY OF EDUCATION

PREFACE

The vision of the Ministry of Education is the provision of quality Education and Training for sustainable development. In order to realize this, all Ministry programmes are anchored on International, Regional and National commitments. In this regard these commitments have been factored in MOE's legal, policy and strategy documents.

The Ministry of Education has had an increased focus on the quality of Education in lower primary, particularly in the areas of literacy and numeracy. It is against this background that Tusome National literacy programme supported by USAID and DFID was conceptualized. The implementation of the programme, started in 2015 in all Public Schools and 1,000 Alternative Provision of Basic Education & Training (APBET) Institutions with an aim of improving literacy learning outcomes in lower primary.

The Tusome Midline Evaluation study was undertaken to monitor the progress of learner competencies towards the achievement of the set goal. The study findings were prepared on the basis of the baseline survey.

The report has highlighted key findings, challenges and recommendations. It is envisaged that the findings will enhance stakeholders' understanding of the programme success and gaps towards mitigation, ownership and sustainability. I am therefore calling upon all stakeholders to work together towards the implementation of the recommendations for the success of the programme.

Dr. Belio R. Kipsang, CBS
PRINCIPAL SECRETARY
STATE DEPARTMENT OF BASIC EDUCATION

ACKNOWLEDGMENT

The Tusome Midline Evaluation study was undertaken in September – October, 2016 through a collaborative effort between the Ministry of Education, TSC and Partners.

I wish to appreciate the strategic leadership of the **Cabinet Secretary Fred Matiang'i** PhD, EGH and the **Principal Secretary Dr. Belio R. Kipsang**, CBS towards the undertaking of the evaluation study. Through their commitment the recommendations articulated in the report will be implemented for programme impact and sustainability.

The Ministry of Education appreciates the financial support provided by USAID for the evaluation exercise. I also wish to take this opportunity to thank the Technical Working Teams from the Management Systems International (MSI), Ministry of Education and TSC for the successful completion of the study.

It is my hope that the findings and recommendations will be useful to stakeholders in the improvement of quality of education.

Leah K. Rotich (Mrs.), MBS
DIRECTOR GENERAL
STATE DEPARTMENT OF BASIC EDUCATION

EXECUTIVE SUMMARY

EVALUATION PURPOSE AND AUDIENCE

The purpose of the Tusome External Evaluation is to establish measurements for an evaluation of the five-year (2014–2019) Tusome ("Let's Read" in Kiswahili) programme.

The evaluation is a non-experimental cross-sectional study with measurements at three time points: baseline, midline and endline. The evaluation compares reading outcomes at the baseline (pre-test) to those at the midline and endline (post-tests). In addition, it examines pupil, teacher, head teacher, school and household factors for their relationships to reading outcomes and any changes in those relationships over time.

The main audiences for the study are the following groups: 1) the Government of Kenya (GOK) and Ministry of Education (MOE); 2) USAID and DFID; and 3) Research Triangle Institute (RTI International), the implementing partner. Other stakeholders include the Teachers Service Commission (TSC), semiautonomous government agencies and county governments.

EVALUATION QUESTIONS

USAID/Kenya and East Africa (KEA) asked the evaluation team to address the following evaluation questions:

1. What proportion of students can demonstrate they can read grade-level text (within Kenya's curricular goals) by the end of Standards 1 and 2?
2. What are the levels of Classes 1 and 2 pupils on reading subtasks?[1]
3. What school-level and institutional factors influence reading outcomes when implementing at scale, and how?
4. What community-level factors influence reading outcomes when implementing at scale, and how?
5. To what extent have the Tusome Early Grade Reading (EGR) activity components been implemented in schools nationwide?
6. To what extent can any incremental changes in early grade reading outcomes throughout Kenya be correlated with or attributed to the scale-up of Tusome?

EVALUATION METHODS

Management Systems International (MSI) led the Tusome baseline study using multiple data collection methods, including an early grade reading assessment (EGRA); surveys of pupils, teachers, head teachers, curriculum support officers (CSO) and households; and classroom observation. The EGRA assessment tool was developed during the baseline and includes eight subtasks in English and six subtasks in Kiswahili. The midline included developing additional data collection tools, revising the baseline surveys, recruiting and training supervisors and enumerators, administering the tool and surveys in the same sample schools as the baseline, ensuring quality control, establishing the reliability of the assessment tool, and analyzing the data.

For the midline, the evaluation team assessed pupils from the same 204 schools sampled for the baseline. Through discussions with USAID, MOE and RTI, the evaluation team created the sampling

1 Evaluation question revised in consultation with USAID. See Evaluation Questions in the body of the report for more information.

frameworks and set up the design for a national sample in 2015. Using a three-stage cluster sampling procedure from a sampling frame of 22,154 formal public schools and 1,000 non-formal (or Alternative Provision of Basic Education and Training – APBET) schools, the evaluation team drew a clustered, random sample, resulting in a target of 4,896 total pupils comprising 2,448 boys and 2,448 girls divided equally between Class 1 and Class 2.

EVALUATION STRENGTHS AND LIMITATIONS

The evaluation methodology and implementation resulted in valid, reliable data for the midline evaluation, including the changes from baseline to midline. The data collection tools and the analyses were sufficient for answering the midline evaluation questions. The tools covered a variety of aspects of the Tusome reading interventions by collecting data from the pupils, teachers, head teachers, communities, CSOs, and education officials.

This evaluation was designed to collect data from the same schools as the baseline to document change over time. While MSI collected data from all 204 schools in the sample, technical difficulties resulted in data from one school being lost. The evaluation team was able to include more than 99 percent of schools in the midline database.

A few limitations to the survey data should be taken into consideration when interpreting the pupil, teacher, head teacher and CSO survey results. Some pupil groups (e.g. by some of the age and language groups) and most teacher and head teacher groups had small sample sizes, so any conclusions for those groups should be made with a high degree of caution. Confounding may lead to inaccurate interpretations, for example, a group of teachers with higher pupil scores may teach pupils in urban areas who generally have higher scores. Inconsistencies emerged when different kinds of respondents answered similar questions, for example, teachers and head teachers responded differently when asked if schools had libraries, and socioeconomic measures varied between pupil and parent responses.

Some limitations also existed with the household survey data, which were collected over the phone from Nairobi. Due to the team's efforts to maximize the survey of the pupils' households given the time, budget, and logistical constraints, the evaluation reached 49 percent of households. While this meant that data were not collected from half of the households, this was a high percentage of households – based on the team's experiences with similar surveys – reached in a cost-effective manner.

Finally, the timeline of the study should be noted as a potential limitation. USAID recommends collecting the baseline data before the start of interventions and then the midline (and endline) data at the same time point as the baseline data during the subsequent school year(s). However, due to various issues, the baseline data were collected in June and July 2015, shortly after Tusome started working with Class 1 teachers, but before it started working with Class 2 teachers. Then the midline data were collected in September and October 2016, somewhat after the same time point in the school year as the baseline data and thus providing slightly more time for learning during the school year (even taking into consideration the school break in August). However, after consultations with USAID, the evaluation team did not make adjustments to the student scores to compensate for these issues.

FINDINGS

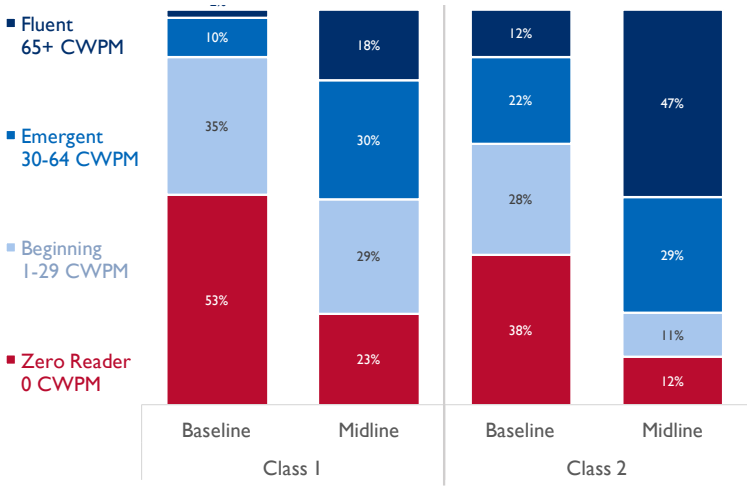
The evaluation team's key findings are listed below in responding to each evaluation question. A detailed analysis of findings is in the body of the report.

Evaluation Question 1: What proportion of students can demonstrate they can read grade-level text (within Kenya's curricular goals) by the end of Standards 1 and 2?

The Class 2 benchmarks for reading performance in English are 30 to 64 correct words per minute (CWPM) for emergent readers and 65 or more CWPM for fluent readers. At midline, 30 percent of Class 1 pupils are emergent readers and 18 percent are fluent readers, while 29 percent of Class 2 pupils are emergent readers and 47 percent are fluent readers in English.

As shown in Figure 1, English reading performance improved between baseline and midline, with a lower percentage of pupils in the zero and beginning reader categories and a higher percentage in the emergent and fluent reader categories.

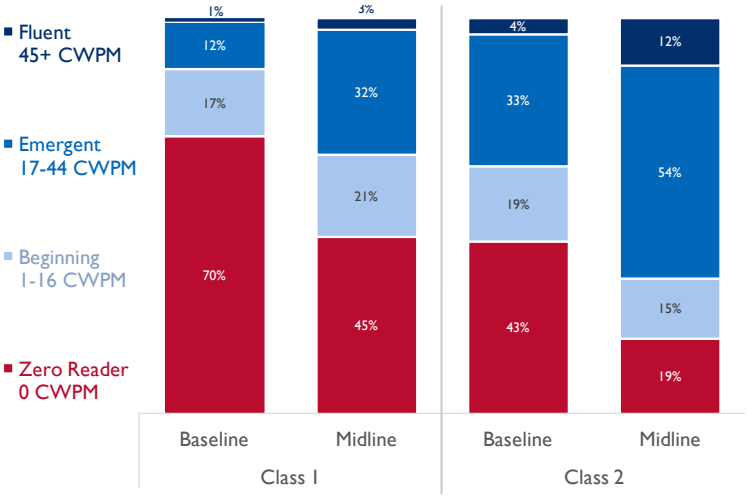
Figure 1: English Reading Performance Categories



The Class 2 benchmarks for reading performance in Kiswahili are 17 to 44 CWPM for emergent readers and 45 or more CWPM for fluent readers. At midline, 32 percent of Class 1 pupils are emergent readers and 3 percent are fluent readers, while 54 percent of Class 2 pupils are emergent readers and 12 percent are fluent readers in Kiswahili.

As shown in Figure 2, Kiswahili reading performance also improved between baseline and midline, with a lower percentage of pupils in the zero and beginning reader categories and a higher percentage in the emergent and fluent reader categories.

Figure 2: Kiswahili Reading Performance Categories



Evaluation Question 2: What are the levels of Classes 1 and 2 pupils on reading subtasks?

Pupils have shown improvements on all EGRA subtasks in both languages and classes between baseline and midline. The raw scores are shown below in Table 1 for English and Table 2 for Kiswahili. All gains between baseline and midline are statistically significant at the 0.01 level.

Table 1: English Raw Reading Scores

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Phoneme segmentation	1.1	3.8	2.6*	0.6	5.0	4.5*
Letter sound knowledge	15.1	26.3	11.3*	10.2	32.6	22.4*
Invented/non-word decoding	5.7	10.4	4.7*	10.4	18.6	8.3*
Vocabulary	5.9	7.8	1.9*	8.2	10.2	1.9*
Passage reading (A)	10.6	22.3	11.7*	23.8	43.6	19.9*
Reading comprehension (A)	0.2	0.5	0.3*	0.5	1.0	0.5*
Passage reading (B)	9.7	22.0	12.4*	21.8	44.2	22.5*
Reading comprehension (B)	0.2	0.8	0.6*	0.6	1.7	1.2*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 2: Kiswahili Reading Scores

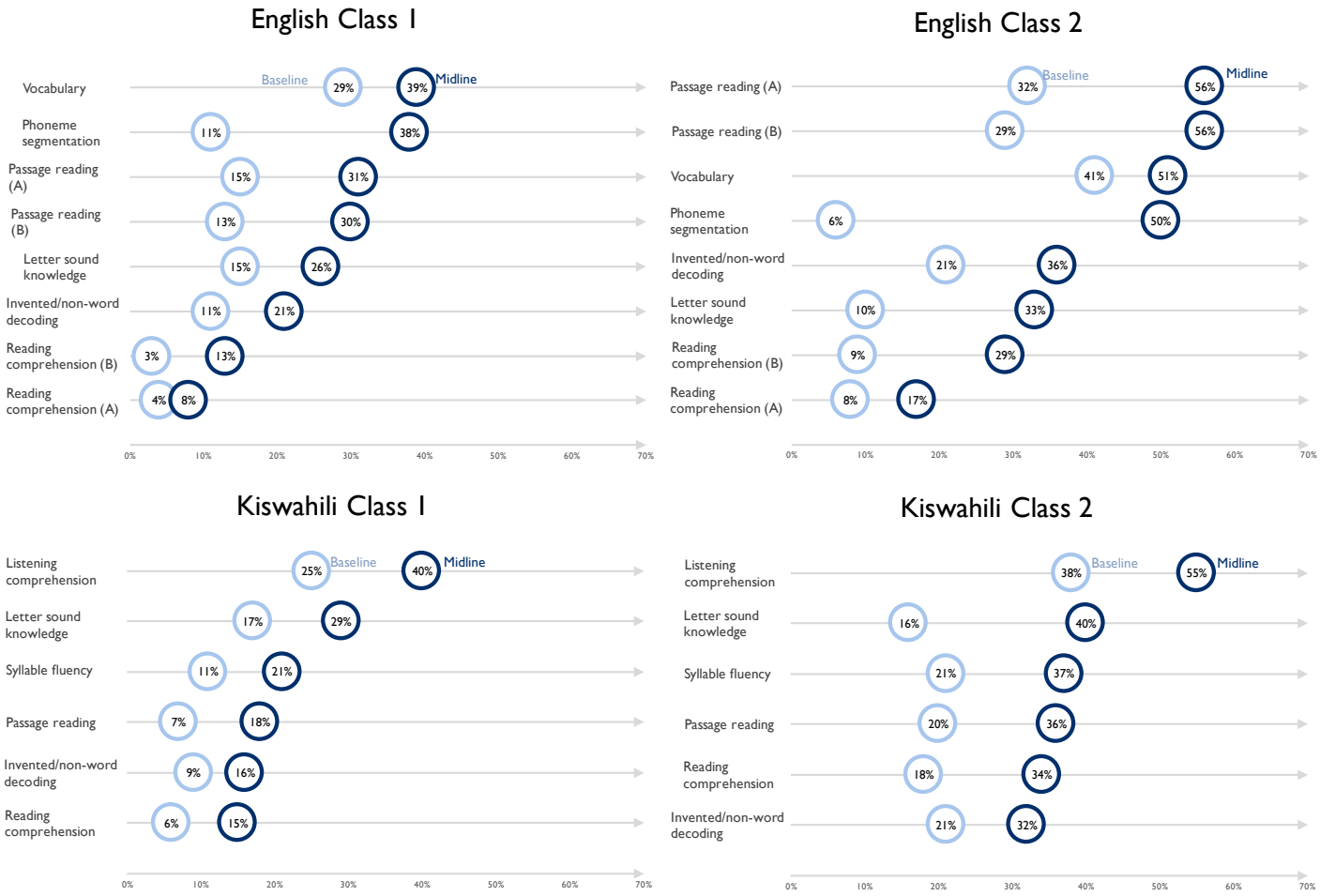
Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Letter sound knowledge	16.6	29.7	13.1*	16.2	39.7	23.4*
Syllable fluency	11.0	21.5	10.4*	20.9	37.5	16.6*
Invented/non-word decoding	4.7	8.3	3.6*	10.2	16.1	5.8*
Passage reading	4.9	12.2	7.3*	13.5	24.5	11.0*
Reading comprehension	0.4	0.9	0.5*	1.1	2.0	1.0*
Listening comprehension	1.2	2.0	0.8*	1.9	2.0	0.9*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Figure 3 below illustrates pupil performance in terms of percent reading scores at baseline and midline. In other words, for each subtask, the average raw score is divided by the total possible score, thus allowing for comparisons across the subtasks in the same metric.

In English, Class 1 pupils performed best on the vocabulary and phoneme segmentation subtasks, followed by the passage reading subtasks. Class 2 pupils performed best on the passage reading subtask, followed by the vocabulary and phoneme segmentation subtasks. In Kiswahili, pupils in both classes performed best on the listening comprehension subtasks. In both languages, scores for reading comprehension – the most difficult subtask – improved but remained the lowest in terms of overall performance.

Figure 3: Percent Correct Reading Scores



Evaluation Question 3: What school-level and institutional factors influence reading outcomes when implementing at scale, and how?

The evaluation team examined various pupil, teacher, head teacher and school variables and their associations with reading outcomes, namely oral reading fluency (ORF). Several school-level and institutional factors were found to be associated with better reading outcomes, including:

- Access to reading materials in the school;
- Access to a reading teacher's guide;
- Practice reading aloud and silently at school;
- Increased frequency of CSO observations;
- Increased frequency of lesson plans being reviewed;
- Full-day school shifts (versus half-day);
- Classroom libraries;
- Pupils of the correct age range (5 to 9 years old); and
- Female teacher or head teacher.

Evaluation Question 4: What community-level factors influence reading outcomes when implementing at scale, and how?

The evaluation team examined various pupil and household characteristics, and their associations with reading outcomes (i.e., ORF). Results from two characteristics – socio-economic status (SES, measured through both student questionnaires and household interviews) and parent education levels – are provided below.

SES did not have a strong association with ORF, except for the highest income households. For SES measured through the student questionnaire, the pupils in the upper part of the SES scale tended to have higher ORF scores, though the differences were often small. For the Class 1 pupils, the ORF gains between baseline and midline were similar for the bottom three SES groups, but almost twice as large for pupils in the highest SES group. For Class 2 in both languages, the gains were fairly similar across all SES groups. In contrast to Class 1, the lowest SES group showed the largest gains in Class 2.

The evaluation team also examined SES by constructing a wealth index from the household interviews. As illustrated in Figure 4, there was little difference in ORF scores for households in the first four quintiles, that is for 80 percent of the population. The highest quintile (top 20 percent of households) is associated with markedly higher ORF scores.

Figure 4: Oral Reading Fluency by Wealth Quintiles

In general, higher levels of education for the mothers and fathers are positively related to higher English and Kiswahili ORF scores. The trends are more pronounced for the English ORF scores than for the Kiswahili ORF scores, and at the highest levels of parental education.

Evaluation Question 5: To what extent have the Tusome Early Grade Reading (EGR) activity components been implemented in schools nationwide?

The evaluation team looked at data on implementation components collected through the teacher, head teacher and CSO interviews, as well as through classroom observation, and found a high level of national implementation. Key findings include:

- Ninety-eight percent of teachers have received at least some Tusome training. Thirty-eight percent of Class 1 and 48 percent of Class 2 teachers reported participating in five or more Tusome training sessions.
- Eighty-three percent of head teachers reported that they had received reading instruction training in the past 12 months.

- Ninety-nine percent of teachers had a Tusome teacher's guide in their classroom.
- Ninety-seven percent of Class 1 and 95 percent of Class 2 classrooms had at least one Tusome pupil's book per student.
- Ninety-six percent of classrooms had at least one exercise book per student.
- Eighty-four percent of Class 1 and 82 percent of Class 2 teachers reported being observed about once per term by a CSO.
- Ninety-six percent of Class 1 and 90 percent of Class 2 teachers reported being observed about once per term by their head teachers.
- Fifty-four percent of CSOs had observed 15 or more lessons in the last 30 days.

Evaluation Question 6: To what extent can any incremental changes in early grade reading outcomes throughout Kenya be correlated with or attributed to the scale-up of Tusome?

As Tusome is a national programme that is being implemented in all schools simultaneously, there is no control or comparison group to compare the activity's gains. However, the evaluation team did examine the effect sizes seen during the prior pilot study, PRIMR, to contextualize the gains seen between the Tusome baseline and midline.

The effect sizes for Tusome between baseline and midline range from 0.40 to 1.07 for Class 1 and from 0.41 to 2.57 for Class 2. In social science research, an effect size of 0.5 is considered to show a large impact. These effect sizes were higher than those of PRIMR, which ranged from 0.28 to 0.68 for Class 1 and 0.30 to 0.78 for Class 2.

The evaluation team also looked at the results of other USAID reading programmes in the region. Though direct comparisons between this study and other regional results are difficult due to methodological differences in programmes and assessments, the Kenya results were found to be about twice as high as those in Tanzania.

CONCLUSIONS

Based on the findings above, the evaluation team reached the following conclusions:

- The Tusome approach is having a strong, positive influence on reading outcomes, with relationships between project implementation and reading outcomes.
- Reading outcomes for Class 1 and 2 pupils greatly improved during the one-year period between the baseline and midline evaluations. While impressive gains have been made, continuing with the Tusome approach will be critical to sustaining or improving on those gains.
- The Tusome project has achieved a high level of national implementation of the activities at each level of the education system. Given that project activities such as CSO observations, in-service training and access to materials are associated with higher ORF scores, the high level of implementation across all schools appears to be a key part of its success. The effect sizes seen during the PRIMR pilot have been at least sustained, and in most cases strengthened, in the national scale up of Tusome.
- The evaluation methodology and implementation resulted in valid, reliable data for the midline evaluation, including the changes from baseline to midline.

RECOMMENDATIONS

Based on its fieldwork, data and workshops, the evaluation team has several recommendations.

The evaluation team recommends that the **Tusome project**:

1. Continue a high level of implementation fidelity in its support for materials, instruction and supervision in early reading activities to further increase reading gains.
2. Conduct additional analysis using the midline dataset to see what programmatic insights can be used for improved activity implementation.

The evaluation team recommends that **USAID/KEA**:

3. Use the findings of this evaluation in its continued support of materials, instruction and supervision in early grade reading activities.
4. Share the evaluation findings in other USAID early grade reading projects beyond Kenya to increase regional and international collaboration and learning.

The evaluation team recommends that the **MOE**:

5. Continue its support of early reading activities and evaluations to ensure further ministry ownership of the Tusome implementation and results.
6. Set benchmarks and targets for reading comprehension, in addition to the ORF benchmarks, to monitor pupil progress in comprehension over time.

The evaluation team recommends that the **team tasked with the endline evaluation**:

7. Use the data collection tools, sampling plan and data collection schedule used at midline to ensure valid, reliable and interpretable data.
8. Continue the strong collaboration with the MOE for the implementation of the study, including tools revision, training and data collection.

EVALUATION PURPOSE AND QUESTIONS

EVALUATION PURPOSE

The purpose of the Tusome External Evaluation is to establish initial measurements for an evaluation of the five-year (2014–2019) Tusome ("Let's Read" in Kiswahili) programme.

The evaluation is a non-experimental cross-sectional study with measurements at three time points: baseline, midline, and endline. The evaluation compares reading outcomes at the baseline (pre-test) to those at the midline and endline (post-tests). In addition, pupil, teacher, head teacher, school, and household factors are examined for their relationships to reading outcomes and any changes in those relationships over time.

The main audiences for the study are 1) the Government of Kenya (GOK) and Ministry of Education (MOE); 2) the U.S. Agency for International Development (USAID) and the U.K.'s Department for International Development (DFID); and 3) RTI International (RTI, the implementing partner). Other stakeholders include the Teachers Service Commission (TSC), semiautonomous government agencies and county governments.

EVALUATION QUESTIONS

USAID/Kenya and East Africa (KEA) asked the evaluation team to address the following evaluation questions:

1. What proportion of students can demonstrate they can read grade-level text (within Kenya's curricular goals) by the end of Standards 1 and 2?
2. What are the levels of Classes 1 and 2 pupils on reading subtasks?
3. What school-level and institutional factors influence reading outcomes when implementing at scale, and how?
4. What community-level factors influence reading outcomes when implementing at scale, and how?
5. To what extent have the Tusome Early Grade Reading (EGR) activity components been implemented in schools nationwide?
6. To what extent can any incremental changes in early grade reading outcomes throughout Kenya be correlated with or attributed to the scale-up of Tusome?

Evaluation question two in the Statement of Work was "What proportion of students are able to answer comprehension questions after reading grade level text (within Kenya's curricular goals) by the end of Standards 1 and 2?" This question was revised in consultation with USAID based on comments to the draft evaluation report in order to address a wider range of factors in pupil reading performance.

The evaluation team's findings, conclusions and recommendations are detailed by evaluation question in the respective sections.

PROJECT BACKGROUND

The 2013-2018 National Education Sector Plan (NESP) Implementation Plan notes that after the GOK passed a reform package in 2003 that guaranteed free primary education, pupil enrollment increased dramatically, with near gender parity. However, the quality of instruction received in schools suffered, including in the core skill of reading, as the increase in enrollment was not accompanied by an increase in supportive resources.

This lack of reading skills negatively impacts academic performance across subjects, retention and repetition of grades, which all have major implications for cost and for the achievement of Kenya's Vision 2030 goals. The NESP Implementation Plan includes raising literacy and numeracy levels as one of its focuses. In line with GOK priorities and USAID's strategic focus on early grade reading in its 2011 Education Strategy, Tusome addresses the need to improve learning outcomes for young children in all Kenyan schools, including public formal, public non-formal (Alternative Provision of Basic Education and Training, or APBET) and low-cost private schools, both of which teach the content in the approved Kenya Institute for Curriculum Development (KICD) syllabi.

Starting in 2007, USAID/KEA and MOE ran a one-year randomized control trial in 40 schools in the Malindi district. Building on these findings, USAID/KEA funded a three-year applied research programme, Primary Math and Reading Initiative (PRIMR). According to the PRIMR Final Report, this activity reached 56,036 pupils across 547 formal public schools and APBET institutions in Nairobi, Kiambu, Nakuru, and Kisumu.

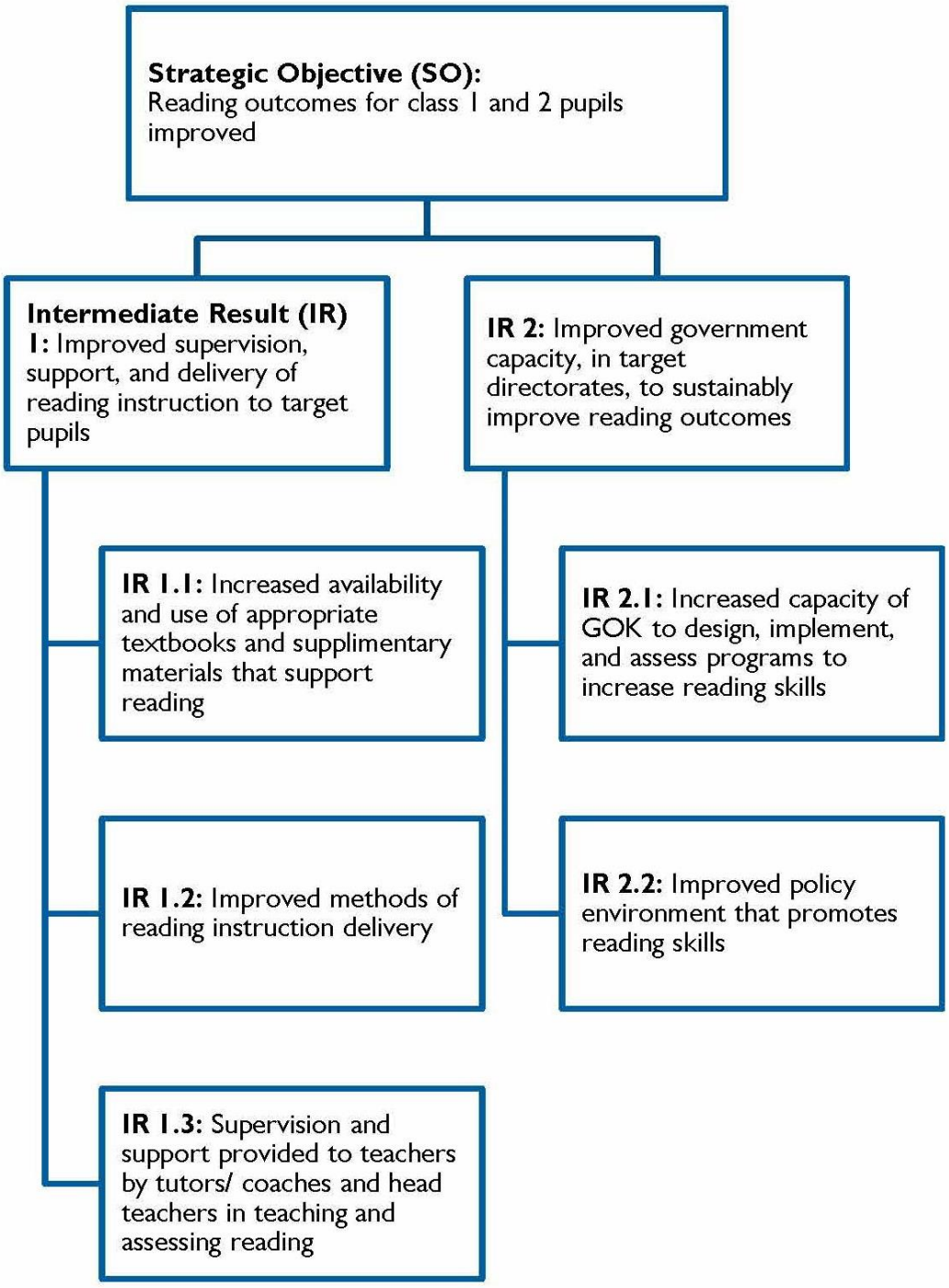
Following PRIMR, the MOE requested a national expansion of the PRIMR model. USAID/KEA awarded Tusome to RTI in 2014 in order to scale up the intervention nationally. This four-year, $55 million basic education initiative is a collaborative effort between USAID/KEA and DIFD to improve the reading skills of the approximately 5.4 million individual Kenyan children who began primary school during the 2015-2017 school years. The Tusome programme is intended to 1) scale up the previous (2011-2014) PRIMR pilot activity and 2) increase the capacity of the GOK to deliver and administer early grade reading programmes nationwide. Tusome will continue through July 2018, and has integrated options for transition to government ownership.

As detailed in Tusome's Performance Management Plan, the main strategic objective is "Reading outcomes for Class 1, Class 2 and Class 3 pupils improved." The means for achieving this objective are outlined in the two intermediate results: 1) improved supervision, support and delivery of reading instruction to target pupils and 2) improved government capacity, in target directorates, to sustainably improve reading outcomes. (Figure 5) The intended beneficiaries include:

1. 7.4 million primary school pupils (7.25 million children in public schools and 150,000 children in APBET institutions;
2. 76,000 Class 1, Class 2 and Class 3 teachers (covering all 22,344 public schools and 1,500 APBET institutions);
3. 23,844 primary school head teachers (22,344 head teachers in public schools and 1,500 head teachers in APBET institutions);
4. 1,376 curriculum support officers (CSOs) (1,292 for public schools and 84 for APBET institutions);[2] and
5. 1,500 senior education personnel.

2 Referred to as "Teacher Advisory Centre (TAC) tutors" in the Statement of Work and Performance Monitoring Plan.

Figure 5: Tusome Results Framework



EVALUATION METHODS AND LIMITATIONS

Management Systems International (MSI) led the Tusome baseline study using multiple data collection methods, including an early grade reading assessment (EGRA); surveys of pupils, teachers, head teachers, CSOs and households; and classroom observation. The EGRA tool was developed during the baseline. The midline included developing additional data collection tools, revising the baseline surveys, recruiting and training supervisors and enumerators, administering the tool and surveys in the same sample schools as the baseline, ensuring quality control, establishing the reliability of the assessment tool and analyzing the data. These steps are described below.

EVALUATION TEAM

The evaluation team consisted of a multidisciplinary group of international and national education, data collection, analysis and technology experts. The team was led by MSI, supported by data collection subcontractor Research Solutions Africa. Team members included:

- Evaluation Team Leader – Elizabeth Freudenberger
- Local Technical Expert – Charles Munene Kiura
- Technical Advisor – Jeff Davis
- Statistician – Idalia Rodriguez Morales
- Trainer – Sarah Fuller
- Quality Control Officers, Field Supervisors, and Enumerators
- Technology and Programming Specialists
- Kenya Support Project home office and field office staff

DATA COLLECTION TOOLS

The Tusome midline consisted of six data collection tools, shown in Table 1. The EGRA tool and the pupil, teacher and head teacher surveys were developed, piloted, revised and validated during the baseline assessment in 2015 in collaboration with the MOE. For a full description of the tool development process, please see the baseline technical report.[3]

Table 3: Midline Data Collection Tools

Data Collection Tools	Timeline	Collected at Baseline	Evaluation Questions
Early Grade Reading Assessment (EGRA)	October	Yes	1, 2, 6
Pupil Survey	October	Yes	3, 5, 6
Classroom and School Observation	October	No	3, 5, 6
Head Teacher and Teacher Survey	October	Yes	3, 5, 6
CSO Interview	October	No	3, 5
Household Survey	November / December	No	4, 6

The EGRA included 14 subtasks, eight in English and six in Kiswahili, which Table 2 details. For English, there were four pre-reading subtasks (phoneme segmentation, letter sound knowledge, invented/non-word decoding and vocabulary) and four reading subtasks (two passages each with passage reading and

3 Management Systems International (October 2015, revised January 2016). Tusome revised baseline study. Nairobi, Kenya: Kenya Support Program (KSP).

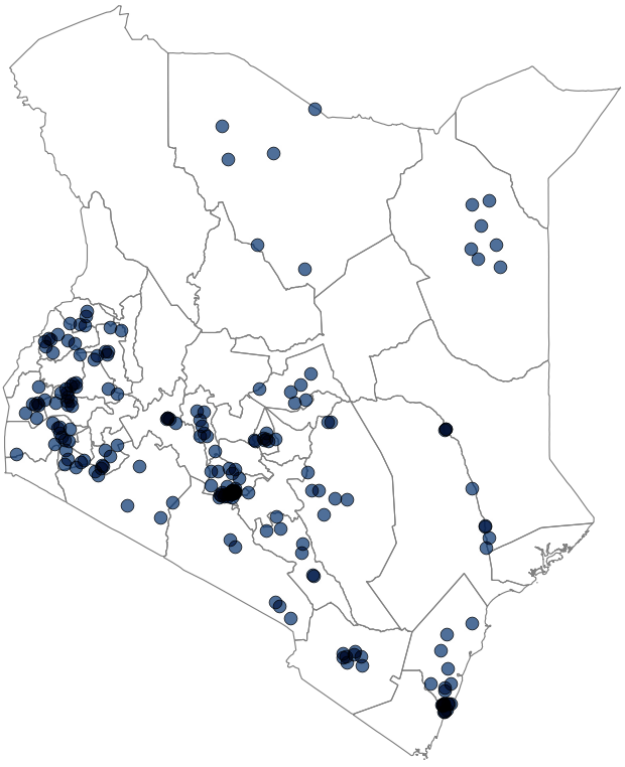
comprehension). Kiswahili had four pre-reading subtasks (letter sound knowledge, syllable fluency, invented/non-word decoding and listening comprehension) and two reading subtasks (passage reading and reading comprehension). Some of the subtasks were untimed and others were timed. For the untimed tasks, the pupils were presented with a series of items, for example, identifying vocabulary words or answering comprehension questions. For the timed tasks, the pupils were given one minute to perform a subtask, for example, naming letter sounds or orally reading a passage.

Table 4: EGRA: Number of Items per Subtask

Subtask	Timed?	English	Kiswahili
Phoneme segmentation		10	
Letter sound knowledge	🕒	100	100
Syllable fluency	🕒		100
Invented/non-word decoding	🕒	50	50
Vocabulary		20	
Passage reading (A)	🕒	70	68
Reading comprehension		6	6
Passage reading (B)	🕒	70	
Reading comprehension (B)		6	
Listening comprehension			5

A description of the EGRA subtasks and copies of all other data collection tools are in Annex III: Data Collection Tools.

Figure 6: Map of Sampled Schools



SAMPLING

Sample Design

For the midline, the evaluation team assessed pupils from the same 204 schools sampled for the baseline. Through discussions with USAID, MOE and RTI, the evaluation team created the sampling frameworks and set up the design for a national sample in 2015. Using a three-stage cluster sampling procedure from a sampling frame of 22,154 formal public schools and 1,000 non-formal (or Alternative Provision of Basic Education and Training – APBET) schools, the evaluation team drew a random sample as described below, resulting in a target of 4,896 total pupils comprising 2,448 boys and 2,448 girls divided equally between the two classes:

1. 26 (of 47) counties covering all eight former provinces;
2. 204 schools comprising 174 public schools and 30 APBET institutions; and
3. 24 pupils per school, with 12 (six boys and six girls) each in Classes (Standards) 1 and 2.

APBET institutions are regulated by the MOE under the Education Standards and Quality Assurance (ESQAC) Basic Education Act and use the Kenya Institute for Curriculum Development's (KICD) syllabi. The MOE has stipulated a minimum of 30 percent of the teachers at APBET institutions having a relevant teacher-training certificate from a recognized institution. Note that at the time of the midline data collection, Tusome supported APBET institutions located in informal settlements of the urban areas of Nairobi, Kisumu, and Mombasa, and started supporting APBET institutions in Eldoret in January 2017. As such, their characteristics may be inherently different than the public schools, which are in both urban and rural locations.

For each sample school, the evaluation team aimed to interview the Class 1 and Class 2 teacher, the head teacher and the associated CSO.

For the household survey, MSI collected contact information from the head teachers of the pupils during the assessment data collection. The sampling frame included all pupils' households.

For additional details on the sampling design, actual sample and weighting, see Annex II: Sampling.

EVALUATION STRENGTHS AND LIMITATIONS

Strengths

The data collection tools and the analyses were sufficient for answering the midline evaluation questions. The tools covered a variety of aspects of the Tusome reading interventions by collecting data from the pupils, teachers, head teachers, communities, CSOs, and education officials.

Limitations and Their Mitigation

This evaluation was designed to collect data from the same schools as the baseline to document change over time. While MSI collected data from all 204 schools in the sample, technical difficulties resulted in data from one school being lost. The evaluation team was able to include more than 99 percent of schools in the midline database.

A few limitations to the survey data should be taken into consideration when interpreting the pupil, teacher, head teacher and CSO survey results.

- Some pupil groups and most teacher and head teacher groups had small sample sizes, so any conclusions for those groups should be made with a high degree of caution.
- Confounding may lead to inaccurate interpretations, for example, a group of teachers with higher pupil scores may teach pupils in urban areas who generally have higher scores.
- Some of the group percentages do not sum to 100 percent, for example, home language percentages were based on responses to questions about individual languages ("Do you speak Kiswahili at home?").
- "Don't know" or "Other" response categories often included invalid responses, so scores were not reported for them, for example, "Q: Do you practice silent reading at school?" "A: I read at home."
- Inconsistencies emerged when different kinds of respondents answered similar questions, for example, teachers and head teachers responded differently when asked if schools had libraries, and socioeconomic measures varied between pupil and parent responses.

There are also some limitations with the household survey data, which were collected over the phone from Nairobi. While MSI made every effort through careful planning to maximize the survey of the pupils' households, the evaluation reached 49 percent of households given its time, budget and logistical constraints. Based on the team's experience with other similar surveys, this was actually a high

percentage of households reached in a cost-effective manner. See Annex II: Sampling for further discussion about differences between the full pupil population and household survey respondents.

Finally, the timeline of the study should be noted as USAID recommends collecting the baseline data before the start of interventions and then the midline (and endline) data at the same time point as the baseline data during the subsequent school year(s). An original baseline was conducted in 2014. Due to issues with data quality, the baseline was redone in 2015. The baseline data were collected in July 2015, shortly after Tusome started working with Class 1 teachers, but before Tusome started working with Class 2. The goal of the timing for the revised baseline was to limit the possible effect of the interventions on the baseline scores, but also capture as much learning during the school year as possible. The data collection in the second half of July measured pupils' ability levels about three-quarters of the way into the instructional part of the school year.

For the midline, data were collected more towards the end of the school year in September and October 2017, that is at a later point in the school year than the baseline data collection. However, there was only slightly more time for learning due to the school break in August and the beginning of September. After consultations with USAID, the evaluation team did not find adequate reason for adjusting pupil scores due to the later time point of the midline data collection (or the small amount of learning that may have taken place by Class 1 pupils at the time of the baseline).

EGRA TEST VALIDITY AND RELIABILITY

Test Validity

Validity was assured through the test development process that involved close collaboration between the MOE and the evaluation team. The model test selection, a test development workshop, pilot testing, test revision and a test validation workshop with the MOE were critical to establishing test validity. The process successfully created a version of EGRA that measured reading skills in English and Kiswahili for the Kenyan context. The test also complied with USAID requirements for collecting data that would allow for measuring progress toward its global Goal 1 indicators.

Test Reliability

The main indicator of reliability for psychometric tests is Cronbach's alpha, which estimates the internal consistency reliability of a test for a particular test administration. It indicates the extent to which subtasks or items that are designed to measure a particular construct are able to deliver consistent scores. The range for Cronbach's alpha is 0.00 to 1.00, with higher values indicating better (or more desirable) reliability. Values of 0.80 and above are considered acceptable. The evaluators calculated the alphas separately for each grade level and language for both baseline and midline data collection rounds using the percentage of correct scores for the subtasks. Table 3 shows the results.

Table 5: Test Reliabilities by Grade Level

Language	Number of Subtasks	Class 1		Class 2	
		Baseline	Midline	Baseline	Midline
English	8	0.92	0.91	0.92	0.90
Kiswahili	6	0.89	0.89	0.90	0.88

For English, the values ranged from 0.90 to 0.92. For Kiswahili, the values ranged from 0.88 to 0.90. These values indicate strong reliability for each of the languages, grade levels and data collection rounds,

especially considering that reliability estimates are generally lower when the number of subtasks is smaller, such as with the eight English and six Kiswahili subtasks on this version of EGRA.

Inter-Rater Reliability

For the inter-rater reliabilities (IRRs), the team conducted a sample-based study during data collection to report how consistently the assessors rated the students' performance on EGRA and how much variation occurred in their scores due to lack of consistency of the assessors' ratings. A higher IRR estimate for a study would indicate greater confidence in the data. An IRR estimate of 0.60 is considered "good" and 0.75 and above is considered "excellent."

For the midline, a subset of 397 students (8.2 percent of the sample) was included in the IRR study. Two enumerators assessed each student in the subset. The team calculated two IRR statistics – Kappa and Intra-Class Correlation (ICC) – for the test and for each subtask. For the test, the Kappa estimate was 0.80 and the ICC estimate was 0.79, or "excellent." For the subtasks, the Kappa estimates ranged between 0.61 and 0.98, which indicate "good" to "excellent"; seven subtask estimates were "good" and seven were "excellent."

Subtask Quality and Reliability

At the subtask level, the team calculated two statistics: 1) subtask-total correlations for the quality (or discrimination) of the subtasks and 2) Cronbach's alpha for the reliability of the untimed subtasks.

The subtask-total correlation provides an indication of whether the subtask is able to discriminate between high- and low-achieving pupils. For each language, these were calculated by correlating the percentage of correct scores for each subtask and the grand mean for all subtasks (total score). Subtasks are considered to have acceptable quality if this correlation is 0.20 or above.

Cronbach's alpha for the subtasks is similar to the alpha for the test, except that the subtask is treated as a testlet. In other words, it is calculated using the items within the subtask as opposed to the subtasks within the test. For instance, with phoneme segmentation, the evaluators calculate the alpha using the percentage of correct scores for each item and the percentage of correct scores for the subtask. Since these are subtasks instead of tests, values of 0.70 and above are considered acceptable in this type of subtask analysis. Note that the coefficients were calculated only for the untimed tasks, since the similarity of the items on the timed tasks will always lead to high alphas.

Subtask-total correlations and the alpha coefficients were calculated separately for each grade level and language at the midline. For English (Table 4), all subtask-total correlations were well above the minimum standard, indicating high-quality subtasks. All alpha coefficients (for the untimed subtasks only) were above 0.70, indicating strong internal consistency reliability at the subtask level.

Table 6: English Subtask-Total Correlations and Alpha Coefficients

Subtask	Class 1		Class 2	
	Subtask-Total	Alpha Coefficient	Subtask-Total	Alpha Coefficient
1. Phoneme segmentation	0.61	0.94	0.63	0.93
2. Letter sound knowledge	0.66	--	0.56	--
3. Invented/non-word decoding	0.87	--	0.83	--
4. Vocabulary	0.78	0.88	0.73	0.87
5a. Passage reading (A)	0.93	--	0.91	--

Subtask	Class 1		Class 2	
	Subtask-Total	Alpha Coefficient	Subtask-Total	Alpha Coefficient
5b. Reading comprehension (A)	0.73	0.79	0.74	0.82
6a. Passage reading (B)	0.93	--	0.91	--
6b. Reading comprehension (B)	0.79	0.87	0.78	0.88

For Kiswahili (Table 5), the subtasks were also of high quality, with subtask-total correlations well above 0.20 for all six subtasks. The alphas (again, for the untimed subtasks only) were above 0.70, indicating good internal consistency reliability.

Table 7: Kiswahili Subtask-Total Correlations and Alpha Coefficients

Subtask	Class 1		Class 2	
	Subtask-Total	Alpha Coefficient	Subtask-Total	Alpha Coefficient
1. Letter sound knowledge	0.80	--	0.72	--
2. Syllable fluency	0.89	--	0.86	--
3. Invented/non-word decoding	0.82	--	0.80	--
4a. Passage reading	0.89	--	0.88	--
4b. Reading comprehension	0.86	0.77	0.84	0.79
5. Listening comprehension	0.55	0.73	0.52	0.71

See Annex VI: Psychometric Analyses for more information on the correlations and item quality and reliability.

DATA COLLECTION

MSI information technology (IT) specialists adapted an electronic data collection application that they had developed for previous USAID-funded projects. The evaluation team, including the quality control officers (QCOs), piloted the application and the IT specialists conducted quality control prior to data collection. MSI selected a local subcontractor, Research Solutions Africa (RSA), to administer the tests and surveys. Data collection was split into two cohorts to allow for smaller training groups.

The evaluation team provided extensive training to the RSA leadership team, supervisors and enumerators so that administration of the tests and surveys would adhere to international standards of quality. This training took place in two five-day workshops prior to data collection. It included scripted practice, during which the evaluation team provided detailed training, checked the enumerators' inter-rater reliability (IRR) and retrained enumerators whose ratings did not agree with the gold standards. In general, retraining was minor; nearly all of the QCOs, supervisors and enumerators had previously participated in training with IRR-type agreement analysis from MSI during the baseline or with RTI during the PRIMR EGRA data collections. The training included two practice days in Nairobi schools.

A total of 24 QCOs, 24 supervisors and 96 enumerators working in 24 teams were selected to perform data collection in the schools. The first 12 teams were trained for one week from September 26–30 and collected data during the three weeks from October 3–20. The other 12 teams were trained for one week from October 3–7 and collected data during the two weeks from October 10–20.

DATA ANALYSIS

An MSI statistician cleaned and analyzed the data using Stata statistical software, with quality assurance by an MSI psychometrician. The team created Excel tables to prepare this technical report. The statistician, psychometrician and the evaluation team leader reviewed each table for data quality and consistency.

For the sampling weights, the calculations of the final weights were based on the inverse of the overall probabilities of selection. The calculations took into consideration the stratification (counties, public/APBET) and the number of students per school.

The EGRA findings were validated during a series of workshops in March 2017 between the evaluation team, USAID, MOE, and RTI. These validation workshops provided valuable insight into the findings and helped inform the evaluation team's conclusions and recommendations.

Upon approval of this report, MSI will submit the full datasets to USAID/KEA electronically. MSI will also submit all required reports and the public-use file (cleaned, finalized and de-identified dataset) to the Secondary Analysis for Results Tracking (SART) portal.

KEY FINDINGS

For findings related to EGRA, all data are disaggregated by grade level and language. For each language, the findings for Classes 1 and 2 are presented together, either in the same table or in adjacent tables, to compare the levels of pupils by grade. The results by language are presented in separate tables since they should not be compared; English and Kiswahili have different structures, so pupils might learn at different rates even with the same level of instruction. All results are presented at the national level and should not be disaggregated to counties, as the sample sizes are not large enough for those comparisons to be valid.

The EGRA results for passage reading and comprehension are presented under evaluation questions 1 and 2 respectively. Results for the other EGRA tasks are included in Annex V: EGRA Results.

Some of the data were also disaggregated by other variables. In particular, the EGRA data were disaggregated by school type (public and APBET) and gender (male and female). The teacher and head teacher data were disaggregated by demographics (gender, qualifications, years of experience, etc.) and by survey variables (instructional methods, facilities, etc.). All results were analyzed using descriptive statistics (frequencies, percentages, raw score means, etc.). For the pupils, inferential statistics (t-tests) were used to compare results on the group variables, with the significance level set at $p < .01$ based on the level used in the power calculations for the sampling. Statistically significant findings were indicated with an asterisk next to the mean score of the higher-performing group. Inferential statistical tests (t-tests and ANOVAs) on the teacher, head teacher, CSO, classroom, and household data are not reported due to small sample sizes.

EVALUATION QUESTION 1

What proportion of students can demonstrate they can read grade-level text (within Kenya's curricular goals) by the end of Standards 1 and 2?

KEY FINDINGS

The Class 2 benchmarks for reading performance in English are 30 to 64 correct words per minute (CWPM) for emergent readers and 65 or more CWPM for fluent readers.

- At midline, 30 percent of Class 1 pupils are emergent readers and 18 percent are fluent readers in English.
- At midline, 29 percent of Class 2 pupils are emergent readers and 47 percent are fluent readers in English.

The Class 2 benchmarks for reading performance in Kiswahili are 17 to 44 CWPM for emergent readers and 45 or more CWPM for fluent readers.

- At midline, 32 percent of Class 1 pupils are emergent readers and 3 percent are fluent readers in Kiswahili.
- At midline, 54 percent of Class 2 pupils are emergent readers and 12 percent are fluent readers in Kiswahili.

In 2012, the RTI PRIMR team collaborated with the MOE and the Kenya National Examinations Council (KNEC) in setting Class 2 reading benchmarks. After discussing and analyzing options, they established draft benchmarks for oral reading fluency (ORF) in English and Kiswahili, expressed in correct words per minute (CWPM).[4]

As Table 6 shows, the benchmarks were set with three cut-scores (beginning, emergent and fluent), which were then used for placing each pupil's performance into one of four reading categories (zero, beginning, emergent, and fluent readers). The fluency benchmarks helped determine whether pupils were reading at grade level, that is whether they could read grade-level text with proficiency. The English fluency benchmark was set at 65 CWPM and Kiswahili fluency at 45 CWPM for Class 2 pupils. The reason cited for this difference was that Kiswahili is an agglutinative language. While there are no corresponding benchmarks for Class 1 pupils, the evaluation team used the same benchmarks to analyze the results of both classes in order to show improvements over time.

Table 8: ORF Performance Categories for English and Kiswahili

Category	English CWPM	Kiswahili CWPM
Fluent reader	65+	45+
Emergent reader	30-64	17-44
Beginning reader	1-29	1-16
Zero reader	0	0

English

The percentages of pupil scores by performance category were based on the ORF scores from English passage A (with the standard EGRA administration). Table 7 shows that performance improved between baseline and midline, with a lower percentage of pupils in the zero and beginning reader categories and a higher percentage in the emergent and fluent reader categories.

The proportion of non-readers (or zero readers) decreased substantially from baseline to midline. In Class 1, 53 percent of the pupils cannot read a single word correctly at baseline, which decreased by over half to 23 percent at midline. In Class 2, the percentage of non-readers decreased by over two thirds from 38 percent at baseline to 12 percent at midline.

4 RTI International (2014). USAID/Kenya primary math and reading initiative: Final report.

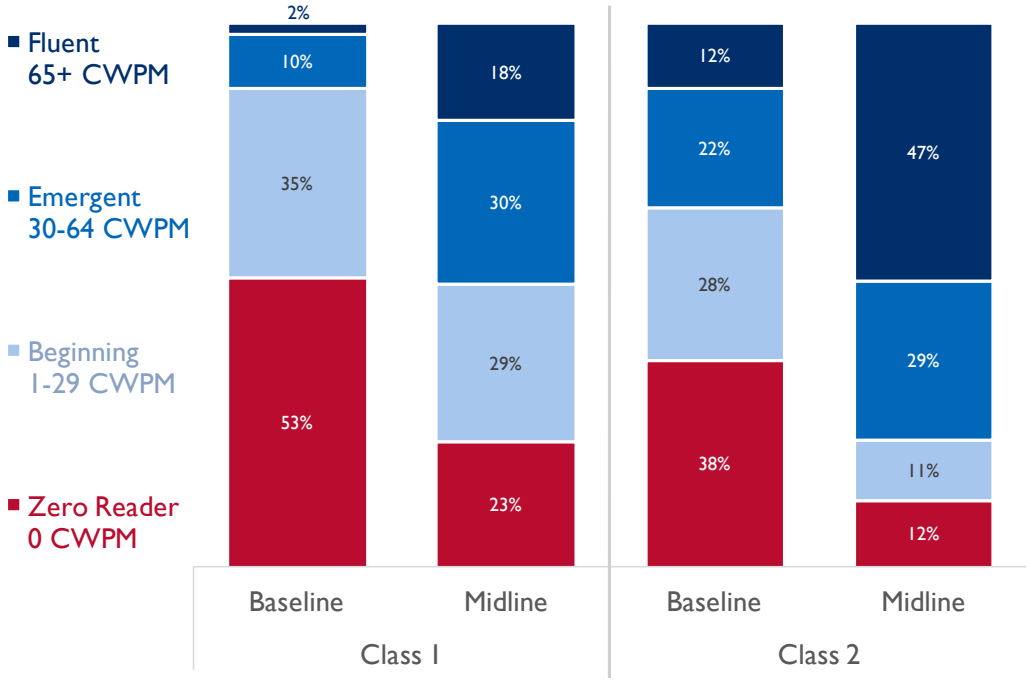
The percentage of emergent and fluent readers increased between baseline and midline. For Class 1, fluent went from 2 percent at baseline to 18 percent at midline. For Class 2, it rose from 12 percent to 48 percent. Similarly, for Class 1, emergent readers increased from 10 percent to 30 percent. For Class 2, emergent readers increased from 22 percent to 30 percent.

Table 9: English Oral Reading Fluency Performance Categories

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Fluent reader	2.2%	17.9%	15.7%	11.6%	47.5%	35.9%
Emergent reader	9.8%	30.2%	20.4%	22.2%	29.4%	7.2%
Beginning reader	35.2%	29.0%	-6.2%	28.2%	11.3%	-16.9%
Zero reader	52.8%	22.9%	-29.9%	37.9%	11.8%	-26.1%

Figure 7 shows bar graphs with Classes 1 and 2 at baseline and midline. Clearly, the percentages of scores in the emergent and fluent reader categories increased substantially from baseline to midline.

Figure 7: English Reading Performance Categories, Baseline and Midline



The evaluation team also analyzed the performance categories by disaggregating the results by school type and gender. In general, girls scored slightly higher than boys and APBET institutions scored higher than public schools.

By school type, pupils in the APBET institutions had substantially fewer scores in the lower categories and more scores in the upper categories than the public schools did at midline. The percentages of pupils with zero scores in the APBET institutions were about 2 percent in Class 1 – compared to 23 percent in the public schools – and about 3 percent in Class 2 – compared to 12 percent in the public schools. Similarly, about 65 percent of the Class 1 scores in the APBET institutions were in the

fluent category – compared to 17 percent in public schools – and about 86 percent of the Class 2 scores were in the fluent category – compared to 47 percent in public schools (Table 8).

Table 10: English Midline ORF Scores by Performance Category, Class and School Type

Class	School Type	Zero	Beginning	Emergent	Fluent
1	Public	23.1%	29.3%	30.3%	17.3%
	APBET	1.7%	5.7%	27.3%	65.3%
2	Public	11.9%	11.4%	29.7%	47.0%
	APBET	2.7%	2.7%	8.9%	85.7%

By gender, the female pupils had somewhat fewer scores in the lower categories and more scores in the upper categories than the male pupils did at midline. The percentages of pupils with zero scores were about 10 percent lower for females than males in Class 1 and about 6 percent lower for females than males in Class 2. Similarly, the percentages of pupils with fluent scores were about 2 percent higher for females than males in Class 1 and about 7 percent higher for females than males in Class 2 (Table 9).

Table 11: English Midline ORF Scores by Performance Category, Class and Gender

Class	Gender	Zero	Beginning	Emergent	Fluent
1	Male	28.0%	28.9%	26.6%	16.5%
	Female	17.7%	29.2%	33.9%	19.3%
2	Male	14.5%	11.1%	30.3%	44.1%
	Female	9.2%	11.5%	28.5%	50.9%

Kiswahili

The percentages of pupil scores by performance category were based on the ORF scores from the Kiswahili reading passage. Table 10 shows that performance in Kiswahili ORF has improved between baseline and midline but is in general lower than in English.

In Class 1, 70 percent of the pupils count not read a single word at baseline. This decreased to 45 percent at midline. In Class 2, the percentage of non-readers decreased by over half, from 43 percent to 19 percent, at midline.

As in English, the percentages of emergent and fluent readers also increased in Kiswahili. About 3 percent of the Class 1 pupils and 12 percent of the Class 2 pupils demonstrated fluency at midline, an increase from 1 percent and 4 percent at baseline respectively. For Class 1, emergent readers increased from 12 percent to 32 percent. For Class 2, emergent readers increased from 33 percent to 54 percent.

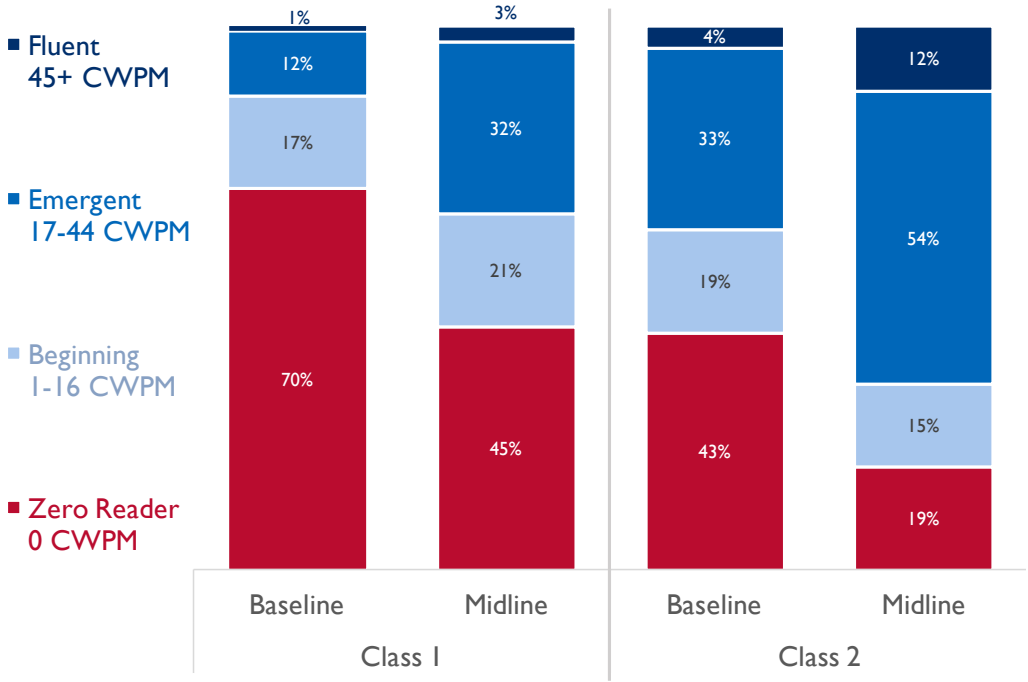
Table 12: Kiswahili Oral Reading Fluency Performance Categories

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Fluent reader	0.7%	2.6%	1.9%	4.3%	11.8%	7.5%
Emergent reader	12.0%	31.7%	19.7%	33.3%	54.4%	21.0%
Beginning reader	17.5%	21.0%	3.6%	19.0%	14.9%	-4.2%

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Zero reader	69.9%	44.7%	-25.2%	43.3%	18.9%	-24.4%

Figure 8 shows the percentages of Kiswahili scores in the different categories in Classes 1 and 2 at baseline and midline, with substantial increases in the emergent and fluent categories and decreases in the zero reader categories.

Figure 8: Kiswahili Reading Performance Categories, Baseline and Midline



As with English, the evaluation team conducted further analyses of the performance categories in Kiswahili by disaggregating the results by school type and gender. In general, these analyses were consistent with the trends from the earlier analyses of average scores for the groups.

By school type, the pupils in the APBET institutions had substantially fewer scores in the lower categories and more scores in the upper categories than the public schools did. The percentages of pupils with zero scores in the APBET institutions were about 10 percent in Class 1 – compared to 45 percent in the public schools – and about 4 percent in Class 2 – compared to 19 percent in the public schools. Similarly, about 15 percent of the Class 1 scores in the APBET institutions were in the fluent category – compared to 2 percent in the public schools – and 35 percent of the Class 2 scores were in the fluent category – compared to 12 percent in the public schools (Table 11).

Table 13: Kiswahili Midline ORF Scores by Performance Category, Class and School Type

Class	School Type	Zero	Beginning	Emergent	Fluent
1	Public	45.2%	21.1%	31.4%	2.4%
	APBET	9.9%	19.6%	55.4%	15.1%

Class	School Type	Zero	Beginning	Emergent	Fluent
2	Public	19.1%	15.0%	54.3%	11.5%
	APBET	3.9%	4.4%	57.0%	34.7%

By gender, female pupils had slightly fewer scores in the lower categories and more scores in the upper categories than the male pupils did. The percentages of pupils with zero scores were about 6 percent lower for females than males in Class 1 and about 2 percent lower for females than males in Class 2. Similarly, the percentages of pupils with fluent scores were about 1 percent higher for females than males in Class 1 and about 7 percent higher for females than males in Class 2 (Table 12).

Table 14: English Midline ORF Scores by Performance Category, Class, and Gender

Class	Gender	Zero	Beginning	Emergent	Fluent
1	Male	47.5%	21.9%	28.5%	2.1%
	Female	41.8%	20.2%	34.9%	3.1%
2	Male	20.0%	15.6%	54.6%	9.8%
	Female	17.9%	14.1%	54.2%	13.9%

EVALUATION QUESTION 2

What are the levels of Classes 1 and 2 pupils on reading subtasks?

KEY FINDINGS

- Pupils have shown improvements on all subtasks in both languages and classes.
- In English, Class 1 pupils performed best on the vocabulary and phoneme segmentation, followed by the passage reading subtasks. Class 2 pupils performed best on passage reading, followed by the vocabulary and phoneme segmentation subtasks.
- In Kiswahili, both classes performed best on listening comprehension.
- In both languages, scores for reading comprehension – the most difficult subtask – has improved but remained the lowest in terms of overall performance.
- When given the opportunity to read aloud, read silently and retain the text, pupils can answer more comprehension questions than when only given the opportunity to read aloud.

The EGRA consisted of fourteen subtasks designed to assess the pupils' reading skills from phonemic awareness to reading comprehension as described above in Data Collection Tools.

To answer evaluation question two, the evaluation team looked at both the raw scores and the percent correct scores. For raw scores, the untimed tasks are reported in terms the number of items that the pupil got correct. The scores for the timed tasks are reported in terms of adjusted raw scores; these scores were adjusted upwards if the pupil completed the task prior to the end of one minute. For raw scores, an asterisk denotes that the differences are statistically significant at the $p<.01$ level. Percent correct scores divide by the number of items answered correctly for each subtask by the total number of items. This statistic allows comparisons on performance between subtasks.

Disaggregated results by school type (public and APBET) and gender (male and female) are included in Annex V: EGRA Results. In general, girls' scores were slightly higher than average and boys' scores were slightly lower than average, though these differences were minimal. Scores for APBET institutions are higher than public schools, though gains between baseline and midline are similar or higher for public schools.

English

Pupils have shown improvement on all English subtasks in both classes between the baseline and midline. The raw scores are shown in Table 13 below. All gains between baseline and midline are statistically significant at the 0.01 level.

Table 15: English Raw Reading Scores

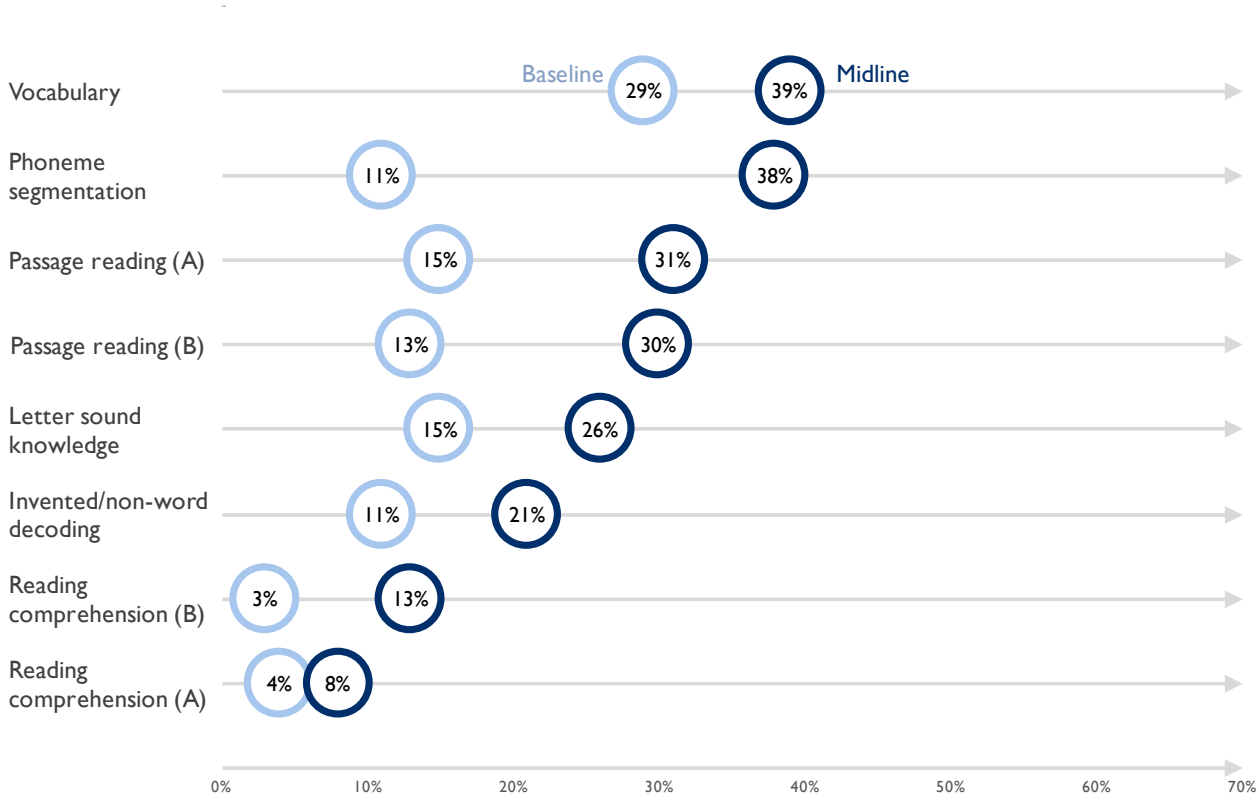
Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Phoneme segmentation	1.1	3.8	2.6*	0.6	5.0	4.5*
Letter sound knowledge	15.1	26.3	11.3*	10.2	32.6	22.4*
Invented/non-word decoding	5.7	10.4	4.7*	10.4	18.6	8.3*
Vocabulary	5.9	7.8	1.9*	8.2	10.2	1.9*
Passage reading (A)	10.6	22.3	11.7*	23.8	43.6	19.9*
Reading comprehension (A)	0.2	0.5	0.3*	0.5	1.0	0.5*
Passage reading (B)	9.7	22.0	12.4*	21.8	44.2	22.5*
Reading comprehension (B)	0.2	0.8	0.6*	0.6	1.7	1.2*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

To analyze performance across subtasks, the evaluation team also looked at the percent correct scores. As shown in Figure 9, Class 1 the pupils performed best on the vocabulary and phoneme segmentation subtasks, with an average of 39 and 38 percent correct at midline respectively. This was followed by the two passage reading subtasks, with an average score of 30 and 31 percent correct. As with baseline, the lowest scores were in reading comprehension, though Class 1 pupils have shown improvement in this area. Of particular note is the difference between the two types of reading comprehension subtasks, which is discussed further below.

Class 1 pupils saw the most improvement in phoneme segmentation, increasing scores by 27 percentage points from 11 percent at baseline to 38 percent at midline. Passage reading also showed strong gains, with a 16 and 17 percent gain for passages B and A respectively.

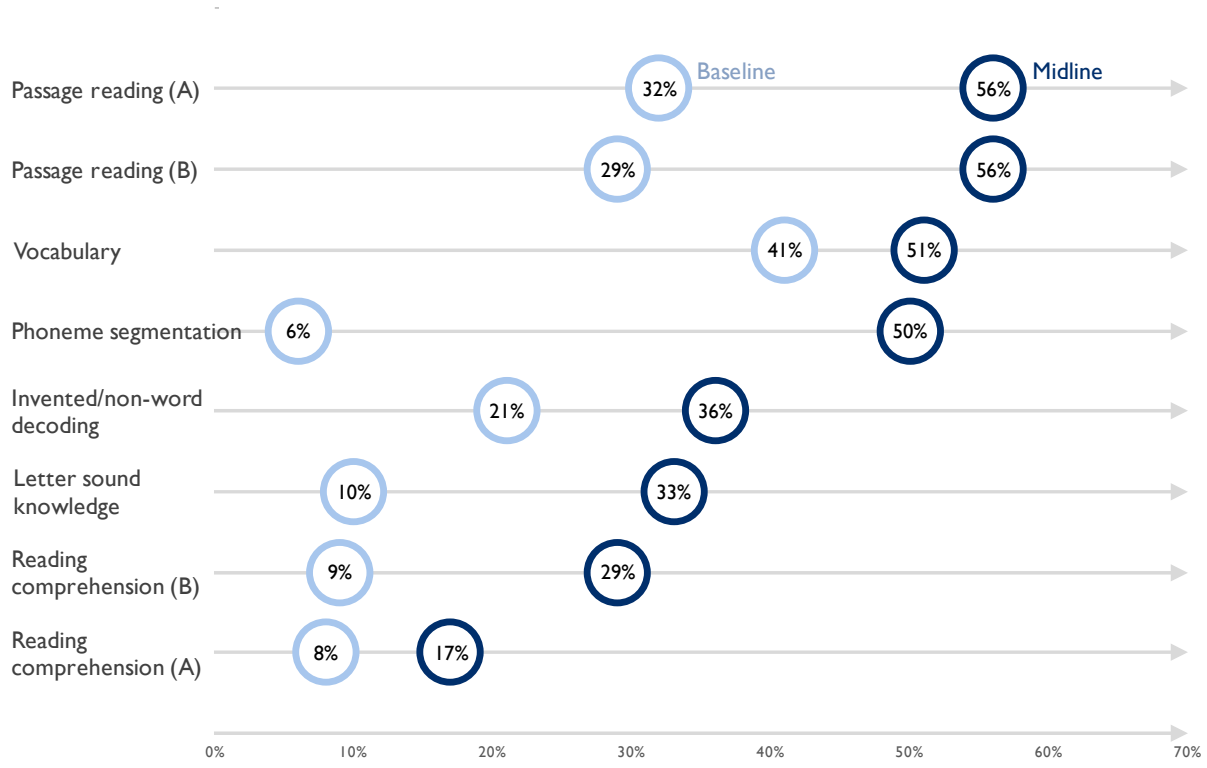
Figure 9: English Class 1 Percent Correct Reading Scores



As Figure 10 shows, Class 2 pupils showed the strongest performance in the passage reading subtasks, at 56 percent correct at midline. As with Class 1, they also showed strong performance on the vocabulary and phoneme segmentation subtasks with 51 and 50 percent correct respectively. As with Class 1, reading comprehension scores are lower than other tasks but showed strong improvement.

Class 2 pupils demonstrated an impressive 44 percentage point improvement in phoneme segmentation from 6 percent correct at baseline to 50 percent correct. They also showed strong improvement in the two passage reading subtasks of 24 and 27 points and letter sound knowledge of 23 points. Class 2 pupils also demonstrated a 20 point improvement in the reading Comprehension B subtask as discussed below.

Figure 10: English Class 2 Percent Correct Reading Scores

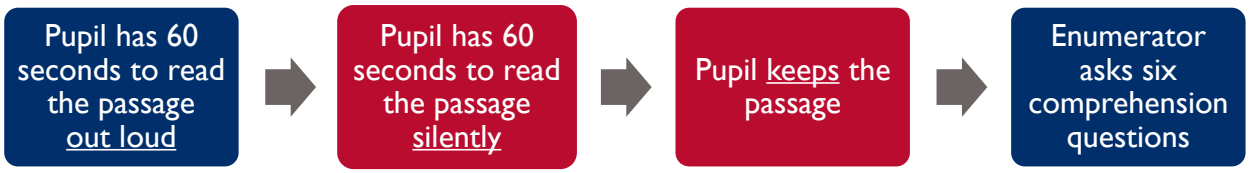


Of particular note in this study, the EGRA included two types of reading comprehension in English. Reading Comprehension A is the standard EGRA subtask, as Figure 11 shows.

Figure 11: Reading Comprehension A

Reading Comprehension B is a custom subtask type developed at the request of the MOE. This subtask incorporates silent reading, a skill taught under the Tusome methodology. Figure 12 shows that it differs from the standard subtask in two ways. First, the pupil is given 60 seconds to re-read the passage silently after reading it orally. Second, the pupil retains the passage during the comprehension questions and may go back to the text as a reference.

Figure 12: Reading Comprehension B



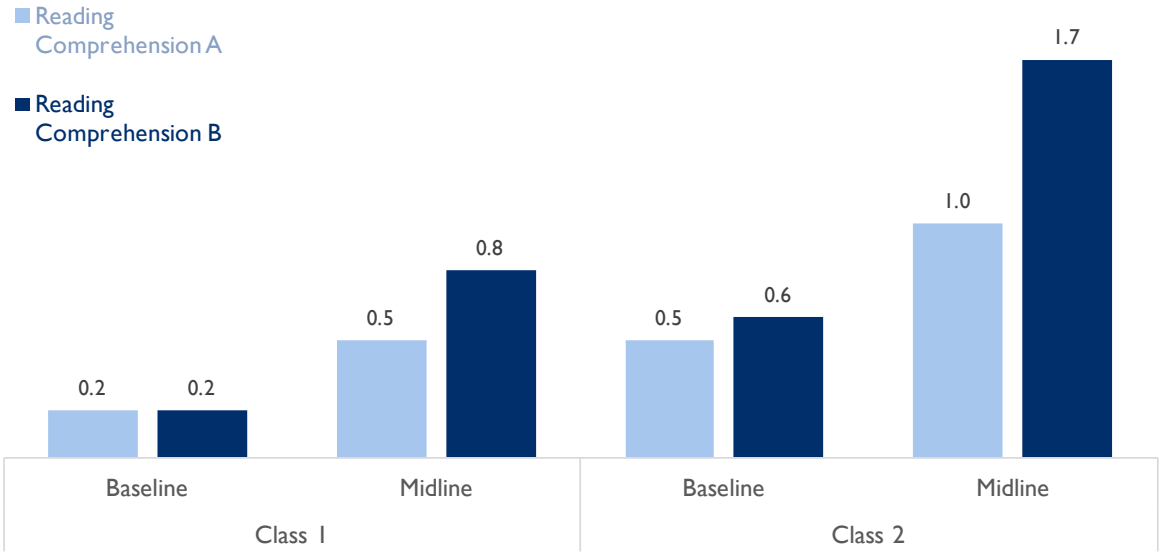
The difference between the two comprehension tasks at baseline and midline are summarized in Table 14 as the average number of correct responses out of six comprehension questions. The average comprehension score for English started at 0.2 for Class 1 and at 0.5 to 0.6 for Class 2 at baseline for both types of reading comprehension subtasks. For Comprehension A, these scores more than doubled to 0.5 correct responses for Class 1 and 1.0 for Class 2. For Comprehension B (with silent reading), the improvement was more pronounced, with an average score of 0.8 correct responses for Class 1 and 1.7 for Class 2.

Table 16: English Reading Comprehension Raw Scores

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Reading Comprehension A	0.2	0.5	0.3	0.5	1.0	0.5
Reading Comprehension B	0.2	0.8	0.6	0.6	1.7	1.1

Figure 13 illustrates that while the average raw score for the two English comprehension tasks were similar at baseline, pupils showed stronger improvement for reading Comprehension B (with silent reading) than A (without silent reading) at midline.

Figure 13: English Reading Comprehension Raw Scores



Kiswahili

Pupils have shown improvement on all Kiswahili subtasks in both classes between the baseline and midline. The raw scores are shown in Table 15. All gains between baseline and midline are statistically significant at the 0.01 level.

Table 17: Kiswahili Reading Scores

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Letter sound knowledge	16.6	29.7	13.1*	16.2	39.7	23.4*
Syllable fluency	11.0	21.5	10.4*	20.9	37.5	16.6*
Invented/non-word decoding	4.7	8.3	3.6*	10.2	16.1	5.8*
Passage reading	4.9	12.2	7.3*	13.5	24.5	11.0*
Reading comprehension	0.4	0.9	0.5*	1.1	2.0	1.0*
Listening comprehension	1.2	2.0	0.8*	1.9	2.0	0.9*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

As with English, the evaluation team used the percent correct scores to analyze performance across the Kiswahili subtasks. The gains shown between baseline and midline for Class 1 and Class 2 are illustrated in Figures 14 and 15 respectively.

In both classes, the pupils' best performance was on listening comprehension, followed by letter sound knowledge. The largest improvement was in letter sound knowledge for Class 2 pupils, increasing from an average of 16 percent correct at baseline to 40 percent correct at midline. Large gains were also seen in listening comprehension which rose from 25 to 40 percent correct in Class 1 and from 38 to 55 percent correct in Class 2.

As at baseline, reading comprehension remains among the lowest for both classes in terms of performance but scores have shown improvement between baseline and midline. This improvement is particularly pronounced in Class 2, which went from an average reading comprehension score of 18 percent correct at baseline to 34 percent correct at midline.

Figure 14: Kiswahili Class 1 Percent Reading Scores

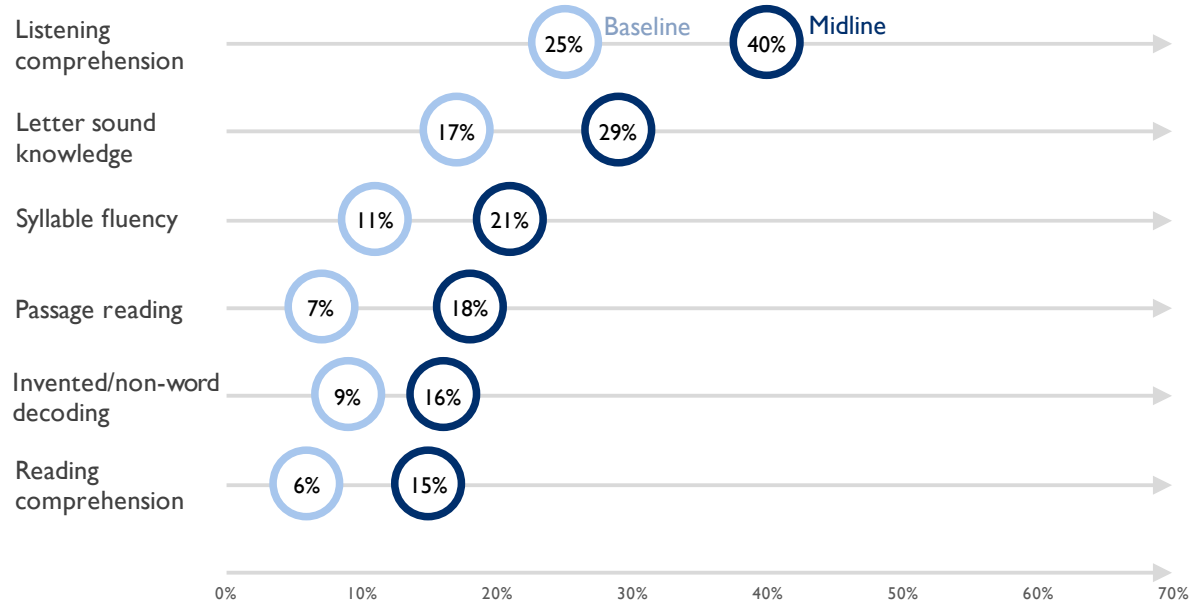
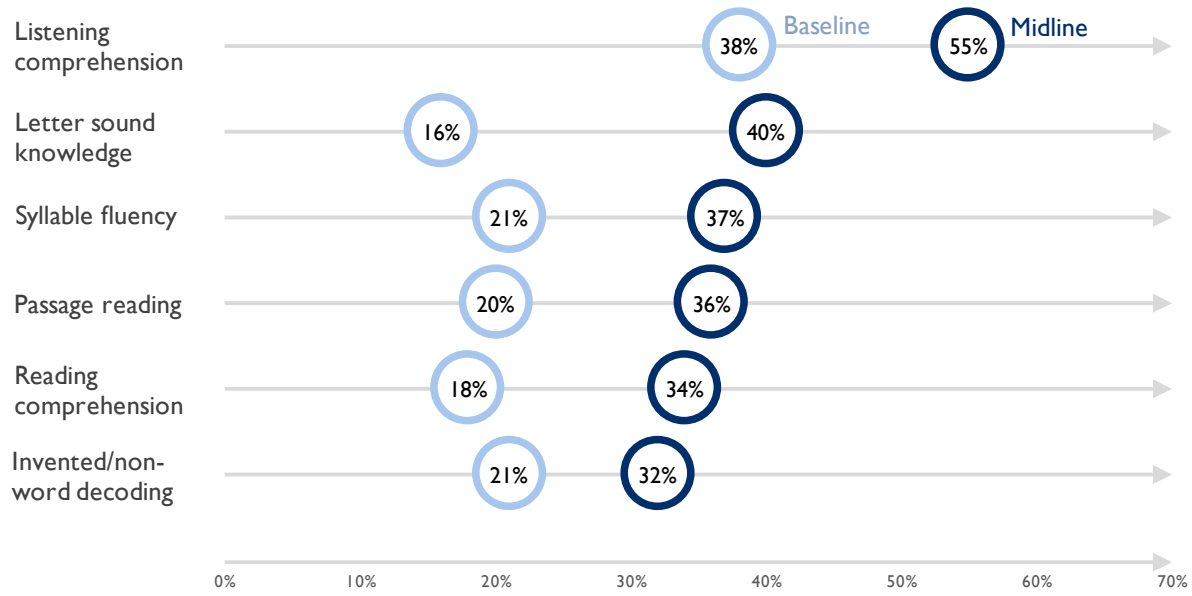


Figure 15: Kiswahili Class 2 Percent Reading Scores



EVALUATION QUESTION 3

What school-level and institutional factors influence reading outcomes when implementing at scale, and how?

KEY FINDINGS

Several school-level and institutional factors were associated with better reading outcomes (ORF), including:

- Access to reading materials at home and in the school;
- Access to a reading teacher's guide;
- Practice reading aloud and silently at school;
- Increased frequency of CSO observations;
- Increased frequency of lesson plans being reviewed;
- Full-day school shifts (in comparison to half-day);
- Classroom libraries;
- Pupils of the correct age range (5–9 years old); and
- Female teacher or head teacher.

The following tables provide information on pupil, teacher, head teacher, and school variables and their associations with reading outcomes (or ORF).[5] The tables show the percentages for each category and the average ORF scores for those categories. The data presented are not comprehensive, but selected

5 All ORF scores for English are from the Passage Reading A task.

for the purposes of this report. Even though all of the data are not presented, some overlaps exist in the pupil, teacher and head teacher indicators. As noted under Strengths and Limitations, this survey data has limitations, but provides valuable contextual information.

Pupils

Findings from the pupil questionnaire are grouped into three sections: pupil characteristics; pupil reading and materials; and pupil-school characteristics.

Pupil Characteristics

As Table 18 shows, pupils with certain characteristics had significantly higher scores. Generally, the differences in the groups and the sample sizes needed to be large enough in order to show significance. Pupils in the appropriate age range (ages 5–9) for both grade levels had higher ORF scores than either the underage or over-age pupils in English in both classes and in Kiswahili in Class 2. Those with Kiswahili as their main home language had higher ORF scores than pupils with "other" home languages in English in Class 1 and in both languages in Class 2. Pupils with English as their main school language had higher ORF scores in both classes. Pupils who attended pre-school had higher ORF scores than those who did not (question was asked only of the Class 1 pupils) in English and Kiswahili.

Table 18: Pupil Characteristics

Characteristic	Group	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Age (in years)	Below 5	0.1%	15.7	8.7	0.2%	43.0	26
	5-9	90.1%	22.9*	12.6	76.2%	46.5*	25.9*
	Above 9	9.8%	16.9	9.1	23.6%	34.3	19.9
	Distribution / Trend						
Home language	Kiswahili	27.0%	26.4*	13.6	29.1%	50.0*	27.5*
	English	4.0%	27.5	15.7	4.8%	40.8	23.6
	Other	72.6%	22.0	12.4	79.1%	43.1	24.2
School language	Kiswahili	75.0%	23.3	12.9	81.6%	43.8	24.5
	English	32.3%	28.9*	15.6*	44.7%	51.3*	27.9*
	Other	13.7%	14.6	7.1	9.1%	32.8	18.8
Pre-school attendance	No	13.5%	12.2	7.3	--	--	--
	Yes	81.7%	24.5*	13.3*	--	--	--

Note: The asterisk indicates a statistically significant difference between responses at the p<.01 level

Pupil Reading and Materials

Table 19 shows the results on pupil reading and materials at midline. The "Yes" responses were generally associated with higher passage reading scores. Having English and/or Kiswahili books and other materials at home was associated with higher ORF scores. Having someone read aloud at home did not make a difference for the Class 1 pupils, but did for Class 2 pupils. Silent story reading at home, practice reading aloud to the teacher or another pupil and practice reading silently at school were all associated with higher ORF scores.

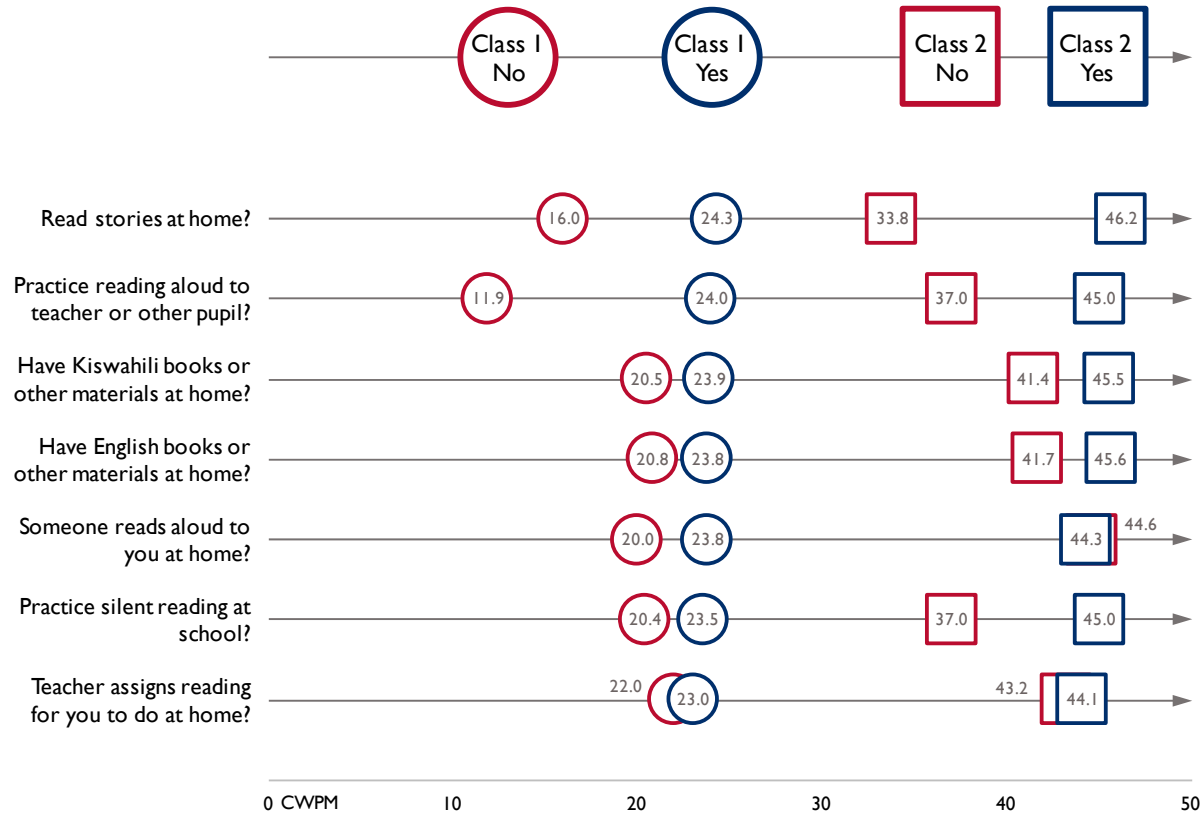
Table 19: Pupil Reading and Materials

Question	Response	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Have English books or other materials at home?	No	35.0%	20.8	11.9	36.1%	41.7	23.6
	Yes	61.7%	23.8	12.7	62.8%	45.6	25.5
Have Kiswahili books or other materials at home?	No	32.4%	20.5	11.6	33.3%	41.4	23.2
	Yes	64.5%	23.9*	12.9	65.7%	45.5*	25.6*
Someone reads aloud to you at home?	No	26.9%	20.0	10.6	31.3%	44.6	24.8
	Yes	68.4%	23.8*	13.1*	66.5%	44.3	25.0
Read stories at home?	No	18.3%	16.0	8.7	14.4%	33.8	19.8
	Yes	77.8%	24.3*	13.4*	83.5%	46.2*	25.7*
Practice reading aloud to teacher or other pupil?	No	8.2%	11.9	5.8	7.0%	37.0	19.7
	Yes	85.8%	24.0*	13.2*	90.9%	45.0	25.3*
Practice silent reading at school?	No	18.9%	20.4	10.9	15.4%	39.9	22.7
	Yes	74.7%	23.5	13.1	82.2%	45.1	25.3
Teacher assigns reading for you to do at home?	No	13.3%	22.0	12	12.4%	43.2	24.1
	Yes	83.0%	23.0	12.6	86.3%	44.1	24.7

Note: The asterisk indicates a statistically significant difference between responses at the p<.01 level

Figure 16 illustrates the differences in average ORF for those students **with** and **without** access to reading materials or practices. As noted, pupils who have access to reading materials and practice reading tend to have higher passage reading scores. The largest differences in performance are for pupils who read stories at home and those who practice reading aloud to their teacher or other pupils in school. For Class 2 pupils, silent reading at school also has a large difference in performance.

Figure 16: Pupil Reading and Materials



Pupil-School Characteristics

Table 20 details pupil-school characteristics. No clear trend exists in the effect of class size on reading outcomes. The pupils in the full-day shifts had higher scores than those in half-day shifts did. The results from comparisons involving single-grade vs. multi-grade classrooms were inconclusive due to small sample sizes for the multi-grade group.

Table 20: Pupil-School Characteristics

Characteristic	Group	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Class size	Below 21	3.7%	32.7	18.1	4.1%	42.7	25.2
	21-25	5.2%	16.5	9.1	5.6%	53.2	28.3
	26-30	4.7%	34.7	20.1	5.2%	52.2	28.1
	31-35	6.8%	31.3	16.5	7.8%	43.4	24
	36-40	8.6%	21.3	12.1	6.4%	41.1	25.6
	Above 40	71.0%	20.5	11.2	70.9%	42.6	23.9
	Distribution / Trend						

Characteristic	Group	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
School shift	Full day	59.4%	26.4	14.2	61.0%	47.6	26.2
	Half day	40.6%	16.0	9.3	39.0%	36.9	22.0
Multi-grade classrooms	No	97.7%	22.2	12.1	97.5%	43.6	24.4
	Yes	2.3%	17.8	17.0	2.5%	37.9	26.6

Teachers

Findings from the teacher questionnaire are presented in the following three sections: teacher characteristics; teacher reading materials and instruction; and teacher-school characteristics.

Teacher Characteristics

Table 21 has information on the teacher characteristics. The teachers' gender had an influence on the ORF scores; the scores by the pupils who were taught by female teachers were higher than those taught by male teachers.

The teachers' highest qualification was somewhat related to ORF. In general, a higher qualification had a positive effect on ORF for Class 1. However, the relationship was inconsistent at Class 2. Results for years of experience were inconclusive.

Table 21: Teacher Characteristics

Characteristic	Group	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Gender	Male	23.8%	19.0	11.8	18.6%	30.0	17.5
	Female	76.2%	24.2	13.0	81.4%	46.4	25.9
Highest qualification	Untrained	5.2%	18.8	12.1	1.0%	78.6	42.4
	P1 (Cert.)	43.4%	21.7	12.4	41.2%	41.1	23.2
	Diploma/S1	37.9%	24.5	12.9	39.7%	48.1	25.9
	Bachelor's	7.1%	28.3	17.3	12.7%	44.5	26.0
	Masters	0.0%	--	--	1.3%	37.8	20.7
	Other	6.5%	--	--	4.1%	--	--
	Distribution / Trend						
Years of experience	Below 6	16.5%	13.7	7.4	16.3%	33.5	19.0
	6-9	18.2%	13.6	7.3	16.6%	30.2	17.7
	10-19	20.8%	18.0	9.0	25.7%	31.5	18.7
	20-29	30.3%	14.5	7.6	25.7%	32.7	18.1
	Above 29	14.3%	19.9	10.6	15.6%	36.4	20.5

Characteristic	Group	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
	Distribution / Trend						

Teacher Oversight, Materials and Training

Table 22 shows the results from an analysis of teacher oversight, materials and training.

For the most part, a higher frequency of CSO observations is associated with higher pupil ORF scores. Pupils of Class 1 teachers who were observed just once per year averaged 19 CWPM in English and 10 CWPM in Kiswahili. In comparison, pupils of Class 1 teachers who were observed once per month averaged 26 CWPM in English and 13 CWPM in Kiswahili. The majority (60 percent) of Class 1 teachers were observed once per term, with an average of 22 CWPM in English and 13 in Kiswahili. The data from the Class 2 teachers show a similar trend. The pupils of teachers observed at least once per week were the highest, however, the number of responses in this category were less than 1 percent for Class 1 and 2 percent for Class 2.

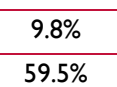
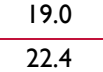
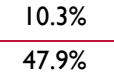
No clear pattern emerged for head teacher observations of teachers in either class or language. Most frequently, CSOs and head teachers observed teachers once per term or once per month.

ORF scores were higher for teachers who reported having a teaching guide. The vast majority of teachers reported having a teaching guide for Kiswahili (92 percent in Class 1, 96 percent in Class 2) and English (93 percent in Class 1, 94 percent in Class 2).

ORF scores were clearly higher for teachers who participated more frequently in Tusome training sessions. Most teachers participated in three to six training sessions (77 percent in Class 1, 76 percent in Class 2). Very few teachers did not participate in any training sessions (1 percent in Class 1, 2 percent in Class 2).

Finally, the ORF scores of pupils in schools that had classroom libraries were higher than those in schools with either school libraries or no libraries. However, about three-fourths of schools did not have a functioning library of either kind.

Table 22: Teacher Reading Materials and Instruction

Question	Response	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Frequency of observations from CSO?	Never	7.5%	21.4	13.2	8.5%	38.9	24.4
	1 per year	9.8%	19.0	9.6	10.3%	40.3	20.6
	1 per term	59.5%	22.4	12.9	47.9%	42.4	23.9
	1 per month	22.7%	26.4	13.4	30.6%	47.6	26.4
	1 per week	0.5%	44.5	20.5	2.6%	58.1	31.2
	Distribution / Trend						

Question	Response	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Frequency of observations from head teacher?	Never	2.4%	28.1	14.8	4.1%	51.9	21.5
	1 per year	1.7%	31.2	12.1	5.5%	44.5	25.5
	1 per term	49.4%	23.8	13.2	46.0%	42.2	23.8
	1 per month	36.9%	14.9	10.7	30.7%	48.3	27.1
	1 per week	9.6%	23.3	12.5	13.2%	44.3	25.2
	Distribution / Trend						
Guides for teaching Kiswahili?	Yes	91.6%	23.7	13.0	95.6%	43.5	24.6
	No	8.4%	15.0	9.3	4.4%	41.7	18.0
Guides for teaching English?	Yes	93.1%	23.4	13.0	93.5%	44.1	24.9
	No	6.9%	14.9	7.9	6.5%	35.2	18.1
Number of Tusome teacher training sessions?	None	1.1%	8.0	4.7	1.8%	27.3	26.6
	1-2 sessions	16.1%	16.7	9.3	11.2%	36.7	19.6
	3-4 sessions	45.0%	23.2	12.1	38.3%	40.6	23.7
	5-6 sessions	32.2%	25.9	14.5	37.2%	47.6	25.9
	More than 6	5.5%	22.2	17.8	11.4%	52.3	27.6
	Distribution / Trend						
Have a functioning library?	No	75.5%	22.0	11.8	72.5%	44.5	24.8
	In school	17.2%	23.8	14.2	19.1%	38.4	22.1
	In classroom	6.5%	30.7	17.8	7.4%	55.9	30.2
	In both	0.8%	23.1	12.5	1.1%	59.9	26.6

Head Teachers

Findings from the head teacher questionnaire are organized into three sections: head teacher characteristics; head teacher training and instructional supervision; and head teacher-school characteristics. For this questionnaire, the enumerators interviewed the head teachers. If the head teachers were not available, enumerators interviewed the deputy head teachers. For the midline, 83 percent of respondents were head teachers and 17 percent were deputy head teachers. Also, note that confounding factors may influence some of these data. For instance, head teachers with certain characteristics may be more often placed in schools that tend to have higher pupil ORF scores.

Table 23 provides results on head teacher characteristics. The pupils of schools with a female head teacher scored higher than those in schools with a male head teacher. Results for years in the position were inconclusive. In general, the pupils of schools with a head teacher who had a higher level of qualification scored higher than those with a head teacher who had a lower level of qualification. The exception is with the scores of Class 2 pupils in schools with an untrained head teacher, though the sample sizes for that category are small.

Table 23: Head Teacher Characteristics

Characteristic	Group	Percent	Class 1		Class 2	
			English ORF	Kiswahili ORF	English ORF	Kiswahili ORF
Gender	Male	83.4%	21.5	12.2	41.7	23.5
	Female	16.6%	29.5	15.6	55.4	29.3
Years in position	0-5	40.5%	16.1	8.9	33.7	19.4
	6-9	23.1%	15.9	7.8	30.1	16.8
	10-19	29.5%	15.7	8.0	32.3	18.5
	20-29	6.9%	13.0	7.0	36.9	21.6
	Distribution / Trend					
Highest qualification	Untrained	1.8%	18.0	9.6	52.1	26.7
	P1 (Cert.)	15.9%	22.0	13.4	43.7	24.2
	Diploma/S1	48.1%	22.1	12.3	44.3	25.0
	Bachelor's	30.0%	24.6	13.5	42.6	23.3
	Masters	1.8%	54.2	19.2	64.7	33.1
	Other	2.4%	--	--	--	--
	Distribution / Trend					

Table 24 provides information on head teacher training and instructional supervision in the school. Pupils at schools where head teachers have received training in school management have slightly higher ORF scores. On the other hand, results show little association between a head teacher having received training in reading and ORF. Similarly, almost no association exists between the frequency with which the head teacher observes teachers within the school and ORF scores. On checking lesson plans, the highest scores appear to be in schools that check plans once per week and once per day.

Table 24: Head Teacher Training and Instructional Supervision

Question	Group	Percent	Class 1		Class 2	
			English ORF	Kiswahili ORF	English ORF	Kiswahili ORF
Training in school management?	Yes	51.5%	24.9	13.8	45.9	25.2
	No	48.5%	21.6	11.9	42.6	24.1
Training in reading instruction?	Yes	82.8%	23.2	13.0	44.4	24.5
	No	17.2%	22.7	11.8	44.7	26.1
Responsible for teacher observation?	Head Teacher	84.1%	22.2	12.5	43.4	24.2
	Deputy Head Teacher	50.7%	25.6	14.0	45.9	25.7
	Other	8.2%	32.8	17.1	54.9	29.7
	No One	1.4%	25.2	11.0	61.3	30.6
Frequency of teacher observation per term?	Never	0.0%	--	--	--	--
	1 x	33.3%	25.8	13.7	47.5	25.9
	2 x	30.1%	19.3	11.3	45.4	25.0
	3 x or more	36.6%	23.1	13.0	39.8	23.0
	Distribution / Trend					
Responsible for lesson plan review?	Head Teacher	84.1%	24.2	14.3	46.9	24.8
	Deputy Head Teacher	50.7%	22.9	12.6	44.0	24.8
	Other	8.2%	42.7	21.9	59.8	31.8
	No One	1.4%	37.3	18.5	71.6	37.4
Frequency of lesson plan review?	Never	1.3%	12.0	5.1	58.7	31.2
	1 per year	3.2%	14.7	6.9	26.3	16.5
	1 per 2-3 months	19.0%	18.7	11.8	36.3	19.6
	1 per month	19.6%	21.8	12.4	40.9	23.3
	1 per 2 weeks	24.0%	23.5	12.5	44.8	25.3
	1 per week	28.6%	26.7	14.6	50.1	28.3
	1 per day	4.3%	25.4	13.5	50.2	27.1
	Distribution / Trend					

Table 25 summarizes head teacher-school characteristics. Results for pupils in schools with and without a functional library show small and inconsistent differences. The differences based on where the school gets its textbooks are inconclusive, as are the differences between schools with electricity, a feeding program and a computer room.

Table 25: Head Teacher-School Characteristics

Question	Group	Percent	Class 1		Class 2	
			English ORF	Kiswahili ORF	English ORF	Kiswahili ORF
Functioning school library?	Yes	17.9%	22.3	14.7	44.7	24.7
	No	82.1%	23.1	12.3	43.9	24.5
Who provides textbooks?	Ministry of Education	78.0%	22.3	12.5	43.2	24.5
	NGO	22.9%	19.9	11.7	43.4	24
	School	4.0%	16.0	8.5	28.1	16.1
	Board of Management	2.8%	14.5	9.9	39.9	25.3
	Parents	2.2%	23.5	11.2	47.7	23.5
	County Government	2.2%	23.1	11.9	41.4	24.6
	Other	17.8%	28.8	14.2	50.7	25.7
Electricity?	Yes	83.0%	23.0	12.7	44.7	25.0
	No	17.0%	23.1	13.5	39.1	22.0
Feeding program?	Yes	44.1%	24.9	13.5	45.1	25.3
	No	55.9%	21.4	12.1	43.1	23.9
Computer room?	Yes	45.9%	22.5	12.4	42.6	24.6
	No	54.1%	23.3	13.0	45.2	24.6

EVALUATION QUESTION 4

What community-level factors influence reading outcomes when implementing at scale, and how?

KEY FINDINGS

- Socio-economic status did not have a strong association with ORF, except for the wealthiest households.
- Levels of parental education tend to be associated with higher ORF scores for those parents who completed secondary or higher.

To assess community-level factors that influence reading outcomes, the evaluation team 1) interviewed pupils at the end of the EGRA testing and 2) conducted a household survey of pupils' parents. These findings are presented in the two sections that follow.

Pupils

Tables 26 and 27 provide information on pupil socioeconomic status (SES) collected from the pupils in relation to ORF scores in English and Kiswahili at baseline and midline. The SES index is a scale from 0 to 11, calculated from 11 equally weighted questions from the pupil questionnaire. Zero is the lowest SES category, and 11 is the highest.

For baseline and midline scores, the pupils in the upper part of the SES scale tended to have higher ORF scores, though the differences were often small. These scores have a larger spread and show more inequality than the SES findings from the household survey, as noted below.

The relationship between changes in ORF between baseline and midline and SES are mixed. For the Class 1 pupils in both languages, the gains between baseline and midline were similar for the bottom three SES groups (0-9), but almost twice as large for pupils in the highest SES group (10-11). For Class 2 in both languages, the gains were similar across all SES groups. In contrast, the lowest SES group (0-3) shown the largest gains in Class 2.

Table 26: Pupil Socio-economic Status, English ORF Scores

SES Group	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
0-3 (lowest)	7.5	19.6	12.1	20.4	42.8	22.4
4-6	11.9	23.6	11.7	23.4	43.5	20.1
7-9	13.7	26.8	13.1	30.9	46.7	15.8
10-11 (highest)	12.8	34.4	21.6	28.8	47.6	18.8
Trend						

Table 27: Pupil Socio-economic Status, Kiswahili ORF Scores

SES Group	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
0-3	3.7	11.3	7.6	12.3	24.2	11.9
4-6	5.5	12.8	7.3	13.5	24.4	10.9
7-9	6.1	13.1	7.0	16.5	26.0	9.5
10-11	5.0	20.0	15.0	13.7	25.1	11.4
Trend						

Households

The findings on households collected through the phone interviews are presented on parents' level of education and socio-economic status. The response rate was 49 percent of assessed students' households. These pupils tended to have higher ORF scores than the pupils in households that were not reached, which should be taken into consideration when interpreting the findings. However, there was not much difference in pupil-reported SES between households that were reached and not reached.

Parents' Level of Education

Table 28 shows the data on parents' level of education and ORF. In general, higher levels of education for the mothers and fathers are positively related to higher English and Kiswahili ORF scores. The

trends are more pronounced for the English ORF scores than for the Kiswahili ORF scores. This is especially true for pupils whose parents' have secondary education or higher.

Table 28: Parents' Level of Education

Question	Group	Class 1			Class 2		
		Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Mother's highest level of education attained	Non-formal	7.7%	23.1	13.3	7.1%	47.7	25.5
	Incomplete primary	33.8%	22.0	11.4	30.6%	41.8	23.9
	Primary	33.3%	23.3	13.0	37.5%	47.7	26.2
	Incomplete Secondary	9.9%	30.3	18.6	11.8%	47.3	26.9
	Secondary	9.2%	32.8	15.8	9.4%	53.5	28.3
	Diploma	3.1%	45.8	28.7	2.7%	62.2	31.7
	Undergraduate or higher	0.7%	36.1	16.8	0.3%	91.9	42.0
	Distribution / Trend						
Father's highest level of education attained	Non-formal	6.0%	24.5	12.9	5.8%	39.6	22.4
	Incomplete primary	20.4%	18.9	10.4	20.0%	42.1	24.2
	Primary	29.6%	24.7	13.9	30.7%	47.5	25.7
	Incomplete Secondary	7.9%	27.6	15.0	10.6%	42.8	24.8
	Secondary	16.8%	26.7	14.5	17.7%	49.6	26.5
	Diploma	5.3%	40.6	21.7	5.1%	57.2	32.9
	Undergraduate or higher	2.1%	44.4	25.7	1.6%	71.1	35.9
	Distribution / Trend						

Figures 17 and 18 illustrate the difference in ORF in English and Kiswahili for Class 2 based on the parents' levels of education.

Figure 17: Class 2 English ORF by Parents' Level of Education

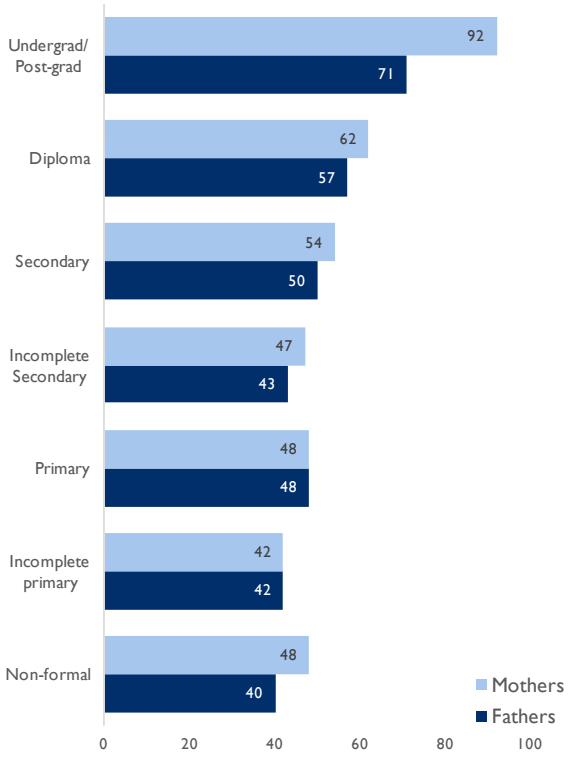
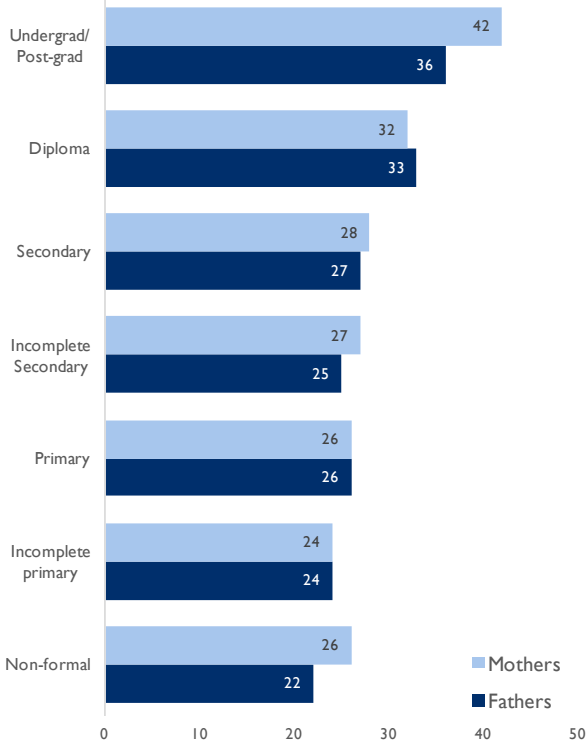


Figure 18: Class 2 Kiswahili ORF by



Household Socio-Economic Status

To establish a measure of socio-economic status, the evaluation team used a simplified version of the 2014 Kenya Demographic and Health Survey questions to create a wealth index.[6] Households were then grouped into quintiles to compare socio-economic status and ORF. Note that the percentages of households do not equal 20 percent per quintile in Table 29 because of weighting.

As noted above, the distribution of ORF scores in relation to SES is narrower in the household survey than in the student questionnaire. The evaluation team's interpretation of this difference is that it may be due to the differences in wording in the questionnaires, as the distribution of student SES was similar for those households reached versus those households not reached by the phone survey (see Annex II: Sampling).

The table shows little difference in ORF scores for households in the first four quintiles, that is for 80 percent of the population. The highest quintile (top 20 percent of households) is associated with markedly higher ORF scores. The evaluation team could not conduct a trend analysis for these data, as the household survey was only conducted at midline.

6 Kenya National Bureau of Statistics (December 2015). Kenya Demographic and Health Survey. Nairobi, Kenya. http://dhsprogram.com/publications/publication-FR308-DHS-Final-Reports.cfm

Table 29: Household Socio-Economic Status

Wealth Quintile	Class 1			Class 2		
	Percent	English ORF	Kiswahili ORF	Percent	English ORF	Kiswahili ORF
Lowest	19.4%	22.7	12.9	20.8%	42	23.2
Second	18.8%	21.2	13.4	21.7%	42.1	25.3
Middle	21.4%	22.2	11.8	17.9%	43.7	25.1
Fourth	20.5%	25.3	13.0	19.7%	45.0	24.7
Highest	20.0%	33.9	18.4	19.9%	60.8	31.1

Figures 19 and 20 illustrate the similarities between ORF scores for the first four quintiles of the index.

Figure 19: ORF by Wealth Quintiles, English

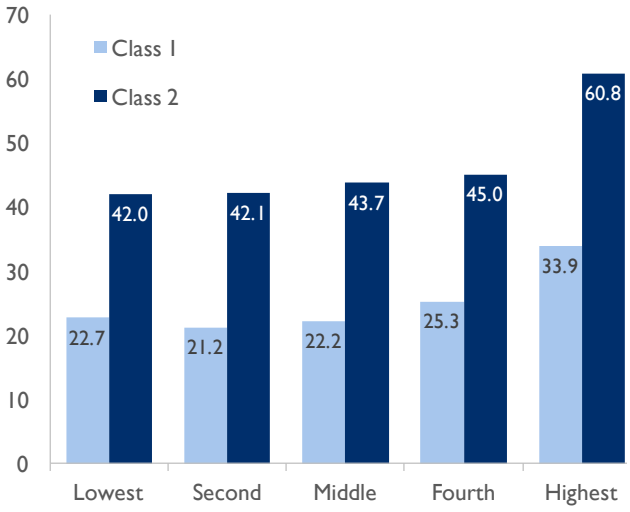
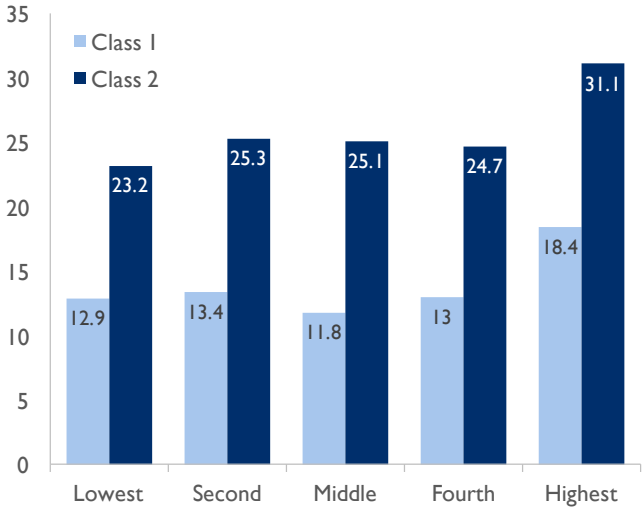


Figure 20: ORF by Wealth Quntiles, Kiswahili



EVALUATION QUESTION 5

To what extent have the Tusome Early Grade Reading (EGR) activity components been implemented in schools nationwide?

KEY FINDINGS

- At midline, 98 percent of teachers have received at least some Tusome training. Thirty-eight percent of Class 1 and 48 percent of Class 2 teachers reported participating in five or more Tusome training sessions.
- At midline, 83 percent of head teachers reported that they had received reading instruction training in the past 12 months.
- At midline, 99 percent of teachers had a Tusome teacher's guide in their classroom.
- At midline, 97 percent of Class 1 and 95 percent of Class 2 classrooms had at least one Tusome pupil's book per pupil.
- At midline, 96 percent of classrooms had at least one exercise book per pupil.
- At midline, 84 percent of Class 1 and 82 percent of Class 2 teachers reported being observed about once per term by a CSO.
- At midline, 96 percent of Class 1 and 90 percent of Class 2 teachers reported being observed about once per term by their head teachers.
- At midline, 54 percent of CSOs had observed 15 or more lessons in the last 30 days.

To establish the extent to which Tusome Early Grade Reading activity components have been implemented nationwide, this evaluation focused on Tusome's Intermediary Result (IR) 1: Improved supervision, support, and delivery of reading instruction to target pupils. The findings from the teacher survey, classroom observation, and the CSO survey are presented in the three sections below organized in relation to the three related sub-IRs:

- IR 1.1: Increased availability and use of appropriate textbooks and supplementary materials that support reading
- IR 1.2: Improved methods of reading instruction delivery
- IR 1.3: Supervision and support provided to teachers by tutors/coaches and head teachers in teaching and assessing reading

Information on the IRs can be found in the three sections below, with some of the information on the IRs provided in more than one section. As noted above in Project Background, Tusome's goal is to reach all public school Class 1 and Class 1 pupils, teachers, and head teachers, as well as a limited number of APBET Class 1 and Class 1 pupils, teachers, and head teachers. The findings provide information on whether this has been attained.

Materials

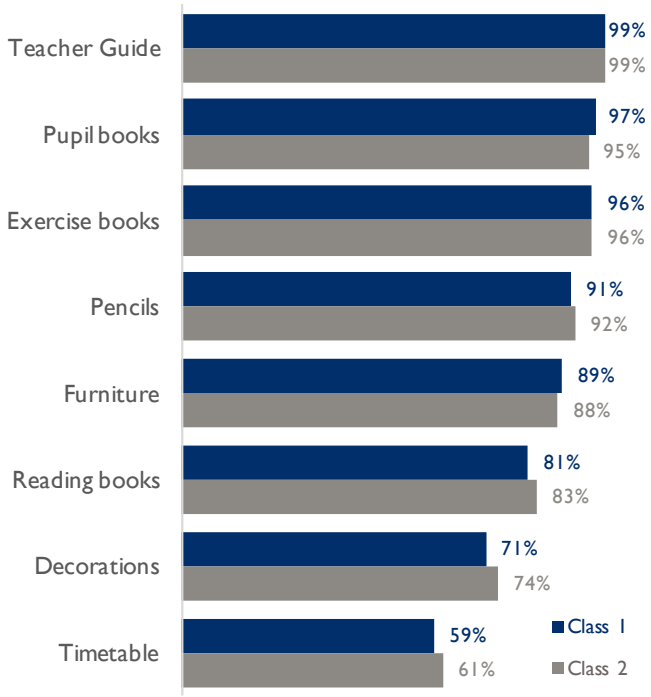
As noted in Evaluation Question 3, 92 percent of Class 1 teachers and 96 percent of Class 2 teachers reported having a teaching guide for Kiswahili reading instruction during the teacher survey. For English reading instruction, 93 percent of Class 1 teachers and 94 percent of Class 2 teachers reported having a teaching guide. During the classroom observation, the evaluation team observed that 99 percent of teachers had a Tusome teacher's guide in their classroom. Ninety-seven percent of Class 1 and 95

percent of Class 2 classrooms had at least one Tusome pupil's book per pupil. Ninety-six percent of classrooms had at least one exercise book per pupil. Differences in materials between Class 1 and Class 2 classrooms were minimal. Additional materials are noted in Table 30 and Figure 21.

Table 30: Materials Observed in Classroom

Observation	Class 1	Class 2
Tusome teacher's guide in classroom	98.8%	99.2%
Tusome pupil's book for each pupil	96.8%	95.4%
Exercise books for each pupil	96.0%	95.8%
Pencils for each pupil	90.6%	92.0%
Child-sized tables and chairs	89.4%	88.3%
Reading books for the pupils	80.8%	82.7%
Decorations/materials on the walls	71.3%	74.2%
Timetables on the wall	59.0%	61.2%

Figure 21: Materials Observed in Classroom



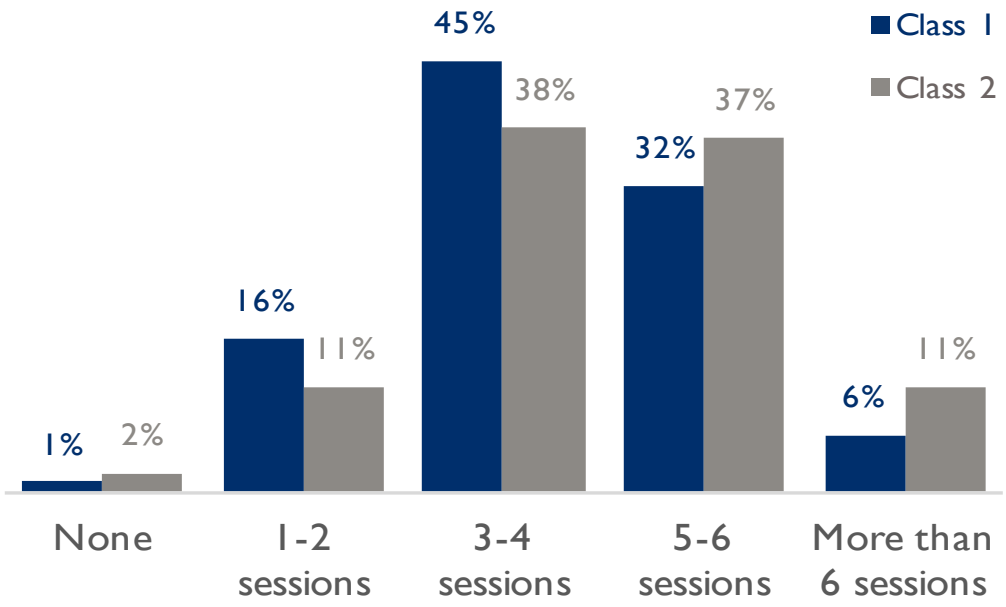
Reading Instruction Training and Delivery

As Table 31 and Figure 22 show, 83 percent of Class 1 and 87 percent of Class 2 teachers reported participating in 3 or more Tusome training sessions. The rest of the teachers either participated in one to two sessions or did not attend any sessions at all.

Table 31: Teacher Participation in Tusome Training

# of Tusome Training Sessions Attended	Class 1	Class 2
None	1.1%	1.8%
1-2 sessions	16.1%	11.2%
3-4 sessions	45.0%	38.3%
5-6 sessions	32.2%	37.2%
More than 6 sessions	5.5%	11.4%

Figure 22: Teacher Participation in Tusome Training



Teachers were asked a series of questions about their reading instruction during the teacher survey. These questions are related to the Tusome methodology. They were written in collaboration with USAID, MOE, and the Tusome team. As Table 32 shows, the majority of teachers reported that they cover all of these curriculum elements five days per week, except for pupils retelling a story that either the teacher or the pupil first reads.

Table 32: Frequency of Teacher Instruction Methods – Number of Days per Week

Task	Class 1							Class 2						
	0	1	2	3	4	5	Distribution	0	1	2	3	4	5	Distribution
Choral repetition/reading	1%	2%	11%	15%	4%	68%		1%	3%	10%	17%	5%	64%	
Copied text from blackboard or textbook	1%	4%	9%	9%	7%	70%		1%	3%	11%	16%	10%	59%	
Pupils retold story you read to them	13%	31%	20%	7%	2%	28%		10%	26%	29%	12%	5%	18%	
Pupils retold story they read	20%	24%	20%	13%	1%	22%		20%	28%	25%	9%	4%	15%	
Pupils sounded out unfamiliar words	2%	4%	18%	18%	6%	52%		5%	5%	17%	20%	9%	44%	
Pupils learned meanings of new words	1%	4%	10%	17%	11%	58%		1%	5%	17%	19%	9%	50%	
Pupils read aloud to teacher	1%	9%	9%	12%	10%	59%		2%	7%	18%	19%	11%	43%	
Pupils answered questions (text you read)	2%	14%	14%	8%	7%	56%		1%	8%	9%	15%	12%	55%	
Pupils answered questions (text they read)	6%	13%	11%	10%	9%	51%		7%	8%	17%	13%	7%	49%	
Pupils assigned reading during school time	5%	14%	13%	9%	4%	56%		12%	13%	15%	18%	4%	38%	
Pupils assigned reading to do at home	9%	14%	6%	9%	2%	61%		7%	14%	5%	8%	4%	63%	

The classroom observation checklist, which was also developed in collaboration with USAID, MOE, and the Tusome team, asked questions on the following issues: teacher focus, instructional content, the teacher's actions, the pupils' actions, and the materials in use. Observations were recorded every three minutes. Table 33 summarizes the results as average percent of observations for the duration of the class period. Figures may exceed 100 percent within categories, as the observers could mark more than one response per period.

Teachers spent the majority (79 percent) of the class time focused on the whole class. Teachers spent the most time on comprehension (32 percent for Class 1, 36 percent for Class 2), followed by vocabulary tasks (26 percent for Class 1, 29 percent for Class 2). They spent the least amount of time on alphabetic principle (10 percent for Class 1, 8 percent for Class 2). The teachers spent their time doing a variety of tasks – lecturing, reading, asking questions, listening to pupils, and monitoring pupils.

Pupils spent a large part of the class period listening to the teacher (40 percent for Class 1, 37 percent for Class 2). They also spent a substantial amount of time on choral reading (25 percent for Class 1, 23 percent for Class 2) and reading out loud (10 percent for Class 1, 13 percent for Class 2).

Teachers were actively using their teacher's guide for over half of the class period (61 percent for Class 1, 62 percent for Class 2). Pupils were actively using their pupil's book for just under half of the class period (42 percent for Class 1, 43 percent for Class 2). There was also frequent use of the blackboard (41 percent for Class 1, 38 percent for Class 2). Other materials, such as exercise books, letter cards and pocket chart, were less frequently used.

Table 33: Classroom Observation

Question	Groups	Class 1	Class 2
Teacher Focus	Whole class	79.2%	78.7%
	One individual pupil	11.4%	13.1%
	Small group	6.2%	6.3%
	Not focusing on pupils	2.9%	2.3%
	Teacher not present/disengaged	0.8%	0.4%
Instructional Content	Comprehension	31.5%	35.9%
	Vocabulary	26.2%	28.9%
	Phonological awareness	18.9%	11.6%
	Fluency	15.3%	18.5%
	Alphabetic principle	9.5%	7.8%
	Teacher not present/disengaged	1.5%	0.8%
Teacher Action	Lecturing/explaining	23.0%	21.6%
	Reading	20.0%	20.9%
	Asking questions	18.5%	17.2%
	Listening to pupils	15.1%	17.1%
	Monitoring pupils	12.9%	15.3%
	Writing	10.8%	9.7%
	Giving feedback	4.4%	3.6%

Question	Groups	Class 1		Class 2	
	Teacher not present/disengaged	1.2%		0.6%	
Pupil Action	Listening to the teacher	39.8%		37.2%	
	Choral reading	25.1%		22.5%	
	Writing	11.4%		13.1%	
	Individual reading out loud	9.80%		13.3%	
	Repeating/recitation	9.5%		8.9%	
	Partner reading	3.0%		3.2%	
	Silent reading	2.8%		3.4%	
	Off task/uninvolved	1.9%		1.5%	
Materials Used	Teachers Guide	60.7%		62.1%	
	Pupils Book	41.7%		43.1%	
	Blackboard	41.2%		38.3%	
	Exercise Books	14.9%		14.5%	
	Letter Cards	4.6%		2.1%	
	Pocket Chart	3.4%		2.6%	
	No material used	2.7%		1.8%	

Teaching Supervision and Support

The Tusome Performance Monitoring Plan (PMP) indicator 1.3.4 is "Percentage of Tusome Class 1 and Class 2 trained teachers observed three times or more per year." As such, the evaluation team used the category of once per term or higher to interpret the findings below.

Table 34 and Figure 23 show that 84 percent of Class 1 and 82 percent of Class 2 teachers reported being observed at least once per term by a CSO. About 23 percent of Class 1 and 33 percent Class 2 teachers reported being observed once per month or more.

Table 34: CSO Observation of Teachers

Frequency of classroom observation by CSO	Class 1	Class 2
Never	7.5%	8.5%
1 per year	9.8%	10.3%
1 per term	59.5%	47.9%
1 per month	22.7%	30.6%
1 per week	0.5%	2.6%

Figure 23: CSO Observation of Teachers

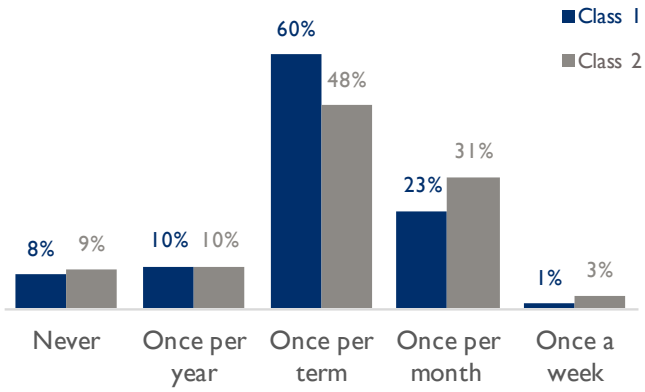
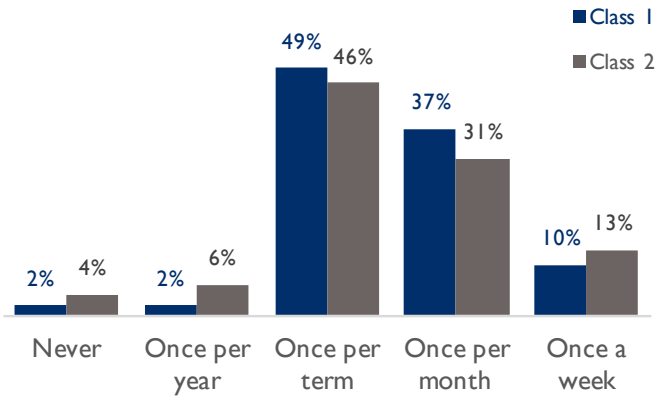


Table 35 and Figure 24 show the frequency of head teacher observation of teachers was higher than CSO observation, with 96 percent of Class 1 and 90 percent of Class 2 teachers reported being observed at about once per term by their head teachers. About 47 percent of Class 1 and 44 percent Class 2 teachers reported being observed once per month or more.

Table 35: Head Teacher Observation of Teachers

Frequency of classroom observation by CSO	Class 1	Class 2
Never	2.4%	4.1%
1 per year	1.7%	5.5%
1 per term	49.4%	46.0%
1 per month	36.9%	30.7%
1 per week	9.6%	13.2%

Figure 24: Head Teacher Observation of Teachers

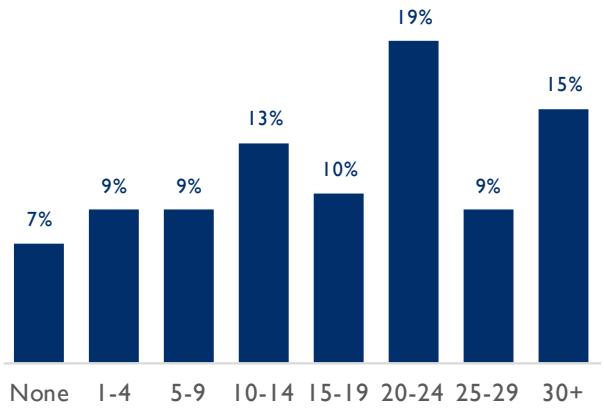


The Tusome indicator for CSO observation indicator 1.3.5 is "Percentage of TAC tutors and instructional coaches observing at least 15 lessons in a month." According to the CSOs who were interviewed, over half (53 percent) of them had observed 15 or more lessons in the last 30 days. As Table 36 and Figure 25 show, 7 percent said that they had not observed any lessons and 15 percent said that they had observed 30 or more lessons.

Table 36: Number of Lessons Observed by CSOs

Lessons observed per month by CSO	Percent
None	6.6%
1-4 lessons	8.8%
5-9 lessons	8.7%
10-14 lessons	13.1%
15-19 lessons	10.2%
20-24 lesson	19.1%
25-29 lessons	8.7%
30 or more	15.3%

Figure 25: Number of Lessons Observed by CSOs



EVALUATION QUESTION 6

To what extent can any incremental changes in early grade reading outcomes throughout Kenya be correlated with or attributed to the scale up of Tusome?

KEY FINDINGS

- Effect sizes from baseline to midline ranged from 0.40 to 1.07 for Class 1 and 0.41 to 2.57 for Class 2, almost all of which would be considered large in social science research.
- Effect sizes for Tusome were higher than those for the PRIMR pilot, which were either moderate or large.
- Though direct comparisons between this study and other regional results is difficult due to methodological differences in programs and assessments, the Tusome results were found to be about twice as high as those in Tanzania.

As Tusome is a national program that is being implemented in all schools simultaneously, there is no control or comparison group to compare the activity's gains. However, the evaluation team did examine the effect sizes seen during the prior pilot study, PRIMR, to contextualize the gains seen between the Tusome baseline and midline. The evaluation team also looked at the results of other USAID reading programs in the region.

Effect Size

In social science research, an effect size of 0.5 is considered to show a large impact.

As shown in Table 37, the effect sizes for Tusome between baseline and midline range from 0.40 to 1.07 for Class 1 and from 0.41 to 2.57 for Class 2 for English and Kiswahili. The highest effect sizes for the early skills of phoneme segmentation and letter sound knowledge were the highest.

Table 37: Tusome Reading Effect Sizes

Subtask	English		Kiswahili	
	Class 1	Class 2	Class 1	Class 2
Phoneme segmentation	1.07	2.57	--	--
Letter sound knowledge	0.71	1.63	0.75	1.32
Syllable fluency	--	--	0.66	0.80
Invented/non-word decoding	0.52	0.68	0.45	0.50
Vocabulary	0.48	0.41	--	--
Passage reading (A)	0.67	0.72	0.75	0.71
Reading comprehension (A)	0.40	0.49	0.62	0.69
Passage reading (B)	0.73	0.86	--	--
Reading comprehension (B)	0.75	0.94	--	--
Listening comprehension	--	--	0.52	0.52

The PRIMR activity was designed to include an impact evaluation based on treatment and control groups. Table 38 shows that the PRIMR endline report showed effect sizes for selected measures of 0.28 to 0.68 for Class 1 and 0.30 to 0.78 for Class 2 in English and Kiswahili. In general, these effect sizes were lower than the Tusome effect sizes.

Table 38: PRIMR Reading Effect Sizes

Subtask	English		Kiswahili	
	Class 1	Class 2	Class 1	Class 2
Letter sound fluency	0.68	0.78	0.57	0.70
Syllable fluency	--	--	0.42	0.45
Oral reading fluency (passage reading)	0.44	0.45	0.41	0.35
Reading comprehension	0.38	0.44	0.45	0.32
Reading at benchmark (% of pupils reading 65+ CWPM)	0.32	0.45	--	--
Reading at benchmark (% of pupils reading 45+ CWPM)	--	--	0.28	0.30

Regional Comparisons of Reading Gains

The evaluation team examined ORF scores under USAID-supported projects in the Eastern and Southern Africa region in order to develop a sense of the relative success of the student gains in Kenya. Please note that making comparisons of reading scores or gains across projects must be done with caution. Any comparisons are going to have validity concerns due to inevitable differences across projects such as:

- Grade levels targeted
- Languages of instruction
- Timing of the assessments
- Intensity of the interventions
- Difficulty levels of the tools
- Sampling plans
- Type of evaluation design

Differences in grade levels, languages, and evaluation designs were judged as insurmountable in terms of making valid comparisons. Each of the projects needed to measure ORF at Grade 2 in the same language (either English or Kiswahili) using a cross-sectional design, that is different students in the same grade level across time (and not a longitudinal design with the same students in different grade levels across time). The other issues were judged as potentially acceptable for the comparisons, though any comparisons could be invalid if the differences across projects were too great. The team did not have enough information on these issues – such as the intensity (i.e., dosage or uptake) of the interventions, or difficulty level of the tools – to necessarily negate the cross-project comparisons.

Keeping these caveats in mind, the evaluation team looked at four other countries in the region that have benefitted from USAID support for possible comparisons with Kenya: Rwanda, Tanzania, Uganda, and Zambia. Unfortunately, it was only possible to make comparisons between Kenya and Tanzania due to the following issues in the other countries.

- Rwanda: Only measured Kinyarwanda at Grade 2, with English at Grade 4.
- Tanzania: Measured English and Kiswahili at baseline but only Kiswahili at midline.
- Uganda: Used a longitudinal evaluation design.
- Zambia: Measured ORF in local languages but not in English.

Kenya and Tanzania both had cross-sectional evaluation designs with two time points and measurements of ORF at Grade 2 in the same language (Kiswahili). The Kenya baseline was in 2015 and the midline in 2016. The Tanzania baseline was in 2013 and the midline in 2016. For Kiswahili, the Kenya ORF benchmark was 45 CWPM and the Tanzania ORF benchmark was 50 CWPM. Note that both are projected to have endlines in 2018.

As shown in Table 39 below, the gains in Kenya were about twice as high as those for Tanzania, with an ORF increase of 12 CWPM in Kenya and 6 CWPM in Tanzania. The percentage of non-readers decreased by 24 percentage points in Kenya and 12 percentage points in Tanzania. There were increases of 8 percent of students in Kenya and 2 percent in Tanzania achieving the ORF benchmarks.

Table 39: Class 2 Kiswahili Reading Gains in Kenya and Tanzania

Characteristic	Kenya			Tanzania		
	Baseline (2015)	Midline (2016)	Difference	Baseline (2013)	Midline (2016)	Difference
ORF Average	16	28	+12	18	24	+6
% of Non-Readers	43%	19%	-24%	28%	16%	-12%
% at Benchmark	4%	12%	+8%	5%	7%	+2%

CONCLUSIONS

Based on the findings, the evaluation team reached the following conclusions:

- The Tusome approach is having a strong, positive influence on reading outcomes, with relationships between project implementation and reading outcomes.
- Reading outcomes for Class 1 and 2 pupils greatly improved during the one-year period between the baseline and midline evaluations. While impressive gains have been made, continuing with the Tusome approach will be critical to sustaining or improving on those gains.
- The Tusome project has achieved a high level of national implementation of activities. Given that project activities such as CSO observations, in-service training and access to materials are associated with higher ORF scores, the high level of implementation across all schools appears to be a key part of its success. The effect sizes seen during the PRIMR pilot have been at least sustained, and in most cases strengthened, in the national scale-up of Tusome.
- The evaluation methodology and implementation resulted in valid, reliable data for the midline evaluation, including the changes from baseline to midline.

RECOMMENDATIONS

Based on its fieldwork, data and workshops, the evaluation team has the following eight recommendations.

The evaluation team recommends that the **Tusome project**:

1. Continue a high level of implementation fidelity in its support for materials, instruction and supervision in early reading activities to further increase reading gains.
2. Conduct additional analysis using the midline dataset to see what programmatic insights can be used for improved activity implementation.

The evaluation team recommends that **USAID/KEA**:

3. Use the findings of this evaluation in its continued support of materials, instruction and supervision in early grade reading activities.
4. Share the evaluation findings in other USAID early grade reading projects beyond Kenya to increase regional and international collaboration and learning.

The evaluation team recommends that the **MOE**:

5. Continue its support of early reading activities and evaluations to ensure further ministry ownership of the Tusome implementation and results.
6. Set benchmarks and targets for reading comprehension, in addition to the ORF benchmarks, to monitor pupil progress in comprehension over time.

The evaluation team recommends that the **team tasked with the endline evaluation**:

7. Use the data collection tools, sampling plan and data collection schedule used at midline to ensure valid, reliable and interpretable data.
8. Continue the strong collaboration with the MOE for the implementation of the study, including tools revision, training and data collection.

ANNEXES

ANNEX I: EVALUATION STATEMENT OF WORK

SECTION C – STATEMENT OF WORK

C.1. BACKGROUND

The USAID/Kenya and East Africa (KEA) Office of Education and Youth (EDY), in collaboration with the United Kingdom's Department for International Development (DFID), is implementing a $55 million basic education initiative to improve the reading skills of the approximately 5.4 million individual Kenyan children who began primary school during the 2015-2017 school years[1]. The project, Tusome (Too-SOH-meh; "Let's Read" in Kiswahili), will continue through March 2018, and has integrated options for transition to government ownership during the out years.

Tusome builds on research-based reading initiatives to create a sustainable and affordable national reading program in Kenya. It seeks to improve the English and Kiswahili skills of Standard 1 and 2 children in approximately 23,600 formal and non-formal[2] public and low-cost private primary schools across the country.[3] Tusome also assists the Government of Kenya (GOK) at the technical and policy levels to sustainably improve reading skills beyond the life of the project.

USAID/KEA support for education development activities has been provided mostly through US-based partners, as local government entities and non-governmental organizations have lacked the capacity to implement literacy development activities at similar scale and quality. As a result, one of the primary objectives of the Tusome award is to build the capacity of the Government of Kenya (specifically the Ministry of Education, Science and Technology (MoEST)) and relevant Semi-Autonomous Government Agencies (SAGAs) to implement literacy activities. USAID/KEA anticipates that, in year 4, Tusome will transition to GOK implementation of activities.

Two Tusome modalities of note are: the *Tusome Partnership Fund* (to leverage private sector support, particularly in the development, production, and distribution of materials) and the *Tusome Youth Fund* (to support youth engagement in improving literacy in the early grades).

C.2. STATEMENT OF WORK

C.2.1. Purpose of the Evaluation

Tusome seeks to improve children's reading skills on a nationwide scale through evidence-based programming. In line with USAID policy, the testing of innovative activities is built into the Tusome design and the previous PRIMR initiative, which developed and tested the methodology that lies at the heart of Tusome. This evaluation will serve to:

1) Document progress towards the stated goals of Tusome during the project period, including supporting achievement of Goal One of the USAID Education Strategy to improve the

[1] Approximately 1.3 – 1.4 million children enter Standard 1 each year.
[2] Non-formal schools – or "low-cost private community schools" – are the predominant form of schooling for children in Kenya's informal urban settlements ("slums"). The 2011 UWEZO report estimates 20% of Kenya's school-age children attend private schools, with rates much higher in slums. These community private schools charge a nominal fee to educate thousands of underserved and poor children who would not otherwise have access to schooling despite a national free primary education policy. There are approximately 1,000 non-formal schools in Kenya, with the majority concentrated in Nairobi, Thika, Nakuru, Eldoret, Kisumu, Mombasa, and Embu. Tusome is working, at Government of Kenya request, in 1,000 of these schools, in addition to 22,600 public primary schools across all 47 counties.
[3] In 1,300 schools in a few select counties – Bungoma, Kisumu, Kiambu, Machakos, Nairobi, and Nakuru – the Primary Math & Reading (PRIMR) program will continue with USAID and DFID funding through August 2014 (USAID-funded schools) and February 2015 (DFID-funded schools). PRIMR will continue implementation in those schools, with the contractor assuming responsibility in 2015.

reading skills of 100 million children worldwide. This discussion must include progress made according to baseline findings.

2) Elucidate the differences in early grade reading outcomes based on contextual factors at the pupil, school and community levels.

The audience for the evaluation is: a) USAID, including headquarters, the Kenya Mission, and other Education Offices around the world; b) Kenya's government, including the MoEST, relevant SAGAs[4], the Teachers Service Commission, and county governments; c) education sector development partners, the private sector, and other stakeholders in and outside of Kenya; and d) the Kenyan public.

This will be a USAID/Forward Quality Evaluation when conducted, so the evaluation design must meet those criteria. A mix of qualitative and quantitative methods applied in a systematic and structured way will be expected. The midline evaluation shall be designed in accordance with Agency guidance (ADS 201 and 203, Evaluation Policy)

C.2.2. Dissemination and Utilization Plan

It is expected that the findings will be used to inform program implementation for early grade reading in and outside Kenya. The initial draft findings will be shared with the USAID/KEA Office of Education and Youth and the implementing partner for necessary review. After the report is finalized, the findings will be shared with relevant offices in USAID/KEA and USAID/Washington, as well as relevant education sector development partners and other stakeholders, including the private sector (e.g., publishing houses and media). For the GOK, the findings will be shared through national and county fora (several counties could be clustered into one region), mainly targeting the MoEST and related SAGAs. Findings of the Early Grade Reading Assessments (EGRAs) will also be shared with the schools.

The deliverables must include presentations to the senior Ministry officials together with the various presentations to other key stakeholders. In addition to the detailed report, a two page report will be required that can be presented to other key stakeholders such as head teachers, teachers, county officials, parents and School Boards of Management. These various smaller reports should take note that the information may be more for information sharing as opposed to the detailed report which is for decision making. The evaluators must develop reporting mechanisms in consultation with USAID to ensure the needs of each stakeholder group is met; indicatively, the contractor must conduct a series of forums at a national and county level to disseminate findings to GOK stakeholders; the results of the Early Grade Reading Assessments (EGRAs) must be shared with schools; and reports must be prepared for USAID in a range of media.

C.2.3. Identification of Intervention(s) to Be Examined

<u>Early Grade Reading Assessments (EGRAs)</u>

[4] SAGAs include the Kenya Institute of Curriculum Development (KICD), the Kenya National Examinations Council (KNEC), the Kenya Institute for Special Education KISE), and the Kenya Education Management Institute (KEMI).

The Tusome midline evaluation must primarily focus on utilizing the EGRA tool in Kiswahili and English to measure students' reading fluency and comprehension.

External evaluations have measured "improved reading skills" using the guidance from USAID's 2011 Education Strategy Technical Notes, revised April 2012[5], which defines them as increases in fluency and comprehension in reading grade level text (at grade 2). Fluency is the ability to read text accurately, quickly, and with good expression and is calculated based on words correct per minute read; while comprehension is understanding the meaning of what has been read. The recommended indicator to measure reading with comprehension is based on the point at which words correct per minute (wcpm) produces 80 percent reading comprehension.

USAID seeks data related to the standard indicator "Percent of learners who demonstrate reading fluency and comprehension of grade level text at the end of grade 2 with USG assistance."

Additionally, the Contractor must test a small sample of longitudinally tracked students both in July and in October to establish the amount of correction needed to estimate project contribution to improvements in reading, as the baseline was conducted after the start of the intervention in 2015.

Contextual Profile

In the context of a national scale-up, inevitably some schools will have better results than other schools, and USAID is interested in understanding this phenomenon in a way that can inform future activities. Thus, in addition to conducting EGRAs, the contractor must collect a contextual profile of the school and its community, and to analyze this data in comparison to the EGRA findings. The idea is that the information collected in the contextual profile will help elucidate the reasons behind this difference in performance. The contractor must propose information to be included, which may include:

- student-level factors (e.g., birth order, background, gender, rural/urban home, participation in early childhood development, access to a school feeding and health programs)
- school-level factors (e.g., pupil/teacher ratio, tutor/school ratio, level of education of teacher/head teacher, class size, language of instruction, teaching and learning materials availability, time on task, ICT integration, electricity, presence and involvement of school management committees)
- community-level factors (e.g., access to cash subsidy, access to clean/safe drinking water, prevalence of HIV/AIDs, stunting, underweight and malaria school ratio), socioeconomic data
- language policy complexities

C.2.4. Evaluation Design

The overall design for this evaluation must be a non-experimental impact evaluation. This approach entails a comparison of three intervention groups: a baseline or pre-intervention

[5] Education Strategy Technical Notes: http://pdf.usaid.gov/pdf_docs/PDACT681.pdf; offerors may also want to review the Education Strategy Implementation Guidance: http://transition.usaid.gov/our_work/education_and_universities/pdfs/2012/ED_implementation_guidance_2011.pdf

group, a midline or intermediate intervention group, and an endline or advanced intervention group.[6] Project impact must be theorized, but not proven, by comparing project indicators and early grade reading outcomes between and within each group, and measured against school-level and community-level factors. This approach must allow researchers to determine the extent of project implementation over the course of 5 years, detect and compare differential changes in early grade reading indicators, and correlate—but not attribute—these changes to project, school, and community variables. The evaluation design described below is based on the assumption that the Tusome Project is similar in design to the PRIMR initiative.

Key Evaluation Questions

The overall research questions for the Tusome Early Grade Reading Activity evaluation are as follows:

1. What proportion of students can demonstrate they can read grade-level text (within Kenya's curricular goals) by the end of Primary 1 and Primary 2?
2. What proportion of students are able to answer comprehension questions after reading grade level text (within Kenya's curricular goals) by the end of Primary 1 and Primary 2?
3. What school-level and institutional factors influence reading outcomes when implementing at scale, and how?
4. What community-level factors influence reading outcomes when implementing at scale, and how?
5. To what extent have the Tusome Early Grade Reading Activity components been implemented in schools nationwide?
6. To what extent can any incremental changes in early grade reading outcomes throughout Kenya be correlated with or attributed to the scale-up of Tusome?

Methodology

Intermediate and advanced intervention groups must be drawn from the same schools and in the same manner and numbers as the pre-intervention group for the midline and endline surveys. Reading fluencies and comprehension levels of these two groups will be compared to the pre-intervention group to determine if any significant change has occurred (Research question 1, 2, and 6). School-level contextual data, classroom instructional practices, and Curriculum Support Officers (CSOs) and education official interviews at the midline will be compared to those at the baseline to document project implementation and account for any changes in enrollment, staffing, facilities, instructional resources, or relevant policies and procedures (Research question 3, 5 and 6).

Community-level contextual factors at the midline must be compared to the baseline to account for any changes in the broader educational and socioeconomic conditions of the communities served by each sampled school and the households in which each student lives (Research question 4 and 6). Multivariate comparisons within and between the pre-intervention and intermediate intervention groups (midline) and within and between the pre-intervention, and intermediate intervention, must be conducted to make correlations between the degree of

[6] As stated earlier, experimental or quasi-experimental methods are not possible at this time because the intervention will be implemented nation-wide in 2014 and thus will not include control schools.

project implementation, school-level contextual factors, and community-level contextual factors with any changes in early grade reading outcomes.

Note on data collection time periods: Data collection must be scheduled to be completed by October 20, 2016, due to academic testing at school year end which will impede data collection.

Data Collection Methods

Seven different data collection methods must be used throughout this evaluation. Except for the desk review, each is designed to collect data from individual children, education officials, the school context, or the community context. The very same set of tools must be used for the baseline, midline and endline surveys.

Table 1: Data collection methods for the five evaluation questions

Data Collection Methods	Evaluation Questions
Desk Review	1, 2, 3, 4, 5, 6
Early Grade Reading Assessment (EGRA)	1, 2, 6
Classroom Observation	3, 5, 6
Head Teacher Interview	3, 5, 6
School Data Protocol	3, 6
Household Survey	4, 6
TAC tutor/County/National Education Official[7] Interview	3, 5

Data Analysis Methods

A number of analysis methods must be employed to check the validity and consistency of the data collected, describe the population from whom data was collected, examine the influence of certain variables on others, and track changes in variables over time. All methods must include in the least disaggregation by gender and geography.

Table 2: Data analysis methods for the five evaluation questions

Data Analysis Methods	Evaluation Questions
Pearson Correlation/Cronbach's alpha	1, 2
Frequency distributions/Cross-tabulations	1, 2, 3, 4, 5, 6
Descriptive statistics	1, 2, 3, 4, 5, 6
Multivariate analysis	1, 2, 3, 4, 5, 6
Trend Analysis	1, 2, 3, 4, 5, 6

C.2.5. Baseline Data Required

[7] This will also include ministers where possible.

An EGRA at the nationwide sample of schools used in the Tusome baseline evaluation, plus associated contextual information, is required. The Tusome midline evaluation must be appropriately linked to the Tusome baseline.

C.2.6. Operating Considerations

Because the implementation area is nationwide, the contractor may face security restrictions when seeking to collect midline information in certain areas, including the northeastern counties of Garissa, Mandera, and Wajir.

C.2.7. Participation

GOK staff must be incorporated into the evaluation in collaboration with USAID. The evaluation team must coordinate and work closely with the MoEST Tusome Technical Team.

C.3. ASSESSMENT SCHEDULE

This evaluation will take place in various primary schools across Kenya. The field work must be completed by mid-October, 2016 due to academic testing at school year end which will impede data collection.

The assessment schedule will be as follows on a nationwide scale (disaggregated by county):

	2016	2017
	October	July-October
Standard 1	Midterm	Adjustment*
Standard 2	Midterm	Adjustment*

** Necessary due to timing of revised baseline*

C.4. DELIVERABLES

Specific deliverables are listed under Section F.5.

[END OF SECTION C]

ANNEX II: SAMPLING

Sampling Methodology

Through discussions with USAID, MOE, and RTI, the evaluation team designed and implemented a sampling process during the 2015 baseline to determine the appropriate sample size and select the schools for the baseline. The objective was to produce a sample that would be nationally representative. The process involved six steps:

- Step 1: Define the sampling frame using lists of public and APBET institutions.
- Step 2: Develop a set of design parameters to determine the sample size.
- Step 3: Enter the parameters into sampling software to calculate the sample size.
- Step 4: Select a nationally representative sample of schools equal to the sample size.
- Step 5: Check on the feasibility of the sample and verify the schools in the field.
- Step 6: Replace a limited number of schools (if needed) and finalize the sample.

The sampling frameworks, which were provided by RTI, included 22,154 public schools and 1,000 APBET (Alternative Provision of Basic Education and Training) schools. There was information on school name, administrative units (county, sub-county, and zone), school code, and number of pupils in class 1.

It is important to ensure that the study is sufficiently powered to detect effects. In determining whether the statistical power is sufficient for the study, it is most critical to randomize an adequate number of groups (e.g., schools) – much more so than the number of individuals per group (e.g., pupils)[7]. Values for several parameters (listed below) were assumed to reach a level of minimum detectable effects (MDE) for the study. The MDE is the smallest true effect that has a good chance of having statistical significance. We typically define an MDE as the effect that has 80 percent power for a two-tailed test of statistical significance of 0.05 (alpha level) for all comparisons. A typical MDE target is 0.20 for randomized groups with approximately 10 to 15 individuals per group.

Our parameters below were set using typical values for statistical power and statistical significance, along with the number of counties that would be reasonable to reach within the time and resource constraints of the revised baseline. The design parameters were as follows:

1. Representative set of counties (K = 24 out of 47 total)
2. Number of pupils per class per school (n = 12)
3. Statistical power set to 0.80
4. Alpha (statistical significance) level set to 0.05
5. Intra-class correlation (rho) set at 0.23 (from the RTI PRIMR pilot results)

Based on these design parameters, the MSI statistician used Optimal Design software to calculate the number of schools for the sample. We found that an average of 8.5 schools for each of the 24 clusters

[7] Bloom, H. (2007). *Sample design for group-randomized trials*. Prepared for the U.S. Institute of Educational Sciences/National Center for Educational Research (IES/NCER) Summer Research Training Institute.

(counties) would result in an MDE = 0.20. This led to a total sample size of 204 schools in Kenya for the EGRA baseline, i.e., 8.5 x 24 = 204 schools, with 12 pupils per class per school. Out of the 204 schools, 174 were public schools and 30 were APBET institutions. Based on a desire for more representation in some of the former provinces, we increased the number of counties (K = 26) for an average of 7.85 schools per county (Table 40).

Using a three-stage cluster sampling procedure with the frameworks, MSI drew random samples. The 204 schools were selected proportionally from each of the sampled counties, with independent samples for public and APBET institutions based on their respective sampling frames. School-level samples were 24 pupils, with 12 (6 boys and 6 girls) in each of Classes 1 and 2. The sampling plan resulted in a target of 4,896 total pupils with 2,448 boys and 2,448 girls, along with two teachers and the head teacher from each school.

Table 40: Sampling Stages and Targets

Stage	Procedure
Stage 1	26 sample counties (out of 47 counties in all 8 former provinces)
Stage 2	204 sample schools (174 public and 30 APBET out of out of 22,154 and 1,000 respectively)
Stage 3	12 sample pupils per class (6 boys and 6 girls in each of Classes 1 and 2)

Actual Sample

Table 41 shows the number of pupils assessed by gender and class at baseline and midline. Also provided is the percentage of the sampling target that was reached. All pupils took both the English and Kiswahili subtasks.

Out of the total of 4,671 pupils assessed at midline, 51 percent were boys and 49 percent were girls. In total, the baseline reached 95 percent of the target number of pupils overall: 97 percent of the target for boys and 94 percent of the target for girls. The midline reached a lower number of pupils than the baseline due to fluctuations in enrollment, though still a sufficiently high number to meet sampling requirements.

Table 41: Pupil Sample by Class and Gender, Baseline and Midline

Class	Sample	Baseline			Midline		
		Male	Female	Total	Male	Female	Total
Class 1	Pupils	1,225	1,202	2,427	1,183	1,144	2,327
	% of Target	100.1%	98.2%	99.1%	96.7%	93.5%	95.1%
Class 2	Pupils	1,226	1,213	2,439	1,183	1,161	2,344
	% of Target	100.2%	99.1%	99.6%	96.7%	95.8%	95.8%
Total	Pupils	2,451	2,415	4,866	2,366	2,305	4,671
	% of Target	100.1%	98.7%	99.4%	96.7%	94.2%	95.4%

For the following tables, the teacher, head teacher and CSO breakdowns by gender do not add to the total, as some respondents did not answer this question.

Table 42 shows the number of teachers surveyed by gender and class at baseline and midline, as well as the percentage of the sampling target that was reached. In all, the evaluation team reached 94 percent of its target at both baseline and midline.

Table 42: Teacher Samples by Class and Gender, Baseline and Midline

Gender	Baseline			Midline		
	Class 1	Class 2	Total	Class 1	Class 2	Total
Male	42	36	196	39	33	76
Female	154	152	188	150	151	312
Total	196	188	384	193	189	382
% of Target	96.1%	92.2%	94.1%	94.6%	92.6%	93.6%

Table 43 shows the number of head teachers surveyed by gender at baseline and midline. The evaluation team reached 98 percent of its target at both midline and endline.

Table 43: Head Teachers Samples by Gender, Baseline and Midline

Gender	Baseline	Midline
Male	151	149
Female	48	44
Total	199	200
% of Target	*97.5%*	*98.0%*

Table 44 shows the number of CSOs surveyed by gender at midline. Note that there were no targets for this group, as CSOs were assigned to multiple schools. Additionally, CSOs were not surveyed at baseline.

Table 44: CSO Samples by Gender, Midline

Gender	Midline
Male	81
Female	42
Total	**130**

Table 45 shows the number of households interviewed. These households represent 49 percent of the pupil sample and were balanced between male and female pupils.

Table 45: Number of Households Interviewed

Class	Sample	Midline		
		Male	Female	Total
Class 1	Pupils	607	574	1181
	% of Target	*49.6%*	*46.9%*	*48.0%*
Class 2	Pupils	597	635	1232
	% of Target	*48.8%*	*51.9%*	*48.8%*
Total	Pupils	1,204	1,209	2,413
	% of Target	*49.2%*	*49.4%*	*49.3%*

During the school data collection, the evaluation team worked with head teachers to collect contact information for the pupils selected for the EGRA assessment. Out of the 4,671 pupils assessed, the team collected phone numbers of one or more parents or guardians for 4,027 pupils. Reasons for not collecting phone numbers included that the head teachers did not have them on record, incomplete numbers on record, or the school not wanting to provide the information.

To conduct the household survey, the evaluation team made three attempts at various times of day and days of the week to contact each pupil's parent or guardian. As Table 46 shows, 45 percent of the total calls made were successful. The top reason for not reaching a household was that the phone was out of reach, meaning the number was valid but switched off at the time of the call.

Table 46: Call Results Per Attempt

Call Status	Call 1	Call 2	Call 3	Total Calls	Percent
Successful	1,948	396	165	2,509	45.2%
Out of reach	1,142	141	391	1,674	30.2%
No response	347	84	104	535	9.6%
Wrong numbers	206	41	46	293	5.3%
Call back	115	22	9	146	2.6%
Out of service	82	15	44	141	2.5%
Network problem	70	1	19	90	1.6%
Language barrier	50	4	10	64	1.2%
Refusals	36	7	11	54	1.0%
Incomplete numbers	15	0	0	15	0.3%
Number busy	12	5	3	20	0.4%
Incomplete interviews	6	0	0	6	0.1%
Total Calls	4,029	716	802	5,547	100%

In order to understand the potential differences between the households that were reached and those that were not, the evaluation team looked at two statistics: the differences of the two groups in terms of ORF and student-reported SES. These differences should be taken into consideration when interpreting the household survey results.

Table 47 shows both the English and Kiswahili ORF scores are higher for the pupils from households that were included in the household survey and those that were not. For English, the pupils from households successfully contacted for the survey scored 5.4 CWPM higher in Class 1 and 6.2 CWPM higher in Class 2 than those whose households were not interviewed. For Kiswahili, these pupils scored 5.4 CWPM higher in Class 1 and 6.2 CWPM higher in Class 2. These differences were statistically significant for Class 1, but not Class 2, pupils.

Table 47: Differences in Pupil ORF for Households Reached and not Reached

Task	Class 1			Class 2		
	Household not reached	Household Reached	Difference	Household not reached	Household Reached	Difference
English ORF	19.7	25.1	5.4*	40.6	46.8	6.2
Kiswahili ORF	10.7	13.9	3.2*	23.2	25.9	2.7

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 48 shows that there was little difference in the pupil-reported SES measures for pupils from households that were included in the household survey and those that were not. None of the differences in these groups were statistically significant.

Table 48: Difference in Pupil-Reported SES for Households Reached and not Reached

Task	Class 1			Class 2		
	Household not reached	Household Reached	Difference	Household not reached	Household Reached	Difference
SES 0-1	8.4%	11.7%	3.3%	8.6%	8.1%	-0.5%
SES 2-3	35.9%	35.0%	-0.9%	31.4%	33.5%	2.1%
SES 4-6	41.4%	40.2%	-1.2%	46.5%	45.7%	-0.8%
SES 7-9	12.8%	11.8%	-0.1%	12.2%	11.0%	-1.2%
SES 10-11	1.5%	1.3%	-0.2%	1.3%	1.8%	0.5%

Note: The asterisk indicates a statistically significant difference at the p<.01 level

ANNEX III: DATA COLLECTION INSTRUMENTS

Description of the EGRA Subtasks

NOTE: As the same EGRA tool will be used for the endline, the tool is not included in this report to ensure test security through the endline. The tool is available by request from USAID/Kenya and East Africa.

Some of the subtasks were administered in both languages and others in either English or Kiswahili. All the subtasks are briefly described below, with information on the possible number correct per subtask. The total numbers of subtasks were eight in English and six in Kiswahili.

English and Kiswahili[8]

The **letter sound knowledge subtask** measures a pupil's ability to identify the sounds of written letters. Pupils are given one minute to identify 100 letter sounds. It is measured as the number correct out of 100 letter sounds.

The **invented/non-word decoding subtask** measures a pupil's ability to pronounce (read) unfamiliar written words. Pupils are given 50 non-words to read within one minute. It is measured as the number correct out of 50 words.

The **passage reading subtask** measures oral reading fluency (ORF), i.e., the ability to read text with accuracy and speed. Pupils are given a short passage (60 to 70 words) to read within one minute. ORF is calculated as the number of correct words read per minute (CWPM).

The **oral reading comprehension** subtask measures a pupil's ability to answer comprehension questions based on a story they have just read. Pupils are asked up to six comprehension questions. It is calculated as the number correct out of six questions.

English Only

The **phoneme segmentation subtask** measures a pupil's ability to identify individual phonemes (sounds) in spoken words. Pupils are given ten words, one after the other, and are asked to say the sounds they hear in the word: e.g. cat = /k//a//t/. It is measured as the number correct out of ten words.

The **vocabulary subtask** measures a pupil's ability to understand the meaning of common spoken words. It is measured as the number correct out of 20 vocabulary items.

Kiswahili Only

The **syllable fluency subtask** measures a pupil's ability to identify written syllables. Pupils are given one minute to identify 100 syllables. It is measured as the number correct out of 100 syllables.

The **listening comprehension subtask** measures a pupil's ability to understand a simple story read out loud by the enumerator. Pupils are asked five listening comprehension questions based on the story. It is measured as the number correct out of five questions.

8 Note that, for English, the pupils were administered two sets of reading passages and comprehension questions (A and B). Passage A was traditional in that the pupils had one minute to read the passage aloud, the passage was removed from them, and then they were asked the comprehension questions. For Passage B, the pupils had one minute to read the passage aloud, another minute to read the passage silently, the passage was left in front of them, and then they were asked the comprehension questions. The goal of the second passage was to assess the pupils using a subtask that would reflect a key type of reading instruction on the project. The second set increased the total number of English subtasks to eight.

Pupil Interview

Thank you very much. Now, I am going to ask you some questions about you and your reading habits.
Asante sana. Sasa nitakuuliza maswali kukuhusu na pia kuhusu mtindo wako wa kusoma.

Ask each question verbally to the pupil, as in an interview. Do not read the response options aloud. Wait for the pupil to respond then write this response in the space provided, or check the box of the option that corresponds to the pupil's response. If there is no special instruction to the contrary, only one response is permitted.

1.	What language does your family speak at home? Familia yako huongea lugha gani nyumbani? *(Tick all that apply)*	☐ (1) Kiswahili ☐ (2) English ☐ (-7) Other ☐ (-8) Don't know/No Answer
2.	What language do you speak at school? Wewe huongea lugha gani ukiwa shuleni? *(Tick all that apply)*	☐ (1) Kiswahili ☐ (2) English ☐ (-7) Other ☐ (-8) Don't know/No Answer
3.	*(If in Class 1)*, Did you go to school before Class 1 (nursery, pre-unit, baby class)? *(Iwapo mwanafunzi ni wa darasa la kwanza)* Ulienda shule yoyote kabla ya kuanza darasa la kwanza? *(Shule ya chekechea)* *(For Class 2 pupils mark "not applicable")*	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer ☐ (-9) Not applicable
4.	Do you have English books or other English reading materials at your home? Una vitabu vya Kiingereza vya kusoma nyumbani?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
5.	Do you have Kiswahili books or other Kiswahili reading materials at your home? Una vitabu vya Kiswahili vya kusoma nyumbani?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
6.	Does anyone read stories aloud to you at your home? Kuna mtu yeyote nyumbani kwenu ambaye hukusomea hadithi kwa sauti?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
7.	Do you read stories at your home? Wewe husoma hadithi nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
8.	Do you ever practice reading aloud to your teacher or to other pupils? Wewe hufanya mazoezi ya kusoma kwa sauti kwa mwalimu au kwa wanafunzi wengine?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
9.	Do you practice silent reading in school? Wewe hufanya mazoezi ya kusoma kimya shuleni?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer

10.	Does your teacher assign reading for you to do at your home? Je, mwalimu wako hukupa mazoezi ya kusoma ukiwa nyumbani?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
11.	Do you have a lamp at home? Kuna taa nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
12.	Is there electricity in your house? Kuna stima nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
13.	Do you watch TV at your home? Je, wewe huangalia TV nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
14.	Do you listen to the radio at your home? Je, wewe husikiliza redio nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
15.	Is there a mobile phone or telephone at your home? Kuna simu ya mkono (mobile) au simu nyingine nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
16.	Is there a computer or laptop or tablet or iPad at your home? Kuna kompyuta (laptop/tablet/ipad) nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
17.	Is there a bicycle at your house? Kuna baiskeli nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
18.	Is there a motorcycle at your home? Kuna pikipiki nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
19.	Do you have many cows or goats or camels or donkeys or sheep at your home, such as more than 10? Kuna ng'ombe au mbuzi au ngamia au punda au kondoo wengi nymbani kwenu kama zadi ya kumi?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
20.	Is there a car or truck or tractor or boat at your home? Kuna gari au lori au trakta au boat/boti nyumbani kwenu?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer

Thank you very much for your participation!
Asante sana kwa kushiriki kwako!

Teacher Interview

We are conducting a study so we can understand how children learn to read. Your school was selected to participate in this study through a random process. We would like your help in this, but you do not have to take part if you do not want to.

- **Your name will not be recorded on this form, nor mentioned anywhere in the survey data or reports.**
- **The name of your school and the class level teach will be recorded, but only so that we can correctly link school, class, and pupil data so as to analyze relationships between children's learning and the characteristics of the settings in which they learn. Your school's name will not be used in any report or presentation. Results of this study will be presented to the MOE and its partners. Your school's name will not be used in any report or presentation.**
- **If you agree to help with this study, then please complete this interview which contains questions regarding your daily activities at school, including your interactions with staff in the school, administrators, pupils and parents.**
- **Do you give your consent?**

Check box if verbal consent is obtained: ☐ ***YES***

A. Date of Assessment:	Day: ☐☐ Month: ☐☐	I. Class:	☐ (1) Class 1 ☐ (2) Class 2
B. Region:		J. Stream Name:	
C. County:		K. Gender:	☐ (0) Male ☐ (1) Female
D. Sub-county:		L. Number of pupils in class:	☐ Boys ☐ Girls
E. Zone/Cluster:		M. Age at last birthday	☐☐☐ Years
F. School Unique Code:	☐☐☐☐☐☐☐☐	N. Enumerator's ID:	☐☐☐
G. School Name:		O. Enumerator's Name and Signature:	
H. School Shift:	☐ (0) Full Day ☐ (1) Morning Only ☐ (2) Afternoon Only		

1.	Which subjects do you teach in this class? *(Mark all that apply)*	☐ (1) Kiswahili ☐ (2) English ☐ (-7) Other
2.	What is your highest professional qualification?	☐ (0) Untrained ☐ (1) P1 ☐ (2) Diploma/S1 ☐ (3) Bachelor of Education ☐ (4) Masters ☐ (-7) Other
3.	For how many years have you been teaching?	I___I___I years
4.	Which classes are you teaching this year: *(Mark all that apply)*	☐ (1) Class 1 ☐ (2) Class 2 ☐ (3) Class 3 ☐ (4) Class 4 ☐ (-7) Other
5.	Do you teach in multi-grade classes?	☐ (1) Yes ☐ (2) No ☐ (-8) Don't know/No Answer
6.	Does your school or classroom have a functioning library?	☐ (0) No → Skip to 9 ☐ (1) Only in school ☐ (2) Only in classroom ☐ (3) In both
7.	Do your pupils visit the library?	☐ (1) Yes ☐ (2) No
8.	Do your pupils borrow books from the library to read at home?	☐ (1) Yes ☐ (2) No
9.	Do you use books, other than textbooks, in your classroom?	☐ (1) Yes ☐ (2) No
10.	Do you give extra time or remediation to weak/deserving pupils?	☐ (1) Yes ☐ (2) No
11.	Over the last 12 months, how often has a head teacher observed you teaching in your classroom?	☐ (0) Never ☐ (1) About once a week ☐ (2) About once per month ☐ (3) About once per term ☐ (4) About once per year ☐ (-8) Don't know/No Answer
12.	Over the last 12 months, how often has a Curriculum Support Officer (coach for APBET) observed you teaching in your classroom?	☐ (0) Never ☐ (1) About once a week ☐ (2) About once per month ☐ (3) About once per term ☐ (4) About once per year ☐ (-8) Don't know/No Answer

13.	Do you have teacher guides for teaching reading for English?	☐ (1) Yes ☐ (2) No
14.	Do you have teacher guides for teaching reading for Kiswahili?	☐ (1) Yes ☐ (2) No
15.	Which method do you use <u>most often</u> to measure your pupils' progress during your classroom instruction of reading?	☐ (0) No measure used ☐ (1) Written assessment ☐ (2) Oral assessment ☐ (3) Check exercise books ☐ (4) Check homework
16.	How do you measure pupil achievement in reading at the end of the academic year?	☐ (0) I do not do a measurement ☐ (1) Oral test ☐ (2) Paper and pencil test
17.	Other than Tusome training, how many times did you receive in-service training in the past 2 years?	☐ (0) None → Skip to 19 ☐ (1) 1-2 sessions ☐ (2) 3-4 sessions ☐ (3) 5-6 sessions ☐ (4) More than 6 sessions
18.	Did you learn how to teach reading during these non-Tusome training(s)?	☐ (1) Yes ☐ (2) No
19.	How many Tusome teacher training sessions have you attended so far?	☐ (0) None → Skip to 23 ☐ (1) 1-2 sessions ☐ (2) 3-4 sessions ☐ (3) 5-6 sessions ☐ (4) More than 6 sessions
20.	Overall, how do you rate the quality of Tusome teacher training?	☐ (0) Very poor quality ☐ (1) Poor quality ☐ (2) High quality ☐ (3) Very high quality
21.	How would you rate the relevance of Tusome training?	☐ (0) Not relevant ☐ (1) Relevant ☐ (2) Very relevant
22.	How would you rate the usefulness of Tusome materials?	☐ (0) Not useful ☐ (1) A little bit useful ☐ (2) Useful ☐ (3) Very useful

Think about the last 5 days of school and indicate how often each of the following activities took place. Choose only one answer per question.

	Activity / Action	Never	1 day a week	2 days a week	3 days a week	4 days a week	5 days a week
23.	The whole class repeated sentences that you said/read first (choral repetition/reading)	0	1	2	3	4	5
24.	Pupils copied text from the blackboard or textbook	0	1	2	3	4	5

25.	Pupils retold a story that you read to them	0	1	2	3	4	5
26.	Pupils retold a story that they read	0	1	2	3	4	5
27.	Pupils sounded out unfamiliar words	0	1	2	3	4	5
28.	Pupils learned meanings of new words	0	1	2	3	4	5
29.	Pupils read aloud to teacher or to other pupils	0	1	2	3	4	5
30.	Pupils answered comprehension questions based on the text you read to them	0	1	2	3	4	5
31.	Pupils answered comprehension questions based on the text they read	0	1	2	3	4	5
32.	Pupils were assigned reading to do on their own during school time	0	1	2	3	4	5
33.	Pupils were assigned reading to do on their own at home	0	1	2	3	4	5

In what class should pupils <u>FIRST BE ABLE TO DEMONSTRATE</u> the following skills? Tick only one box for a question.

	Activity / Action	**Before Class 1**	**Class 1**	**Class 2**	**Class 3**	**Class 4 or later**
34.	Recognize letters and say letter names	0	1	2	3	4
35.	Recognize letters and say letter sounds	0	1	2	3	4
36.	Sound out unfamiliar words	0	1	2	3	4
37.	Read aloud a short passage with few mistakes	0	1	2	3	4
38.	Understand stories they read	0	1	2	3	4

Thank you for your participation! You have been very helpful.

Head Teacher Interview

We are conducting a study so we can understand how children learn to read and do math. Your school was selected through a process of statistical sampling. We would like your help in this, but you do not have to take part if you do not want to.

- Your name will not be recorded on this form, nor mentioned anywhere in the survey data or reports.
- The name of your school will be recorded, but only so that we can correctly link school and pupil data so as to analyze relationships between children's learning and the characteristics of the settings in which they learn. Results of this study will be presented to the MOE and its partners. Your school's name will not be used in any report or presentation.
- If you agree to help with this study, then please complete this interview which contains questions regarding your daily activities at school, including your interactions with staff in the school, administrators, pupils and parents.
- Do you give your consent?

Check box if verbal consent is obtained: ☐ *YES*

A. Date of Assessment:	Day: ☐☐ Month: ☐☐	G. School Unique Code:	☐☐☐☐☐☐☐☐
B. Region:		H. School Name:	
C. County:		I. Gender:	☐ (0) Male ☐ (1) Female
D. Subcounty:		J. Total enrolment in Classes 1 and 2:	____ Class 1 ____ Class 2
E. Zone/Cluster:		K. Enumerator's ID:	☐☐☐
F. School Shift:	☐ (0) Full Day ☐ (1) Morning Only ☐ (2) Afternoon Only	L. Enumerator's Name + Signature:	

1.	What is your position at this school?	☐ (1) Head Teacher ☐ (2) Deputy Head Teacher ☐ (-7) Other
2.	How many years have you been in this position?	I___I___I Years
3.	What is your highest professional qualification?	☐ (0) Untrained ☐ (1) P1 ☐ (2) Diploma/S1 ☐ (3) Bachelor of Education ☐ (4) Masters ☐ (-7) Other
4.	Have you received specialized training or taken courses in school management in the past 12 months?	☐ (1) Yes ☐ (2) No
5.	Have you received specialized training or taken courses on reading instruction in the past 12 months?	☐ (1) Yes ☐ (2) No
School instruction, curriculum and assessment		
6.	Does your school timetable include periods for teaching reading skills in English and/or in Kiswahili?	☐ (1) Yes ☐ (2) No → Skip to 8 ☐ (3) There is no timetable → Skip to 8 ☐ (-8) Don't know/No Answer → Skip to 8
7.	In the timetable, how many periods in a week are there for teaching reading skills?	☐ (1) 0 ☐ (2) 1 to 2 ☐ (3) 3 to 4 ☐ (4) 5 or more ☐ (-8) Don't know/No Answer
8.	Have you supported teachers for Classes 1 and/or 2 on how to teach reading skills?	☐ (1) Yes ☐ (2) No
9.	Are you satisfied with the pupils' reading performance at the end of Classes 1 and 2 in your school?	☐ (1) Yes ☐ (2) No
10.	How do you know whether your pupils are progressing during the academic year? *(Mark all that apply)*	☐ (1) Conduct classroom observations ☐ (2) Monitor pupils' results on tests given by teachers ☐ (3) Evaluate pupils myself ☐ (4) Review pupils' assignments or homework ☐ (5) Teachers provide me with progress reports ☐ (-7) Other ☐ (-8) Don't know/No Answer
11.	Who is responsible for reviewing teachers' lesson plans? *(Mark all that apply)*	☐ (0) No one → Skip to 13 ☐ (1) Head teacher ☐ (2) Deputy head teacher ☐ (3) Senior teacher ☐ (-7) Other

12.	How often are these lesson plans reviewed?	☐	(1) Never
		☐	(2) Once per year
		☐	(3) Once every 2-3 months
		☐	(4) Once every month
		☐	(5) Once every two weeks
		☐	(6) Every week
		☐	(7) Once per day
		☐	(-8) Don't know/No Answer
13.	In your school, who is responsible for observing teachers in their classroom? *(Mark all that apply)*	☐	(0) No-one → Skip to 15
		☐	(1) Head teacher
		☐	(2) Deputy head teacher
		☐	(3) Senior teacher
		☐	(4) Curriculum support officer
		☐	(-7) Other
14.	In one term, how often is a teacher observed in their classroom?	☐	(1) Never
		☐	(2) One time
		☐	(3) Two times
		☐	(4) Three or more times
		☐	(-7) Other
		☐	(-8) Don't know/No Answer
15.	Who provides textbooks for English and Kiswahili for Classes 1 and 2? *(Mark all that apply)*	☐	(1) Ministry of Education
		☐	(2) County Government
		☐	(3) School (via independent funds)
		☐	(4) Parents (individually)
		☐	(5) Board of Management
		☐	(6) NGO
		☐	(-7) Other
16.	What language do teachers **use most often** for reading instruction in your school in Classes 1 and 2 for English reading skills?	☐	(1) English
		☐	(2) Kiswahili
		☐	(3) Local Language
		☐	(-7) Other
17.	What language do teachers **use most often** for reading instruction in your school in Classes 1 and 2 for Kiswahili reading skills?	☐	(1) English
		☐	(2) Kiswahili
		☐	(3) Local Language
		☐	(-7) Other
Information about the school			
18.	Does the school have a functioning library?	☐	(1) Yes
		☐	(2) No → Skip to 22
19.	Who can use the library?	☐	(1) Pupils
		☐	(2) Teachers
		☐	(3) Pupils and teachers
20.	Do teachers have a scheduled library time for their classes?	☐	(1) Yes
		☐	(2) No
21.	Are pupils allowed to take library books home?	☐	(1) Yes
		☐	(2) No

22.	How many of the Classes 1 and 2 teachers have received specific training on teaching reading skills?	☐ (1) None of them ☐ (2) Some of them ☐ (3) Most of them ☐ (4) All of them ☐ (-8) Don't know/No Answer
23.	Does your school have a Board of Management (BOM)?	☐ (1) Yes ☐ (2) No → Skip to 26
24.	If yes, how often did the BOM meet during the last school year?	☐ (1) Never ☐ (2) Once per year ☐ (3) Once a term ☐ (4) Once every month ☐ (5) Once every week ☐ (-8) Don't know/No Answer
25.	For which of the following does the BOM have decision making authority and/or responsibility? (Read possible options) *(Mark all that apply)*	☐ (1) Discuss school management problems ☐ (2) Discuss pupils' problems and solutions ☐ (3) Review progress of school improvement efforts ☐ (4) Review financial situation (budgets) of the school ☐ (5) Manage school infrastructure and equipment ☐ (6) Discuss school curriculum ☐ (7) Raise funds ☐ (8) Manage procurement or distribution of textbooks ☐ (-8) Don't know/No Answer
26.	Does the school have electricity?	☐ (1) Yes ☐ (2) No
27.	Does the school have a feeding program?	☐ (1) Yes ☐ (2) No
28.	Does the school have a computer room?	☐ (1) Yes ☐ (2) No

Thank you for your participation! You have been very helpful.

Curriculum Support Officer Interview

The Ministry of Education (MOE), USAID, and Management Systems International (MSI) are collaborating in a study to measure progress in children's reading levels along with associated factors. Schools in your zone were selected through a process of statistical sampling. We would like your help in giving us some information. But you do not have to take part if you do not want to.

- Your name will not be recorded on this form or mentioned anywhere in the survey data. The results of this survey will be published in the form of collective tables. The information acquired through this instrument will be shared with MOE with the hope of identifying areas where additional support may be needed.
- The name of your zone will be recorded, but only so that we can correctly link school, class, student, teacher and head-teacher data so as to analyze relationships between children's learning and the characteristics of the settings in which they learn. The results of analysis will be used by MOE and USAID to help identify additional support that is needed.
- If you agree to participate in this study, I will ask you questions regarding your work including teaching support and other school-related activities. Please answer completely, truthfully and accurately. This interview should take approximately 10 minutes.
- Do you understand and agree to participate in this study?

CONSENT STATEMENT: ☐ *YES*

1	Region	
2	County	
3	Subcounty	
4	Zone	
6	School ID code for all schools covered by CSO in Tusome Midline Study	
7	CSO's gender	Male ……..……0 Female ……..1
8	For how many years have you been a CSO?	Years: _____
9	How many years have you been a CSO in this zone?	Years: _____
10	What was your title before you became a CSO?	Teacher ……...……..1 Deputy Head Teacher ……..2 Head Teacher ……...…….3 Other (specify): ……...…….4

11	What is your highest professional training level in teaching?	Untrained……..……..……0 P1……..……..……..…… 1 Diploma/S1 …..……..…… 2 Bachelor of Education…….…….. 3 Masters……4 Other (specify}: _____ 5
12	How many public schools are in your zone?	Number: _____
13	How many non-formal primary schools are in your zone?	Number: _____
14	During the last 30 days of school, how many days did you visit schools for classroom observations? (Put zero if none)	Number: _____ ***If Zero, Go To Question 17***
15	When you visit a school, what classes do you observe? *(Multiple responses allowed – Circle all that apply)*	Standard 1..... 1 Standard 2..... 2 Other 3
16	During the last 30 days of school, how many actual lessons did you observe?	Lessons: _____
17	During the last 30 days of school, approximately what percentage of your time was spent on instructional support for Tusome?	Percent:_____
18	During the last 2 years, have you received training in providing instructional support in reading at lower primary level for Tusome?	Yes..... 1 No..... 0 ***If no to question 22***
19	If yes, what type of trainings were they?	School-based…… ……1 Off-site in-service …… ……2 Other (specify)_____3
20	If yes, who organized these trainings? *(Multiple responses allowed – Circle all that apply)*	MOE……..……..……1 Tusome (RTI)……..……..2 Other Project/Donor……...……..……..5 Other (specify): _____6
21	If yes, what was the approximate total number of days of all of this type of training you have received?	_____days
22	During the last 2 years, have you personally organized Tusome in-service training for teachers?	Yes..... 1 No.....0
23	How do you assess the teachers' progress in the schools? *(Multiple responses allowed – Check all that apply)*	Check KCPE results ……..……..1 Check zonal term exams results ……..……..……2 Discuss with the head teacher ……..……..……..…….3 Conduct teacher observations ……..……..……..….4 No assessment procedure……...…….5 Other (specify): _____6
24	During the last month, how many times did you have to cancel a lesson observation in order to attend to other duties? (*Enter zero if none cancelled)*	Days: _____

25	How effective do you think the current CSO system has been?	Mostly effective……..……..1 Somewhat effective…….2 A little effective…….…..3 Not effective……...….4
26	How would you describe the current approach to early grade reading in your schools?	Mostly effective……...1 Somewhat effective…….2 A little effective……...3 Not effective……..……4
27	Are there, or have there been, any early grade reading initiatives in your Zone over the last 2 years except for Tusome?	Yes.....1 No.....0 **If no, go to Q 29.**
28	If yes, please list them here:	_____ _____ _____ _____

Use the scale on the right to rate various aspects of Tusome in improving reading in Class 1 and 2 in your schools. *(Repeat scale on right for each item).*		1= Very Bad	2= Bad	3= Good	4=Very Good
29	Quality of Tusome training for CSOs	1	2	3	4
30	Relevance of CSO trainings received on Tusome	1	2	3	4
31	Frequency/duration of CSO trainings	1	2	3	4
32	Quality of teacher trainings by CSOs	1	2	3	4
33	Relevance of teacher trainings by CSOs	1	2	3	4
34	Quality of the content of teachers' guides	1	2	3	4
35	Quality of the content of pupils' books	1	2	3	4
36	Relevance of cluster/zonal Tusome monthly meetings	1	2	3	4
37	Effectiveness of lesson observations by CSOs	1	2	3	4
38	Effectiveness of Tusome support to CSOs for school visits	1	2	3	4
39	Effectiveness of Tusome approach: "I do"; "We do"; "you do"	1	2	3	4
What is your overall rating of the Tusome initiative from the following perspectives? *(Repeat scale on right for each item).*		1= Not enough	2= Almost enough	3= Enough	4= More than enough
40	Number of whole lessons in English each week	1	2	3	4

41	Number of whole lessons in Kiswahili each week	1	2	3	4
42	Time allowed to deliver a whole lesson in English	1	2	3	4
43	Time allowed to deliver a whole lesson in Kiswahili	1	2	3	4
44	Amount of homework for the pupil in English	1	2	3	4
45	Amount of homework for the pupil in Kiswahili	1	2	3	4
46	Overall amount of work for the pupil in a term	1	2	3	4
47	Overall amount of work for the teacher in a term	1	2	3	4
48	Overall amount of work for the CSO in a term	1	2	3	4
49	Overall amount of support from Tusome to the CSO in a year	1	2	3	4

50	How do you keep track of pupils' performance in reading in Class 1 and 2 in your schools? *[Multiple responses allowed – Circle all that apply]*	Observe pupils in the classroom ……..……..……...1 Monitor pupils' results on tests given by teachers…….2 Review children's assignments or homework …….. 3 Collect progress reports from teachers ……..……..4 Do not keep track ……..5 Other (specify): …… 6
51	Are results from external EGRAs (i.e., those not conducted by Tusome) communicated to you?	Yes.....1 No.....0
52	What do you think are the key strengths of Tusome? *(Multiple responses allowed – Circle all that apply)*	Provision of teachers' guide ……1 Provision of pupils' books……..……..…….2 Trainings for teachers……..…….3 Other (specify) ……5
53	What are the main challenges that you have faced in implementing Tusome? *(Multiple responses allowed – Circle all that apply)*	Quantity of work ……1 Time constraints given the number of activities ……..….2 Short notice in communicating information …….3 Transport facilitation to widely spread schools …… 4 Other (specify) ……5

Classroom and School Observation Checklist

I. Classroom Observation Form

Every 3 minutes, span the classroom and check all the activities that you observed and the materials that pupils and teachers are using. Note: Do not use all of the columns if the period is not 48 minutes. Use only those columns that correspond to the amount of minutes for the class period. For instance, **if the period is 30 minutes, only go as far as the column with 30 minutes.**

Minutes	**3**	**6**	**9**	**12**	**15**	**18**	**21**	**24**	**27**	**30**	**33**	**36**	**39**	**42**	**45**	**48**
I. Teacher Focus (one response each period)																
Whole class																
Small group																
One individual																
Not focusing on																
Teacher not																
II. Instructional Content (one response each period)																
Phonological																
Alphabetic principle																
Fluency																
Vocabulary																
Comprehension																
Teacher not																
III. Teacher Action (one response each period)																
Reading																
Writing																
Lecturing/explaining																
Asking questions																
Listening to pupils																
Monitoring pupils																
Giving feedback																
Teacher not																
IV. Pupil Actions (one response each period)																
Choral reading																
Partner reading																
Individual reading																
Silent reading																
Writing																
Listening to the																
Repeating/recitation																
Off task/uninvolved																
V. Materials Used (mark all being used)																
Blackboard																
Pocket Chart																
Letter Cards																
Pupil's Book																
Exercise Books																
Teacher's Guide																
No material used																

II. After-Observation Form

Immediately after the observed class ends, please answer the following questions:

1. Basic Information

A. Date of Assessment	Day: ☐☐ Month: ☐☐
B. School Unique Code	☐☐☐☐☐☐☐☐
C. Enumerator's ID	
D. Enumerator's Name	
E. Class	☐ (1) Class 1 ☐ (2) Class 2
F. Language Lesson	☐ (1) English ☐ (2) Kiswahili
G. Unique ID from interview	☐☐☐☐☐☐☐☐☐☐☐☐

2. Inventory

Indicate if the classroom has the following features and/or items:

Items	**Yes**	**No**
Child-sized tables and chairs		
Timetable on the wall		
Decorations/materials on the walls		
Reading books for the children		
Exercise books for each pupil		
Pencils for each pupil		
Tusome pupil's books for each pupil		
Tusome teacher's guide		

3. Teacher feedback to pupils

Circle the number of the most common type of feedback by the teacher.

4	3	2	1
Gives feedback about correct and incorrect responses in a manner that encourages further effort.	Gives feedback about incorrect responses only, in a manner that encouraged further effort.	Gives feedback about incorrect responses only, in a manner than does not encourage further effort.	Gives no feedback at all.

4. Did the teacher generally praise the pupils when they tried hard and/or gave the correct response?

Yes ☐ No ☐

5. What did the teacher do when a pupil either gave the wrong response or did not respond at all?

Circle the number of the most common type of action by the teacher.

4	3	2	1
Provided remediation and encouraged the pupil to try again.	Provided remediation but then called on another pupil or otherwise moved on.	Ignored the error and then called on another pupil or otherwise moved on.	Criticized the pupil and then called on another pupil or otherwise moved on.

6. Did the teacher generally follow the direct instructional model (I do, We do, You do) throughout the lesson?

Yes ☐ No ☐

7. Did the teacher start and end the class as scheduled (i.e., on time and/or according to the timetable)?

Yes ☐ No ☐

8. Did the teacher generally use a pre-prepared lesson plan?

Yes ☐ No ☐

9. Did the teacher generally use an adequate pace during instruction (i.e., not too fast or too slow)?

Yes ☐ No ☐

10. Language usage (English lesson only)

Circle the number of the most common type of language usage by the teacher.

4	3	2	1
Integrated English and Kiswahili as appropriate, i.e., depending on the level of understanding of the pupils.	Used code switching (English-Kiswahili or vice versa) only when majority of the pupils did not seem to understand.	Communicated in English – even when learners did not seem to understand – and discouraged use of Kiswahili.	Used home language most of the time, with little integration of English and/or Kiswahili.

Thank you.

Household Survey

A. CONSENT

Good Morning/ Afternoon/ Evening, my name is…….

I am calling you from Research Solutions Africa; an independent research firm based in Nairobi. We are currently partnering with a team of researchers studying in Kenya in collaboration with Ministry of Education.

We received your phone number from your child's school (NAME), where we talked to your child (NAME) about her/his reading habits. We now want to talk to you as the parent to understand more about (CHILD'S NAME) family and how this may influence her/his learning.

Before we begin, I need to give you some information so you can decide if you want to participate in our study, which will take about 15 minutes.

Your name and contact information will be strictly confidential and we will not provide this information to anyone other than the research team.

You may ask questions at any time throughout our interview. Please know that your participation is completely voluntary.

Do you have any questions?

May I begin the interview?

1. Yes CONTINUE
2. No TERMINATE

B. BACKGROUND

B1. Survey ID	
B2. Respondent's name	
B3. Gender of the respondent	1. Male 2. Female
B4. Respondent's telephone contact	
B5. Region	
B6. County	
B7. School code	
B8. Student code	
B9. Survey language	1. Kiswahili 2. English
B10. Interviewer's Name	
B11. Date of interview	
B12. Time of interview	

C. BASIC INFORMATION

C1. In relation to student NAME, are you	1. Mother 2. Father 3. Main adult responsible for NAME 4. Neighbor GO TO C2
C2. Our survey focuses on the parents/ main adult responsible for NAME. is NAMEs parents/ guardian available to speak?	1. Yes (phone is handed over and ask questions in section D) 2. No (ask when either of them will be available to speak and note the date and time for call back)

D. FAMILY CHARACTERISTICS

D1. How many adults and children live in your household? Ask for usual household members	1. Adults …… 2. Children……...
D2. What is the highest level of education attained by NAME's mother? DO NOT READ OUT THE OPTIONS	1. No formal schooling 2. Started but did not finish primary 3. Primary 4. Started but did not finish secondary 5. Secondary 6. Diploma/ Certificate 7. Undergraduate 8. Post graduate 9. Don't know/ No response
D3. What is the highest level of education attained by NAME's father? DO NOT READ OUT THE OPTIONS	1. No formal schooling 2. Started but did not finish primary 3. Primary 4. Started but did not finish secondary 5. Secondary 6. Diploma/ Certificate 7. Undergraduate 8. Post graduate 9. Don't know/ No response
D4. [only ask if interviewee is a caretaker, and neither the mother or father – response option 3 on C2.] **What is your highest level of education? DO NOT READ OUT THE OPTIONS**	1. No formal schooling 2. Started but did not finish primary 3. Primary 4. Started but did not finish secondary 5. Secondary 6. Diploma/ Certificate 7. Undergraduate 8. Post graduate 9. Don't know/ No response
D5. Do you earn an income?	1. Yes 2. No GO TO D6

D6. What is your primary source of income? DO NOT READ OUT THE OPTIONS	1. Professional/ technical/ managerial 2. Clerical 3. Sales and service 4. Skilled manual labor 5. Unskilled manual labor 6. Domestic service 7. Agriculture 8. Other - specify 9. Don't know / no response
D7. Do any other adults in your household earn an income?	1. Yes 2. No GO TO D8
D8. What is their primary source of income? DO NOT READ OUT THE OPTIONS	1. Professional/ technical/ managerial 2. Clerical 3. Sales and service 4. Skilled manual labor 5. Unskilled manual labor 6. Domestic service 7. Agriculture 8. Other - specify 9. Don't know / no response

E. READING

E1. Does NAME currently have access to English books or other English reading materials at your home?	1. Yes 2. No
E2. Does NAME currently have access to Kiswahili books or other Kiswahili reading materials at your home?	1. Yes 2. No
E3. Does NAME currently have access to books or reading materials in any other language at your home?	1. Yes 2. No
E4. Does NAME ever read at home?	1. Yes 2. No **GO TO E6**
E5. How often does NAME read at home? DO NOT READ OUT THE OPTIONS	1. At least once per day 2. A few times per week 3. Once a week 4. Less than once a week
E6. Do you or anyone else in your household read aloud to NAME?	1. Yes 2. No **GO TO SECTION F**
E7. How often do you read aloud to NAME? DO NOT READ OUT THE OPTIONS	1. At least once per day 2. A few times per week 3. Once a week 4. Less than once a week

F. ASSETS

F1. Does your household have the following? READ OUT OPTIONS

	1=Yes	2=No
1.Electricity	1	2
2.A Radio	1	2
3.A Television	1	2
4.A mobile phone	1	2
5.A refrigerator	1	2
6.A solar panel	1	2
7.A table	1	2
8.A chair	1	2
9.A sofa	1	2
10.A bed	1	2
11.A cupboard	1	2
12.A microwave oven	1	2
13.A DVD player	1	2
14.A cassette or CD player	1	2

F2. Do you or any members of your household own the following? READ OUT OPTIONS

	1=Yes	2=No
1.A watch	1	2
2.A bicycle	1	2
3.A motorcycle or motor scooter	1	2
4.An animal-drawn cart	1	2
5.A car or truck	1	2
6.A boat with a motor	1	2

G. HOUSING SERVICES

G1. What is the MAIN source of drinking water for members of your household?	1. Piped water 2. Tube well or borehole 3. Dug well 4. Water from spring 5. Rain water 6. Tanker truck 7. Cart with small tank 8. Surface water (River/Dam/Lake/Pond/stream/Canal/Irrigation Channel) 9. Other 10. Don't know / no response

G2. What kind of toilet does your household have?	1. Flush or pour flush toilet 2. Pit latrine 3. Composting toilet 4. Bucket toilet 5. No facility/ bush/ field 6. Other 7. Don't know / no response
G3. What type of fuel does your household MAINLY use for cooking?	1. Electricity 2. LPG/Natural gas 3. Biogas 4. Paraffin/ Kerosene 5. Coal, lignite 6. Charcoal 7. Wood 8. Straw/ shrubs/ grass 9. Agricultural crop 10. Animal dung 11. No food cooked in household 12. Other 13. Don't know / no response

H. HOUSE CONSTRUCTION

H1. How many rooms in your household are used for sleeping?

Number of rooms……

SURVEY END

Thank you for your time and participating in this study.

ANNEX IV: SOURCES OF INFORMATION

School Data Collection

The evaluation team collected data from 204 schools across all 8 former provinces. At each school, the evaluation team assessed up to 12 Class 1 and 12 Class 2 pupils. The team then interviewed the associated Class 1 and Class 2 teachers, the school's head teacher, and the CSO assigned to the school. The team also observed up to two reading lessons per school depending on the timetable. The distribution of schools by district is listed in Table 49.

Table 49: Schools by County

County	Schools
Bomet	6
Bungoma	8
Garissa	8
Homa Bay	8
Kajiado	5
Kiambu	8
Kilifi	8
Kirinyaga	8
Kisii	8
Kisumu	5
Kitui	8
Makueni	7
Marsabit	6
Meru	6
Mombasa	13
Nairobi	27
Nakuru	5
Nandi	6
Narok	6
Nyandarua	8
Siaya	8
Taita Taveta	8
Trans Nzoia	5
Uasin Gishu	5
Vihiga	7
Wajir	7
Total	**204**

Household Data Collection

After randomly selecting the pupils for the assessment, the evaluation team worked with the head teacher to collect contact information for the pupils' parents and/or guardians. Data was collected over the phone.

Literature Review

The evaluation team used the following documents which were provided by USAID as attachments to the solicitation for evaluation findings.

- J.1. Annual Report (FY 2015)
- J.2. Tusome Statement of Work (Includes ALL Modifications)
- J.3. Tusome Work Plan (2015)
- J.7. Performance Monitoring Plan (2014.12.22) - Aligned to CDCS
- J.8. Primary Math and Reading (PRIMR) - Education Policy Study Report 1
- J.9. Primary Math and Reading (PRIMR) - Education Policy Study Report 2
- J.10. Primary Math and Reading (PRIMR) - Education Policy Study Report 3

The evaluation team also reviewed the following documents.

- Piloting Report – Kenya Tusome Baseline Study (2015)
- Tusome Revised Baseline Study (January 2016)
- Ministry of Education, Science and Technology: National Education Sector Plan, Volume I: Basic Education Programme Rationale and Approach. (2015)
- Uwezo: Are Our Children Learning? Uwezo Kenya Sixth Learning Assessment Report. (2016)

ANNEX V: EGRA RESULTS

The EGRA included 14 subtasks, eight in English and six in Kiswahili. This annex includes the raw scores for all of the 14 EGRA subtasks at baseline and midline, as well as disaggregations by school type and gender.

English EGRA Results

Table 50: English Reading Scores

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Phoneme segmentation	1.1	3.8	2.6*	0.6	5.0	4.5*
Letter sound knowledge	15.1	26.3	11.3*	10.2	32.6	22.4*
Invented/non-word decoding	5.7	10.4	4.7*	10.4	18.6	8.3*
Vocabulary	5.9	7.8	1.9*	8.2	10.2	1.9*
Passage reading (A)	10.6	22.3	11.7*	23.8	43.6	19.9*
Reading comprehension (A)	0.2	0.5	0.3*	0.5	1.0	0.5*
Passage reading (B)	9.7	22.0	12.4*	21.8	44.2	22.5*
Reading comprehension (B)	0.2	0.8	0.6*	0.6	1.7	1.2*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 51: English Class 1 Reading Scores by School Type

Subtask	Public			APBET		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Phoneme segmentation	1.1	3.7	2.7*	4.4	7.2	2.8*
Letter sound knowledge	14.8	26.2	11.4*	31.7	39.1	7.4*
Invented/non-word decoding	5.5	10.2	4.7*	16.6	22.4	5.7*
Vocabulary	5.8	7.7	1.9*	11.8	14.5	2.7*
Passage reading (A)	10.2	21.8	11.7*	38.0	58.2	20.2*
Reading comprehension (A)	0.2	0.5	0.3*	1.5	2.1	0.6
Passage reading (B)	9.3	21.6	12.3*	35.1	58.0	22.9*
Reading comprehension (B)	0.2	0.7	0.6*	1.7	3.0	1.3*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 52: English Class 2 Reading Scores by School Type

Subtask	Public			APBET		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Phoneme segmentation	0.5	5.0	4.5*	3.0	8.0	4.9*
Letter sound knowledge	10.0	32.5	22.5*	23.8	40.7	16.9*
Invented/non-word decoding	10.1	18.5	8.3*	24.6	31.1	6.5*
Vocabulary	8.1	10.1	1.9*	14.5	16.1	1.6
Passage reading (A)	23.2	43.1	20.0*	61.9	81.1	19.2*
Reading comprehension (A)	0.5	1.0	0.5*	2.8	3.2	0.3
Passage reading (B)	21.2	43.7	22.5*	58.2	81.9	23.7*
Reading comprehension (B)	0.5	1.7	1.2*	3.4	4.0	0.5

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 53: English Class 1 Reading Scores by Gender

Subtask	Male			Female		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Phoneme segmentation	1.0	3.5	2.5*	1.2	4.0	2.8*
Letter sound knowledge	14.1	23.9	9.8*	16.0	28.7	12.7*
Invented/non-word decoding	5.1	9.3	4.2*	6.3	11.5	5.2*
Vocabulary	5.9	7.6	1.7*	5.8	8.0	2.1*
Passage reading (A)	9.3	20.1	10.8*	11.9	24.5	12.6*
Reading comprehension (A)	0.2	0.5	0.3*	0.2	0.5	0.3*
Passage reading (B)	8.5	20.0	11.5*	10.8	24.1	13.2*
Reading comprehension (B)	0.2	0.7	0.6*	0.2	0.8	0.6*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 54: English Class 2 Reading Scores by Gender

Subtask	Male			Female		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Phoneme segmentation	0.6	4.9	4.4*	0.5	5.1	4.6*
Letter sound knowledge	9.6	30.5	20.9*	10.8	34.7	23.9*
Invented/non-word decoding	9.6	17.3	7.7*	11.1	20.0	8.8*
Vocabulary	8.1	10.3	2.2*	8.3	10.0	1.6*
Passage reading (A)	21.7	41.0	19.3*	25.9	46.3	20.4*
Reading comprehension (A)	0.5	1.0	0.5*	0.5	1.1	0.5*
Passage reading (B)	20.0	41.4	21.4*	23.5	47.0	23.5*
Reading comprehension (B)	0.6	1.8	1.2*	0.6	1.7	1.1*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Kiswahili EGRA Results

Table 55: Kiswahili Reading Scores

Subtask	Class 1			Class 2		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Letter sound knowledge	16.6	29.7	13.1*	16.2	39.7	23.4*
Syllable fluency	11.0	21.5	10.4*	20.9	37.5	16.6*
Invented/non-word decoding	4.7	8.3	3.6*	10.2	16.1	5.8*
Passage reading	4.9	12.2	7.3*	13.5	24.5	11.0*
Reading comprehension	0.4	0.9	0.5*	1.1	2.0	1.0*
Listening comprehension	1.2	2.0	0.8*	1.9	2.0	0.9*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 56: Kiswahili Class 1 Reading Scores by School Type

Subtask	Public			APBET		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Letter sound knowledge	16.2	29.4	13.2*	39.2	52.2	13.0*
Syllable fluency	10.7	21.2	10.5*	30.8	43.0	12.2*
Invented/non-word decoding	4.5	8.2	3.6*	13.3	18.8	5.5*
Passage reading	4.7	12.0	7.3*	15.7	26.8	11.0*
Reading comprehension	0.3	0.9	0.5*	1.4	2.3	1.0*
Listening comprehension	1.2	2.0	0.8*	2.5	3.0	0.5*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 57: Kiswahili Class 2 Reading Scores by School Type

Subtask	Public			APBET		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Letter sound knowledge	15.9	39.5	23.6*	40.4	55.3	14.9*
Syllable fluency	20.5	37.3	16.7*	41.4	54.2	12.8*
Invented/non-word decoding	10.1	15.9	5.9*	21.5	26.2	4.7*
Passage reading	13.2	24.3	11.1*	29.6	39.0	9.4*
Reading comprehension	1.0	2.0	1.0*	2.6	3.5	0.9*
Listening comprehension	1.9	2.7	0.9*	3.2	3.5	0.4

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 58: Kiswahili Class 1 Reading Scores by Gender

Subtask	Male			Female		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Letter sound knowledge	15.2	27.5	12.4*	18.0	31.9	13.9*
Syllable fluency	10.0	19.7	9.6*	12.0	23.3	11.3*
Invented/non-word decoding	4.1	7.6	3.5*	5.2	9.0	3.8*
Passage reading	4.2	11.0	6.9*	5.7	13.5	7.8*
Reading comprehension	0.3	0.8	0.5*	0.4	1.0	0.6*
Listening comprehension	1.3	2.1	0.8*	1.2	1.9	0.7*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Table 59: Kiswahili Class 2 Reading Scores by Gender

Subtask	Male			Female		
	Baseline	Midline	Difference	Baseline	Midline	Difference
Letter sound knowledge	15.1	37.4	22.3*	17.4	42.0	24.5*
Syllable fluency	19.1	36.9	17.8*	22.6	38.1	15.4*
Invented/non-word decoding	9.4	15.2	5.8*	11.1	16.9	5.8*
Passage reading	12.4	23.3	11.0*	14.6	25.7	11.0*
Reading comprehension	1.0	2.0	1.0*	1.1	2.1	1.0*
Listening comprehension	1.9	2.8	0.9*	1.9	2.7	0.8*

Note: The asterisk indicates a statistically significant difference at the p<.01 level

Histograms of Oral Reading Fluency Results

The histograms below (Figures 26 to 29) show the distributions of ORF scores for English passage reading A and Kiswahili passage reading at baseline and midline. In all the histograms, there are large percentages of scores at the lower end of the distributions and positive skews. The distributions change somewhat from Class 1 to Class 2, with fewer scores at the lower end and slightly less skew. There are more scores at the lower end of the distributions in Kiswahili than in English.

Note that the bars for the histograms contain multiple scores. For instance, the lowest bar for English Class 1 ORF (passage A) contains the zero scores (about 50 percent of the scores) plus other scores from pupils who read from 1 to 9 CWPM (another 13 or 14 percent of the scores).

Figure 26: Oral Reading Fluency Histogram - English Baseline

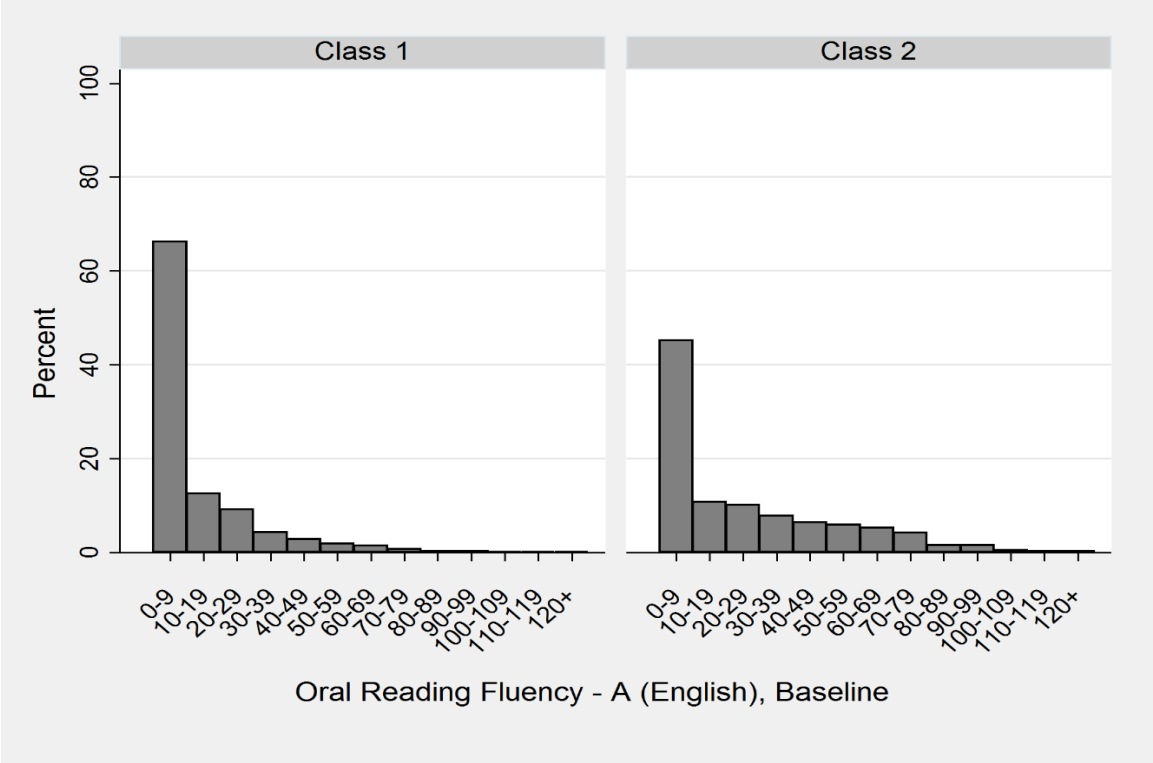


Figure 27: Oral Reading Fluency Histogram - English Midline

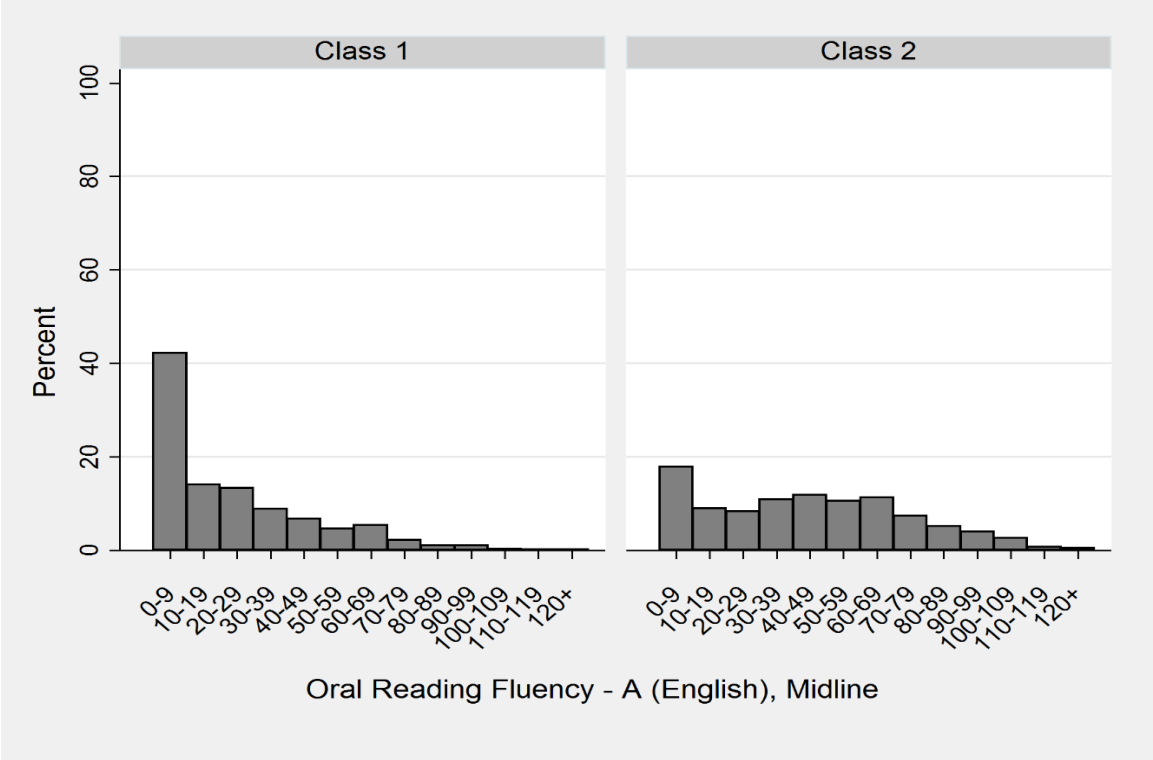


Figure 28: Oral Reading Fluency Histogram - Kiswahili Baseline

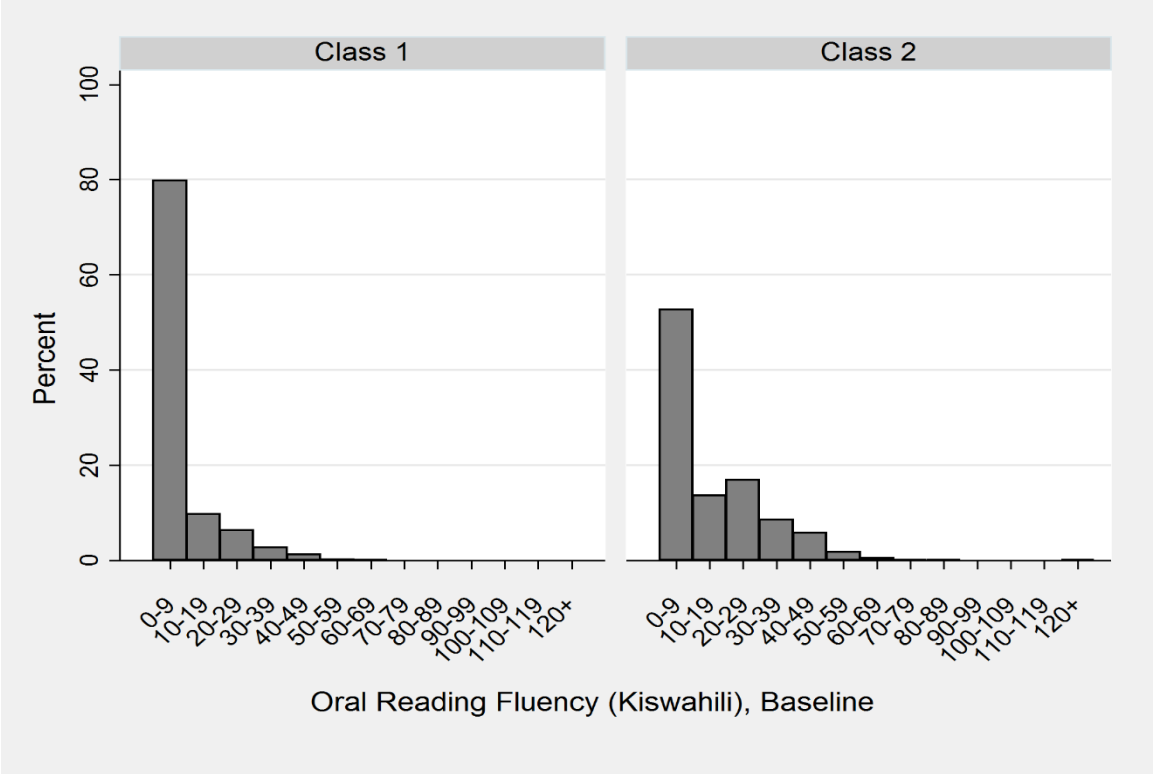
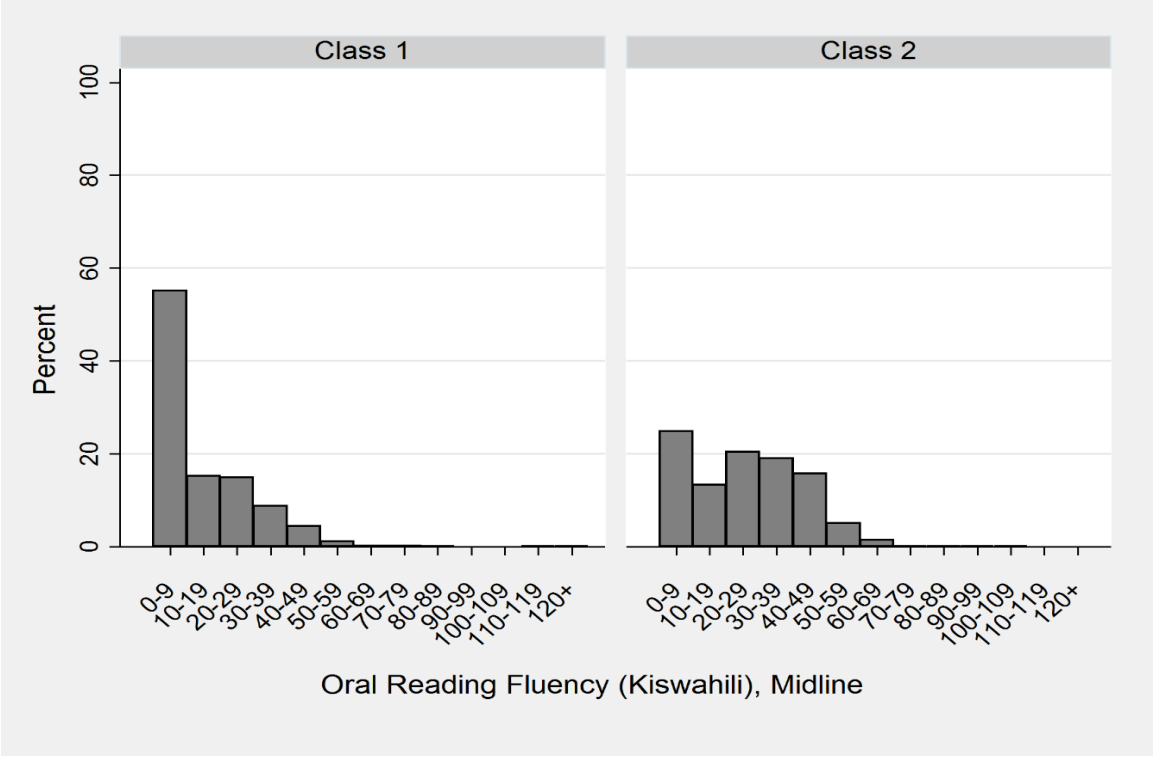


Figure 29: Oral Reading Fluency Histogram - Kiswahili Midline



ANNEX VI: PSYCHOMETRIC ANALYSES

Pearson correlation coefficients were calculated among the subtasks to indicate the consistency of performance by the subtasks on the test. Strong correlations are ideal because they indicate a high degree of consistency. Correlations that are too strong may indicate too much repetition across subtasks.

In addition to the correlations, an item analysis was conducted to determine the psychometric properties (e.g., item difficulty and item-total correlation) of the subtasks. Item difficulty is defined as the percentage of pupils who answered the item correctly. Item-total correlation is defined as the correlation between the correct/wrong scores that pupils received on a given item and the total scores that the pupils received when summing up their item scores. These correlations were corrected so that the given item was removed from the total score when making the calculation to avoid correlating an item with itself. Item difficulties should be between 0.10 and 0.90 and show a range of values within subtasks. Item-total correlation values of 0.20 and above are considered to be psychometrically acceptable.

The psychometric analyses of the subtask correlations and items (untimed only) for each language (English and Kiswahili) and grade level (Classes 1 and 2) are presented in the following sections.

ENGLISH TOOL ANALYSES

Tables 60 and 61 show the Pearson correlation coefficients for the eight subtasks on the English tool for Classes 1 and 2 for the midline. All the correlations were statistically significant and positive ($p < 0.001$). The correlations are moderate to strong across all tasks. For Class 1, the highest correlation (0.97) was between the two passage reading subtasks (A and B), indicating consistent performance in passage reading skills. The two next highest correlations (0.86 and 0.85) were between the passage reading subtasks and non-word decoding, indicating that that the pupils with higher scores in decoding invented/non-words also obtain higher scores in passage reading.

Table 60: English Class 1 Correlation Coefficients

Subtask	1. Phoneme segmentation	2. Letter sound knowledge	3. Non-word decoding	4. Vocabulary	5a. Passage reading (A)	5b. Reading comprehension (A)	6a. Passage reading (B)	6b. Reading comprehension (B)
1. Phoneme segmentation	1							
2. Letter sound knowledge	0.50	1						
3. Invented/non-word decoding	0.43	0.57	1					
4. Vocabulary	0.53	0.49	0.59	1				
5a. Passage reading (A)	0.44	0.57	0.85	0.66	1			
5b. Reading comprehension (A)	0.33	0.34	0.56	0.62	0.69	1		
6a. Passage reading (B)	0.43	0.56	0.86	0.65	0.97	0.67	1	
6b. Reading comprehension (B)	0.37	0.39	0.62	0.65	0.75	0.81	0.76	1

For Class 2, the highest correlation (0.96) was also between the two reading passages. The next highest correlations (0.79 and 0.78), as in Class 1, were between the reading passages and invented/non-word decoding. The correlation (0.79) was equally high between the two reading comprehension subtasks.

Table 61: English Class 2 Correlation Coefficients

Subtask	1. Phoneme segmentation	2. Letter sound knowledge	3. Non-word decoding	4. Vocabulary	5a. Passage reading (A)	5b. Reading comprehension (A)	6a. Passage reading (B)	6b. Reading comprehension (B)
1. Phoneme segmentation	1							
2. Letter sound knowledge	0.45	1						
3. Invented/non-word decoding	0.41	0.45	1					
4. Vocabulary	0.48	0.35	0.53	1				
5a. Passage reading (A)	0.48	0.44	0.79	0.60	1			
5b. Reading comprehension (A)	0.37	0.25	0.49	0.63	0.64	1		
6a. Passage reading (B)	0.48	0.45	0.78	0.59	0.96	0.63	1	
6b. Reading comprehension (B)	0.40	0.29	0.51	0.64	0.70	0.79	0.72	1

Kiswahili Tool Analyses

Tables 62 and 63 show the Pearson correlation coefficients for the six subtasks on the Kiswahili tool for Classes 1 and 2. All the correlations were statistically significant and positive ($p < 0.001$). The correlations are moderate to strong across all tasks. For Class 1, the highest correlation (0.89) was between passage reading and reading comprehension, indicating that the pupils with higher scores in passage reading also obtain higher scores in reading comprehension.

Table 62: Kiswahili Class 1 Correlation Coefficients

Subtask	1. Letter sound knowledge	2. Syllable fluency	3. Non-word decoding	5b. Passage reading	6a. Reading comprehension	6b. Listening comprehension
1. Letter sound knowledge	1					
2. Syllable fluency	0.79	1				
3. Invented/non-word decoding	0.65	0.83	1			
4a. Passage reading	0.68	0.88	0.86	1		
4b. Reading comprehension	0.61	0.79	0.76	0.89	1	
5. Listening comprehension	0.34	0.35	0.29	0.36	0.47	1

For Class 2, the highest correlation (0.86) was also between passage reading and reading comprehension, indicating again that pupils with high scores in passage reading also obtain higher scores in reading comprehension.

Table 63: Kiswahili Class 2 Correlation Coefficients

Subtask	1. Letter sound knowledge	2. Syllable fluency	3. Non-word decoding	5b. Passage reading	6a. Reading compre-hension	6b. Listening compre-hension
1. Letter sound knowledge	1					
2. Syllable fluency	0.71	1				
3. Non-word decoding	0.58	0.81	1			
4a. Passage reading	0.59	0.84	0.84	1		
4b. Reading comprehension	0.52	0.73	0.70	0.86	1	
5. Listening comprehension	0.27	0.28	0.24	0.33	0.49	1

Tables 64 and 65 present the analyses of the untimed items for Classes 1 and 2 in Kiswahili. As with English, only the untimed items – reading comprehension and listening comprehension – were analyzed since the similarity of the timed items within the subtasks would lead to repetition in the statistics. All the Kiswahili items had item-total correlations above the minimum standard of 0.20, indicating acceptable quality (or discrimination) of the items. Most of the correlations were well above the minimum. The item difficulties of the subtasks, with the exceptions of half of the items on Class 1 reading comprehension and one of the items on Class 2 reading comprehension, were between 0.10 and 0.90 and showed a range of values within subtasks.

Table 64: Kiswahili Reading Comprehension Item Statistics

Item	Class 1		Class 2	
	Item Difficulty	Item-Total	Item Difficulty	Item-Total
Q.1	0.28	0.61	0.56	0.62
Q.2	0.17	0.54	0.36	0.56
Q.3	0.10	0.49	0.31	0.58
Q.4	0.04	0.34	0.16	0.47
Q.5	0.02	0.27	0.10	0.41
Q.6	0.00	0.08	0.01	0.21

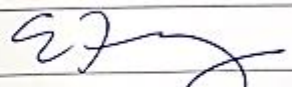
Table 65: Kiswahili Listening Comprehension Item Statistics

Item	Class 1		Class 2	
	Item Difficulty	Item-Total	Item Difficulty	Item-Total
Q.1	0.30	0.37	0.39	0.36
Q.2	0.32	0.46	0.47	0.47
Q.3	0.38	0.41	0.52	0.43
Q.4	0.28	0.43	0.38	0.45
Q.5	0.26	0.39	0.43	0.39

ANNEX VIII: DISCLOSURE OF ANY CONFLICTS OF INTEREST

Name	Elizabeth Freudenberger
Title	Evaluation Team Leader
Organization	Management Systems International
Evaluation Position?	☒ Team Leader ☐ Team member
Evaluation Award Number *(contract or other instrument)*	AID-615-TO-16-00012
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	Tusome external baseline study (2015)
I have real or potential conflicts of interest to disclose.	☐ Yes ☒ No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	29 March 2017

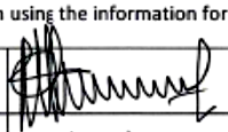
Name	Jeff Davis
Title	Technical Advisor
Organization	Management Systems International
Evaluation Position?	☐ Team Leader ☒ Team member
Evaluation Award Number *(contract or other instrument)*	AID-615-TO-16-00012
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	Tusome external baseline study (2015)
I have real or potential conflicts of interest to disclose.	☐ Yes ☒ No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	29 Mar 2017

Name	Charles Munene Kiura
Title	Local Technical Expert
Organization	Management Systems International
Evaluation Position?	☐ Team Leader ☒ Team member
Evaluation Award Number *(contract or other instrument)*	AID-615-TO-16-00012
USAID Project(s) Evaluated *(Include project name(s), implementer name(s) and award number(s), if applicable)*	Tusome external baseline study (2015)
I have real or potential conflicts of interest to disclose.	☐ Yes ☒ No
If yes answered above, I disclose the following facts: *Real or potential conflicts of interest may include, but are not limited to:* 1. *Close family member who is an employee of the USAID operating unit managing the project(s) being evaluated or the implementing organization(s) whose project(s) are being evaluated.* 2. *Financial interest that is direct, or is significant though indirect, in the implementing organization(s) whose projects are being evaluated or in the outcome of the evaluation.* 3. *Current or previous direct or significant though indirect experience with the project(s) being evaluated, including involvement in the project design or previous iterations of the project.* 4. *Current or previous work experience or seeking employment with the USAID operating unit managing the evaluation or the implementing organization(s) whose project(s) are being evaluated.* 5. *Current or previous work experience with an organization that may be seen as an industry competitor with the implementing organization(s) whose project(s) are being evaluated.* 6. *Preconceived ideas toward individuals, groups, organizations, or objectives of the particular projects and organizations being evaluated that could bias the evaluation.*	

I certify (1) that I have completed this disclosure form fully and to the best of my ability and (2) that I will update this disclosure form promptly if relevant circumstances change. If I gain access to proprietary information of other companies, then I agree to protect their information from unauthorized use or disclosure for as long as it remains proprietary and refrain from using the information for any purpose other than that for which it was furnished.

Signature	
Date	30/03/2017

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, DC 20523