USAID | KYRGYZ REPUBLIC
FROM THE AMERICAN PEOPLE

USAID QUALITY READING PROJECT (QRP): KYRGYZSTAN

Early Grade Reading Assessment (EGRA) Midterm Data Analytic Report

October 09, 2015

A partnership With
American Institutes for Research® (AIR®) and Save the Children International

Contract No.: AID-176-C-13-00001-00

USAID Quality Reading Project: Kyrgyzstan

Early Grade Reading Assessment (EGRA) Midterm Data Analytic Report

Submitted by:
American Institutes for Research®

October 9, 2015

This midterm study of early grade reading assessment is made possible by the support of the American people through the United States Agency for International Development (USAID). The contents are the sole responsibility of American Institutes for Research and Save the Children International and do not necessarily reflect the views of USAID or the United States Government.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AIR	American Institutes for Research
CEATM	Center for Educational Assessment and Teaching Methods
CPD	Continuous Professional Development
EGRA	Early Grade Reading Assessment
IST	In-service Training
KAE	Kyrgyz Academy of Education
MOES	Ministry of Education and Science
NTC	National Testing Center
OECD	Organization for Economic Cooperation and Development
PISA	Programme for International Student Assessment
QRP	Quality Reading Project
TTI	Teacher Training Institutes
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development

ACKNOWLEDGMENTS

The early grade reading assessment (EGRA) has been successfully completed as a result of contributions made by many organizations and individuals. The Ministry of Education and Science (MOES) of the Kyrgyz Republic, National Testing Center (NTC), Pedagogy University, and Kyrgyz Academy of Education (KAE) provided vital guidance and oversight for the whole assessment process. Highly acknowledged also is Mr. Bakirov Artur, Director of NTC, for his constant involvement and follow-up.

This EGRA would have been impossible without the generosity of the American people through the United States Agency for International Development (USAID).

The EGRA was implemented with the technical and logistical assistance of the American Institutes for Research (AIR) and Save the Children International. AIR planned and executed the administration with the support of staff members in both its Washington, DC-based headquarters and the USAID Quality Reading Project office in Kyrgyzstan. Many thanks goes to Mr. Jerome Mindes, Project Manager based in Washington, DC, for his overall role in supporting the activities, to Ms. Barbara Greenwood, Chief of Party, and Ms. Kathryn Fleming, Deputy Chief of Party, Saule Khamzina, Monitoring and Evaluation Manager for technical and administrative leadership, and to all USAID Quality Reading Project staff for their support. Also to Dr. Todd Drummond, Amy Todd, Alvaro Ballabrin-Cabrera, Mauricio Estrada-Matute, and Paul Sirma in the International Development, Evaluation and Research Division at AIR for their collaborative endeavors in producing this Midterm EGRA report.

I. EXECUTIVE SUMMARY

Around the world, the United States Agency for International Development (USAID) continues to invest in improving early grade reading outcomes for children in developing countries. Since 2012, USAID has been actively engaged in the implementation of an Early Grade Reading Assessment (EGRA) in the Kyrgyz Republic to better understand the state of early grade reading in the country. This midterm report presents key findings from the April 2015, midterm EGRA data collection on reading in Grades 2 and 4, Kyrgyz and Russian languages, in four regions in the republic. After a brief introduction to the context of the study, overview of the EGRA subtasks, background on the research design, and methods the study employed, EGRA subtask results are presented across key dimensions of interest.

First, where appropriate, comparison data between EGRA results in 2014 to 2015 are provided to highlight what has changed since 2014. Second, results of several key national indicators of early grade reading are presented. Then, a detailed subtask analysis of the 2015 EGRA is provided along with description of nuances related to the context in the republic and issues in Kyrgyz language reading development. Key dimensions considered for this report include how do EGRA scores differ by gender and school location in the republic? What are the trends in performance on the reading skills subtasks over time across the various groups assessed?

Initial findings indicate that despite some gains in most of the comparable subtasks assessed, the results indicate that more work needs to be done. Approximately 50% of early grade readers reached the 40-words-per-minute benchmark in reading fluency at Grade 2: Fewer met the 80-words-per-minute minimum standard at Grade 4. What has emerged clearly are large score gaps favoring females, often at levels that are statistically significant. Further, the gaps evident at Grade 2 appear to grow wider by Grade 4, especially with the Kyrgyz language cohorts. Difference in performance by school location is also evident. In Russian Grades 2 and 4, as well as Kyrgyz Grade 4, the majority of subtasks show statistically significant differences between urban and rural pupils, with urban pupils being favored in most instances.

The gaps appear to be largest on tasks that require skills and concentration on reading passages, word knowledge, and writing skills (dictations). Several tasks appeared to be relatively easy across all groups (e.g., *Initial Letter Sound*, and the only subtest where males outperformed females in both grades and languages, *Oral Vocabulary).* After presenting the results of EGRA 2015, the midterm report contains an analysis of key findings and poses a series of relevant policy discussion questions to keep the momentum toward growth in reading outcomes in the republic going strong.

II. INTRODUCTION

THE STATE OF READING OUTCOMES IN THE KYRGYZ REPUBLIC

Despite high-literacy attainment in the Soviet period, improving primary reading outcomes has become a new national priority in the Kyrgyz Republic. Recent evidence indicates a steady decline in reading outcomes at multiple grade levels over the last decade. According to a study by the *United Nations Children's Fund*, from 2001 to 2005 results from primary grade literacy assessments showed a 15% decrease in pupils' meeting reading standards with only one half of all pupils meeting basic reading standards (UNICEF, 2005). The results from a 2007 nationally representative assessment at the fourth and eighth grades, the *National Assessment of Educational Quality* (NAEQ), indicated that a majority of pupils at both those levels in the Kyrgyz and Russian language tracks were performing poorly (CEATMb, 2010). USAID's Quality Learning Program (QLP) study in 2010 demonstrated that more than half of all pupils were reading at below grade level. Finally, the Kyrgyz Republics' 2006 and 2009 *Program for International Pupil Assessment* (PISA) scores also revealed that 83% of ear olds were not achieving the minimum PISA reading standards (CEATM, 2010a).

In response to this state of affairs, USAID and the government of the Kyrgyz Republic have intensified their collaboration with the goal of improving reading achievement at the primary levels. With USAID's support, in 2012 the first Early Grade Reading Assessment (EGRA) confirmed previous evidence that pupils were falling behind in the second, third, and fourth grades; especially in the area of reading comprehension (Tvaruzkova & Shamatov, 2012).[1] That first EGRA assessed a nationally representative sample of more than 4,000 pupils and led to a review of teaching practices in the republic. Almost half of the pupils assessed were unable to meet the national standards in reading fluency (measured by the number of words read per minute), the single benchmark of reading skills at that time (ibid, 2012). Though EGRA does not serve as a comprehensive assessment of reading comprehension in the republic, EGRA diagnostic results can serve to focus on the contextual, resource, and pedagogical obstacles to reading development and can stimulate new policy and classroom-level approaches toward improving reading outcomes.

USAID QUALITY READING PROJECT AND EGRA

With support from the USAID Quality Reading Project team, the Ministry of Education and Science (MOES) has taken concrete steps to improve teacher preparation and promote public awareness of the importance of early reading in the republic. In 2013 and 2014 the MOES developed *The Minimum Requirements to Reading in Elementary Schools*, an effort to

[1] The design of the EGRA in the Kyrgyz Republic was influenced by the work done by the 2000 *National Reading Panel*, a U.S. committee of leading scholars that undertook one of the most comprehensive meta-reviews of the reading and literacy literature. *RTI International*, the organization that developed the foundations and structure of the EGRA for use as an assessment template in developing countries, has also influenced EGRA rollout in many contexts around the world. See Gove (2009) for more information about EGRA and how it has been used in various contexts.

develop rigorous reading content standards for the primary school levels. These new minimum requirements cover phonemic awareness, phonics, vocabulary, fluency, reading comprehension, listening comprehension, writing, and what the standards refer to as "literary aspects" (EGRA 2014 Baseline Report). The final iteration of this work was reviewed by a broad constituency of stakeholders during roundtables at the Kyrgyz Academy of Education (KAE), by primary school teachers during in-service training (IST) courses, and with deputy school principals in 2014. IST training materials as well as reading standards for Grades 1 through 4 were approved at the KAE Academic Council's meeting in May 2014. Much has been accomplished.

The establishment in 2013 of the USAID Quality Reading Project built upon these earlier international and domestic initiatives in the Kyrgyz Republic. The project has now rolled out and implemented a comprehensive plan for the development, administration, analysis, and dissemination of EGRA results across the implementation of a longitudinal time series of EGRA subtasks in the Republic. In collaboration with key stakeholders, the USAID Quality Reading Project administered a baseline EGRA in April 2014. In April 2015, a follow-up midterm assessment was conducted and the results of that assessment are the focus of this midterm report. An additional assessment will take place in 2016 and the final EGRA will be in 2017.

Research conducted on primarily monolingual, alphabetic language learners, shows that there are five important components of early grade reading: phonemic awareness, phonics, oral reading fluency, vocabulary, and reading comprehension (NICHD, 2006). The Kyrgyz Republic EGRA has been customized to be linguistically and culturally appropriate for use in the Republic; but core constituent parts in the reading skills domains listed above are included in this EGRA. A fuller explication of the purpose and use of each subtask is presented in later sections of this report.

THE MIDTERM DATA ANALYTIC REPORT

This Midterm Data Analytic Report presents results from the April 2015, midterm EGRA administration. Before the results section, however, a brief overview of the approach employed to effectively implement EGRA and related activities is highlighted. The following sections of the report contain: A brief review of the core USAID Quality Reading Project research methodology; the data collection and sampling plan; information about how the EGRA was designed, developed, administered, and analyzed; results from 2015 scaled for comparability to the 2014 baseline; an explication of the nine EGRA subtasks with accompanying tables, figures, and results; and, an interpretive discussion section to provide readers a clear understanding of how results can be interpreted. Frequency and descriptive statistics are presented where appropriate and key questions for discussion at the policy, classroom, parental, and other stakeholder levels are raised at the end of the report.

III. RESEARCH APPROACH

The 2015 EGRA in the Kyrgyz Republic collects regionally representative sample data on reading progress that enables policy and classroom-level inferences to be drawn to improve reading outcomes.[2] EGRA serves neither as a diagnostic for individual pupils nor as a high stakes examination. Instead, it identifies weaknesses in reading and listening skills that have been proven predictors of future reading success, according to the research that is available. The EGRA initiative in Kyrgyzstan and Tajikistan is some of the first EGRA work in countries with Cyrillic alphabets. Significantly, the project has ensured that the roll out of EGRA also serves as a controlled experiment in which we will examine the effect of comprehensive professional development programming on treatment and control groups across the country.[3] In addition to collecting data on reading skills, a core component of EGRA implementation has been to gather contextual and background data on the pupils assessed and the communities they live in to determine the nature of the relationships between reading skills acquisition and selected school, home, and environmental factors. In the fall of 2015, the project will release the results of an impact study that has analyzed a host of these factors and their relationship to reading achievement outcomes.

DATA COLLECTION PLAN

In April 2014, the USAID Quality Reading Project collected data corresponding to pupil reading outcomes at the pre-intervention stage to establish a baseline. Midterm data on pupil reading progress to monitor progress were gathered in April 2015, which is the focus of this report. Data also will be collected in subsequent years to evaluate impact and change. The team has used cross-sectional and longitudinal research designs for the study. For the cross-sectional design covering Grades 2 and 4, the baseline group will be compared to different groups of pupils at the same schools and the same grade levels in subsequent years.

For the longitudinal design, the same pupils' reading performance at Grade 1 in 2014 will be compared with their performances at Grade 2 in 2015 and at Grade 4 in 2017. A key feature of this design is that pupil reading performances at Grades 1, 2, and 4 are tracked and reported on the same measurement scale on several of the subtasks. The process of bringing reading performance onto the same scale is called *vertical scaling*. Table 1 below shows the assessment data collection plan.

TABLE 1: CROSS-SECTIONAL AND LONGITUDINAL DESIGN

Cohort	2014	2015	2016	2017
Cross-Sectional Design				
1	G2	G2		G2

[2] The data collected are representative in four oblasts: Chui, Jalal-Abad, Bishkek, and Talas (regions in cohort 1 of the intervention).

[3] For more on the USAID/QRP reading interventions, see "Literacy Boost Community Action: Creating a Culture of Reading Outside School Walls" at: http://resourcecentre.savethechildren.se/library/literacy-boost-community-action-creating-culture-reading-outside-school-walls

Cohort	2014	2015	2016	2017
	G4	G4		G4
2 & 3	G2		G2	G2
	G4		G4	G4
Longitudinal Design				
1	G1			
		G2		
				G4

Baseline Analytic Report (2014)

To achieve a more accurate measure of reading outcomes, the USAID Quality Reading Project uses a vertically equated common-matrix sample design for Grades 1, 2, and 4 for several of the subtasks. There is a single form for each grade and language in the baseline, containing a set of core matrix items unique to grade level and a set of common items that appear at the exact same locations in all three grades' instruments. The common set of items brings Grades 1, 2, and 4 reading outcome measures in the baseline onto the same reporting scale and also enables tracking pupils' reading progress from grade to grade.

To measure pupil progress accurately cross-sectionally (i.e., a different cohort of pupils at the same schools in the same grade in different years) and longitudinally (i.e., same pupils in different years) without the tests being exposed, the USAID Quality Reading Project employs different sets of equated assessments in the baseline, mid-term, and at the end of the project. The assessments across different years will be horizontally linked through the same set of common items used for vertical equating. A total of two three-set instruments have been developed for EGRA; one set for Kyrgyz and one set for Russian, with each set consisting of Grades 1, 2, and 4 assessments.

CONTENT OF THE EGRA SUBTASKS

In 2015 the EGRA in the Kyrgyz Republic had nine total subtasks.[4] The 2014 section, *Unfamiliar Words* (2014), was implemented as *Nonsense Words* in 2015.[5] Table 2 below presents the nine tasks (second grade) and seven tasks (fourth grade) and the demands of the tasks for the 2015 assessment in both the Kyrgyz and Russian languages. Recall that four of the nine subtasks for second graders were timed and three of the seven subtasks for fourth graders were timed. Pupils were given a maximum of 120 seconds to complete each timed subtask. These timed tasks were reading letters *Letter Name Recognition*, *Familiar Word*

[4] One subtest from 2014, *Initial Word Sound*, was dropped from the 2015 EGRA because there was no difference in the results between the *Initial Word Sound* and *Initial Letter Sound* subtasks and Russian and Kyrgyz language experts recommended to exclude it.

[5] In the *Unfamiliar Words* task, pupils were asked to read actual words, but ones that were judged to be above the pupils' grade level. Because it is difficult to control exposure to such words, it cannot be determined if decoding skills or memorization/familiarity are being used in the accomplishment of the task. In the *Nonsense Words* task, items are not actual words but ones that resemble the structure of common one and two syllable grade-level words. This task represents a more uncontaminated measure of decoding. The objective of this task is to determine how well pupils can associate written letter combinations with their spoken forms (words) without relying on sight reading (Hirsch, 2003).

Recognition, *Nonsense Words*, and *Reading Passage.* The purpose of the 120-second metric was to provide a broad measure from which to assess and interpret reading fluency. Slower, but nonetheless accurate readers may be penalized on such tasks depending on the metric used to evaluate fluency. As pace in reading can vary over time (e.g., it may be faster or slower at task outset), having a wider measure enabled administrators to record multiple data points; time *at* 1 minute, time *at* 2 minutes, a rate *per minute,* as well as the ability to determine the difference between numbers read in the first and second minutes. The data reported in this midterm report for comparison purposes with 2014 is reading rate *per minute*. At present, analyses on the data collected related to reading pace are ongoing.

The full set of subtasks, administered orally by a trained administrator in one-on-one sessions, required 25 to 30 minutes to conduct. Results from a 10th section added to collect demographic and other background information about the pupils will be reported on in forthcoming EGRA monitoring and evaluation impact studies. Table 2 below provides an overview of the subtasks and describes what pupils were required to do on each. A more detailed description of each subtask is presented in later sections of this report.

TABLE 2: EGRA IN THE KYRGYZ REPUBLIC

2015 Midterm Subtasks		
Subtask (Grade)	**Reading Skills**	**Pupils were asked to:**
1. Letter Name Recognition (2)	Letter identification	*Identify* correctly and *read* aloud letters of the alphabet in lower and upper cases in a 2-minute period (TIMED)
2. Initial Letter Sound (2)	Phonemic awareness, letter–sound correspondence	*Sound out* 10 commonly used letters, randomly arranged, repeating after administrator
3. Familiar Word Recognition (2,4)	Word recognition and decoding	*Read aloud* 40 familiar[6] one- and two-syllable words in a 2-minute period (TIMED)
4. Nonsense Word Recognition (2,4)	Letter–sound correspondence, decoding	*Decode* 40 one- and two-syllable nonsense words in a 2-minute period (TIMED)
5. Oral Vocabulary (2,4)	Basic vocabulary, listening in context	*Identify* 10 objects in pictures after listening to a list of objects read by the administrator using the validated PPVT-R format
6a. Passage Reading (2,4)	Reading fluency	*Demonstrate* oral reading of one short passage in a 2-minute period (TIMED)
6b. Reading Comprehension (2,4)	Reading comprehension of texts	*Demonstrate* reading comprehension by answering 4-5 oral questions from the administrator about the reading passage just read aloud
7. Listening Comprehension (2,4)	Oral language comprehension, vocabulary knowledge	*Demonstrate* listening comprehension by answering 4-5 questions based on a short paragraph read by the administrator

[6] "Familiar" words were identified through an EGRA protocol that requires word analyses on grade- level textbooks to derive counts of the most commonly encountered words.

2015 Midterm Subtasks		
Subtask (Grade)	**Reading Skills**	**Pupils were asked to:**
8. Dictation (2,4)	Oral language comprehension, writing skills	*Listen* to a sentence and *reproduce* it correctly in written form

EGRA DEVELOPMENT AND ADMINISTRATION

For both the baseline and midterm assessments, the USAID Quality Reading Project team adapted EGRA subtasks for the country context using a protocol for the localization of EGRA subtasks. As will become clear under the subtask descriptions below, functional ability with orthography, morphology, phonemes, and phonology is essential to Kyrgyz and Russian language reading development. After reviewing Kyrgyz and Russian language primary grade reading standards, item writers were provided professional development on writing inferential questions related to reading and listening passages. After the items were collated, pilot tests were administered and the initial psychometric properties of the items were examined. Each item was reviewed and analyzed to ensure fairness and balance based on gender and other criteria.

The EGRA midterm assessment was the same instrument as the 2014 baseline with the exception of the changes to the subtasks noted above. A full explication of how the team aligned items to the reading standards, conducted training in item adaptation and development, and piloted and analyzed initial results, can be found in the *EGRA 2014 Baseline Analytic Report for the Kyrgyz Republic.* For more information on how EGRA administrators were trained, see Appendix 2 (this report), and for more on the EGRA administration plan, see Appendix 3 (this report).

SAMPLING PLAN AND FINAL SAMPLE

The midterm EGRA was administered in 60 randomly selected schools drawn from four regions: Bishkek, Chui, Jalal-Abad, and Talas. These four regions comprise cohort 1 of the intervention and serve as the location of project activities rolled out in the first year. Schools were sampled proportionally to the school population in these regions. These 60 schools included 30 schools receiving the USAID Quality Reading Project intervention and 30 control schools not receiving the intervention. Characteristics of the 2015 schools are summarized in Tables 3 and 4 below. Schools were chosen using a stratified sample that allowed for adequate representation from the region in each of the necessary characteristics.

TABLE 3: SCHOOL SIZE AND LOCATION IN THE SAMPLING PLAN

School Size					School Location				
Region	**Small**	**Med.**	**Large**	**Total**	**Region**	**Rural**	**Semi-urban**	**Urban**	**Total**
Bishkek	2	0	2	4	Bishkek	0	0	4	4
Chui	3	13	5	21	Chui	20	1	0	21
Jalal-Abad	10	15	3	28	Jalal-Abad	21	3	4	28
Talas	1	2	4	7	Talas	5	0	2	7
Total	16	30	14	60	**Total**	46	4	10	60

The project administered EGRA to 2,286 pupils in 2015. In the midterm assessment, only Grade 2 and Grade 4 pupils from Cohort 1 were tested, as required by the design. Grade 2 pupils are used for the cross-sectional (horizontal) study (i.e., same grade from one year to the next), and for the longitudinal (vertical) study (i.e., same students progressing from one grade to the next across time). In the Talas, Jalal-Abad, Chui, and Bishkek regions, the same pupils tested in the baseline in Grade 1 were tested in the midline Grade 2.

TABLE 4: SAMPLING PLAN BY LANGUAGES OF INSTRUCTION

	School Language(s) of Instruction					
Region	**Kyrgyz**	**Russian**	**Kyr/Rus**	**Kyr/Uzb**	**Kyr/Rus/Uzb**	**Total**
Bishkek		1	**3** (k = 2, R = 1)			4
Chui	6	3	**12** (K = 6, R = 6)			21
Jalal-Abad	18		**7** (K = 3, R = 4)	**2** (K = 2)	**1** (K = 1)	28
Talas	1	1	**5** (K = 3, R = 2)			7
Total	25	5	27	2	1	**60**

WITHIN SCHOOL PUPIL SAMPLING

To randomly select pupils, enumerators first calculated the gender ratio of male to female pupils in each grade to be tested. The 20 pupils per grade were then divided between boys and girls according to that ratio, with 10 boys and 10 girls generally chosen per grade. Enumerators then calculated an interval by dividing the number of pupils per grade by the number needed for EGRA, separately by gender. The interval was used to randomly select pupils from the pupil roster list. If a selected pupil was not in school that day, or did not consent, the next pupil on the roster was selected. To ensure an adequate sample size for subgroup analysis by language, at least 10 Russian schools, or mixed-language schools that had Russian as a language of instruction, were built into the randomized school sample. If the school was designated as a Russian school for EGRA testing, pupils were given the Russian EGRA. In the case of mixed-language schools, the school was randomly designated (using a random number generated in Excel) as a Russian school for data collection purposes until the 10 Russian-school quota was met.

WEIGHTING THE SAMPLE

Sample weights were calculated and applied to the reported means to adjust for three things: (1) oversampling of control schools compared to the population; (2) oversampling of Russian schools compared to the population; and, (3) the size of the school. This was done because the same number of pupils was sampled per school per grade, thus resulting in oversampling of pupils from small schools compared to the population. For the purposes of weighting, mixed-language schools were treated as Kyrgyz unless they were part of the sample and treated as a Russian-language school. The project made this choice because mixed-language

schools tend to have a Russian sector minority. At present there is no complete data on language of instruction by student or grade for mixed-language schools. Sampling weights may change in the future depending on availability of better school-level administrative data.

ON LANGUAGE DIFFERENCE

A full understanding of the nuances of reading development in a given language is difficult to attain without a modicum of knowledge about the particular language under study and its socio-linguistic context. Although this report does not provide a detailed description of the two languages under investigation, commentary on important linguistic aspects of reading development for Kyrgyz language speakers is provided throughout the text as appropriate.[7] Kyrgyz is an Altaic (Turkic) language spoken by approximately 6 million people in Eurasia and is the state language of the Kyrgyz Republic. Important characteristics include *word agglutination*, *voiced* and *unvoiced letters, vowel harmony*, and a structured *sentence word order* that places verbs at the end of sentences (Hu & Imart, 1989). Relevant to item and instrument development, there are also ongoing socio-linguistic debates on *standardization* and language adaptation (translation) practices in the republic.[8]

[7] Because Russian is a well-known language outside of Eurasia, we provide limited description of the language as other resources are widely available.

[8] Contested issues include an alleged "Russification" of Kyrgyz syntax and expression (especially the use of the conjunction "and" in Kyrgyz); unresolved regional dialect issues that result in inconsistencies in textbooks and other educational materials in regard to rules, syntax, vocabulary, etc.; ideological forces that seek to eliminate *loan words* from the Kyrgyz language; how to incorporate *new words* and concepts into the language, just to name a few key issues. For more detail on these discussions, see Wright (1999), Korth (2004), Korth (2005), and Drummond (2011).

IV. OVERALL RESULTS FROM THE 2015 MIDTERM EGRA

Table 5 presents comparable summary data from the 2014 baseline and 2015 midterm EGRA in the Kyrgyz Republic. Subtasks are presented by language, total scores, and standard deviations for each subtask. Because this is a presentation of subtask results across 2 years, it was important that scores be adjusted to account for any differences in the difficulties of the subtasks from year to year. Item equating was carried out, and several subtask scores in the 2015 midterm were transformed into comparable 2014 scores. This process entailed identifying commonly shared items, or "anchor items," found in the subtasks in both years, and estimating item parameters for the 2014 and 2015 subtasks to determine the stability of the anchor items. The final step was to produce a conversion table in which 2015 scores were converted to 2014 scores.

Once equated, the common scale allowed comparison between assessment years to be made. Some of the subtasks, due to the format of the items and the short time available for assessment (e.g., reading and listening comprehension subtasks each had one core task) could not be equated because the subtask format dictated that new items be developed every year for security reasons. In other words, the reading comprehension text needed to be new every year. In a few cases, equating was not possible due to the properties of the items themselves.

The EGRA subtask scores presented in the table below are presented in one of two ways. First, in the difference column, for subtasks scored by a percent correct, the number represents a percentage point change on a given task from baseline to midterm; that is, if a score went from 90% to 95% during that period, the third "change" column would indicate a 5 percentage point change in score. For timed subtasks, the value in the "difference" column indicates the difference in the number of correct letters or words per minute correct, depending on the subtask. In the data presented below, the timed scores are presented in letter or words read per minute.

There are cautions in regard to interpretation of the comparative data below. Increases or decreases in task scores or proportions of students gaining in proficiency levels cannot directly answer questions about whether the USAID Quality Reading Project intervention has contributed to improved reading outcomes in schools across the republic. To make claims about the impact of project endeavors, the difference in growth rates from the control and treatment populations will need to be examined. Based on what is known about EGRA in other contexts and early grade reading growth in general, some level of natural improvement at this developmental stage should be expected (Gove, 2009).

What needs to be understood is whether the reading outcomes of the pupils who received project support improved more than those who did not. There are intervening factors that must be disentangled before making claims of attribution. In the forthcoming project report, *EGRA Impact Study in the Kyrgyz Republic*, valid inferences about intervention impact on

reading outcomes will be possible as the EGRA design is conducive to the implementation of a controlled experiment: The impact of the intervention will be determined by comparing the control group to the treatment group, while controlling for a host of intervening factors.

TABLE 5: COMPARISON OF BASELINE AND MIDTERM EQUATED SUBTASKS

Subtasks (Grade 2)	Kyrgyz					Russian				
	Baseline		Midterm in Baseline Scale		Difference	Baseline		Midterm in Baseline Scale		Difference
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
	Subtasks (Grade 2)					Subtasks (Grade 2)				
Letter Name Recognition letters per min	**60.44**	20.27	**69.06**	17.80	**8.66**	**57.50**	20.87	**56.57**	19.30	**-0.93**
Initial Letter Sound percent correct	90.07	12.50	93.93	10.86	3.83					
Familiar Word Recognition words per minute	**50.20**	26.99	**56.70**	25.34	**6.50**	**49.15**	21.38	**65.19**	23.19	**16.03**
Dictation percent correct	53.98	25.54	50.24	17.92	-3.76					
	Subtasks (Grade 4)					Subtasks (Grade 4)				
Familiar Word Recognition words per minute	**71.39**	29.34	**88.40**	33.13	17.01	**65.68**	23.20	**92.90**	33.72	27.22

Table 5 above presents the comparative data between the 2014 baseline and 2015 midterm EGRA years. The subtasks that contained anchor items included *Letter Name Recognition, Initial Letter Sound, Familiar Word Recognition (2 and 4),* and *Dictation*. From 2014 to 2015, scores on three of the four subtasks for Kyrgyz language pupils in Grade 2 increased. The increases ranged from the smallest at +3.8 percentage points for *Initial Letter Sound* to the largest at +8.66 letters per minute on *Letter Name Recognition*. On the *Dictation* subtask, scores decreased by 3.7 percentage points.

For the Russian language (Grade 2) there was a substantial increase in scores on the *Familiar Word Recognition*, at + 16.03 words per minute. There was a slight decrease in scores at less than one letter per minute on *Letter Name Recognition*. In Grade 4 (both languages) there were substantial score increases on the *Familiar Word Recognition* subtasks, + 17.01 for the Kyrgyz cohort and +22.72 for the Russian cohort. It is important to reiterate that the scores presented above represent *scaled scores* that take into account the differences in difficulty levels between the 2 assessment years.

For the next set of EGRA subtasks presented, the scores on the subtasks have not been put on a common scale. Although the stated subtask purpose, content, format, and administration for all but one of the subtasks was approximately the same in 2014 and 2015, slight differences in difficulty of the test forms across years make valid comparative inferences impossible. The scores are presented in Tables 6 and 7 for informational purposes only. The scores are presented along with their standard deviations.

TABLE 6: NON-EQUATED SUBTASK RESULTS GRADE 2 (2014 AND 2015)

Subtask		Kyrgyz Language		Russian Language	
		Baseline	Midterm	Baseline	Midterm
Initial Letter Sound percent correct	**2**			**90.2%** (15.4)	**94.4%** (14.0)
Nonsense Words* words per minute	**2**	**24.2** (14.1)	**26.0** (11.1)	**26.8** (11.2)	**30.6** (11.2)
Oral Vocabulary percent correct	**2**	**92.0%** (10.9)	**88.5%** (12.6)	**90.6%** (13.4)	**84.0%** (14.4)
Reading Passage words per min	**2**	**32.2** (18.1)	**40.1** (20.4)	**40.3** (19.2)	**43.9** (19.4)
Reading Comprehension percent correct	**2**	**53.4%** (32.9)	**56.5%** (32.3)	**9.7%** (41.7)	**54.7%** (29.8)
Listening Comprehension percent correct	**2**	**75.1%** (26.4)	**69.5%** (29.4)	**68.4%** (32.7)	**80.9%** (27.7)
Dictation percent correct	**2**			**73.3%** (24.4)	**79.4%** (18.3)
		Sample Sizes: 2015 K2 = 658 2014 K2 = 1627		**Sample Sizes:** 2015 R2 = 324 2014 R2 = 602	

* In 2015 this subtask was *Nonsense Words* while the subtask in 2014 was *Unfamiliar Words*, or the reading of actual words at a higher grade level.

TABLE 7: NONEQUATED SUBTASK RESULTS GRADE 4 (2014 AND 2015)

SUBTASK		Kyrgyz Language		Russian Language	
		Baseline	Midterm	Baseline	Midterm
Nonsense Words* words per minute	**4**	**31.3** (15.8)	**33.9** (13.1)	**38.7** (15.4)	**32.7** (11.9)
Oral Vocabulary percent correct	**4**	**97.1%** (5.8)	**90.7%** (11.1)	**92.8%** (9.6)	**86.9%** (14.3)
Reading Passage words per min	**4**	**68.4** (28.9)	**60.2** (23.8)	**68.4** (25.4)	**62.3%** (26.8)
Reading Comprehension percent correct	**4**	**69.3** (28.4)	**78.4%** (25.4)	**61.7%** (30.5)	**63.1%** (32.7)
Listening Comprehension percent correct	**4**	**68.7%** (27.6)	**66.9%** (29.2)	**87.1%** (22.4)	**71.0%** (28.2)
Dictation percent correct	**4**	**80.0%** (19.7)	**80.3%** (20.7)	85.3% (12.8)	**87.9%** (15.5)
		Sample Sizes: In 2015, K4 = 677 In 2014, K4 = 1691		**Sample Sizes:** In 2015, R4 = 312 In 2014, R4 = 577	

* In 2015 this reading subtask was *Nonsense Words* while the subtask in 2014 was *Unfamiliar Words*, or the reading of actual words at a higher grade level.

BENCHMARKING READING OUTCOMES IN 2014 AND 2015

The midterm results in this report are presented through (1) overall pupil performance on each of the subtasks from 2015 with reference to changes from 2014, where appropriate; (2) pupil progress on meeting benchmarks in two core areas: *Reading Fluency* and *Reading Comprehension*, and (3) the disaggregation of results on each subtask by language, gender and demographic location of school, urban or rural. In this next section, results of the national benchmarking are presented first, followed by a more detailed breakdown of the results from each subtask.

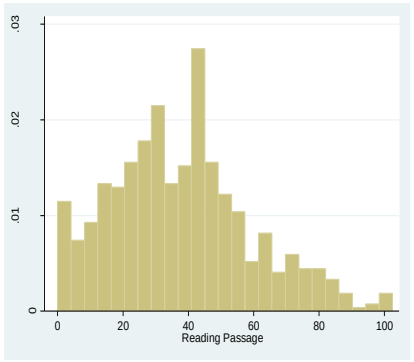
Proficiency scales developed by the MOES both before and during ongoing USAID Quality Reading Project work in several domains provide one means to examine change.[9] The proficiency scales for reading as developed by the MOES include the following performance categories: *Below Basic*, *Standard* (or basic), *Proficient*, and *Advanced*. Attaining the "standard/basic" benchmark does not indicate high proficiency, only that a reader meets a minimum reading requirement. Note that in the *2014 EGRA Baseline Analytic Report*, the results presented were the percentage of pupils attaining the "proficient" (third-level category), not the "standard" (second-level category) as reported here. Without noting this difference in reporting, it might appear that the percentage of those meeting the minimum standard appears to have changed dramatically when in fact the overall data patterns are similar; the 2015 results are simply being reported in accordance with standard practice of the MOES.

For the 2015 midterm report, the metric for establishing whether pupils met the minimum acceptable standard on the construct *Reading Fluency* was based on the 2006 *National Standards for Reading* in the Kyrgyz Republic. Meeting the standard required a reading performance level of 40 words per minute at the second-grade level, and 80 words per minute at the eighth-grade level (Tvaruzkova & Shamatov, 2012).[10] The subtask used to assess fluency was the *Reading Passage* subtask (page 39, this report). Note that the national mean (Kyrgyz Grade 2) is approximately normally distributed with the mean right at the 40 words per minute mark. This indicates that a large proportion of pupils are not yet attaining the basic standard (Figure 1).

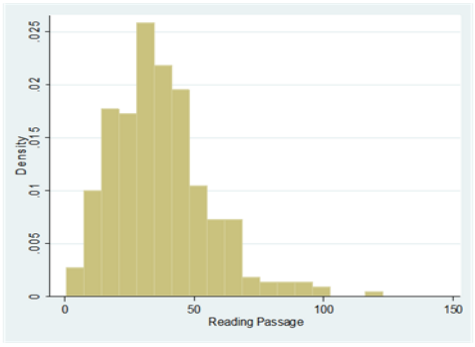
[9] For further reference on the standard-setting methods employed by the USAID Quality Reading Project, see also Livingston & Zieky (1982), Loomis & Bourque (2001), and Cizek & Bunch (2007).

[10] For experimental purposes, data also have been collected on reading rates based on where students find themselves "at the 1 minute" and "at the 2 minute" marks on several of the reading tasks. Based on last year's EGRA, there is some reason to believe that the pace of pupil reading varies across time: Some slow readers may be penalized by a metric that assesses only what they are capable of reading in the first minute. The USAID Quality Reading Project team hopes to conduct further research in this area.

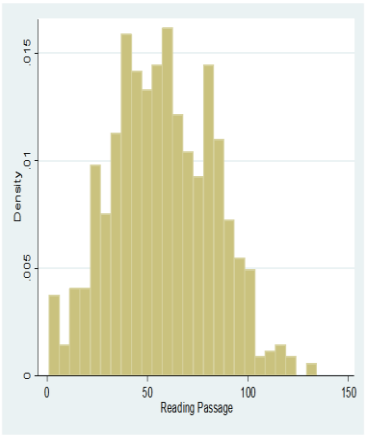
Figure 1. Distributions in Reading Fluency



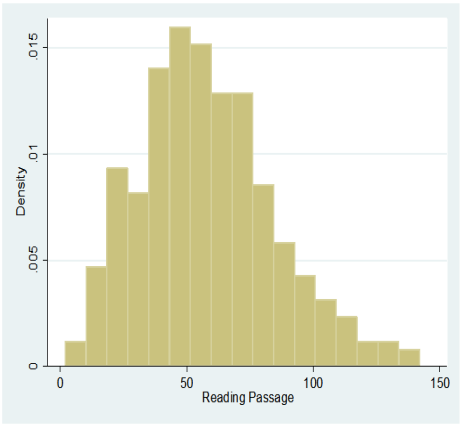
Kyrgyz Grade 2 (2105)



Russian Grade 2 (2015)



Kyrgyz Grade 4 (2015)



Russian Grade 4 (2015)

Table 8 below presents a breakdown of the approximate proportion of pupils meeting the national standard for reading fluency at both Grades 2 and 4, Kyrgyz and Russian languages. Note that the proportions "*meeting the standard"* has increased for both the Russian and Kyrgyz language groups at Grade 2 but slightly decreased for both groups in Grade 4.[11]

TABLE 8: PROPORTIONS OF PUPILS MEETING READING FLUENCY STANDARD

Language Skill	Benchmark	Kyrgyz		Russian	
Reading Fluency	**Standard or Above**	**2014%**	**2015%**	**2014%**	**2015%**
Grade 2	40 words or above	31.2%	49.1%	48.7%	55.3%
Grade 4	80 words or above	35.3%	25.9%	30.6%	23.3%

Attaining the benchmark of *standard* in reading comprehension indicated that a reader met a minimum comprehension requirement based on the *Reading Comprehension* subtask of EGRA (see page 24, this report). Note that in the *2014 EGRA Baseline Analytic Report*,

[11] Caution is in order when interpreting these proportions as absolutes. As noted, in regard to equating across test forms from year to year, while the same type of *Reading Passage* subtask was employed to determine fluency in both years, the actual reading texts changed slightly from year to year as they cannot be reused.

attainment of the reading comprehension standard was reported as composite score of several subtasks. A composite score has not been employed in the 2015 analysis because an empirically demonstrable relationship of predictive validity between each of those subtasks and reading comprehension has yet to be established in the Kyrgyz or Russian language EGRAs. The project will further investigate what predicts reading comprehension success in languages using the Cyrillic alphabet and will seek to develop a more robust set of reading comprehension tasks. For the 2015 EGRA the *Reading Comprehension* subtask alone was employed to assess attainment of the benchmark. The proportion of pupils meeting the benchmark for Grades 2 and 4 are reported for both language groups in Table 9 below.

TABLE 9: PERCENT MEETING READING COMPREHENSION BENCHMARKS BY GRADE AND LANGUAGE

Skill	Benchmark	Kyrgyz		Russian	
Reading Comprehension		**2014%**	**2015%**	**2014%***	**2015%**
Grade 2	K: 3 of 4 correct R: 4 of 5 correct (2015)	42.8%	44.4%		35.7%
Grade 4	R: 3 of 4 correct K: 4 of 5 correct	55.8%	70.6%	50.5%	54.2%

* As the 2014 Grade 2 subtask contained only three questions, results were not used for benchmarking purposes.

V. 2015 MIDTERM EGRA TASKS IN THE KYRGYZ REPUBLIC

PRESENTATION OF THE DATA

Scores on each task are presented by total score, by language group, by gender, and by demographic location of school (urban or rural) for both grades starting with Table 10. Because numerical differences in mean scores can be misleading, score differences were tested for statistical significance by conducting t-tests using Stata software. The t-test assumes a null hypothesis of *equality of means* between groups under study (e.g. male/female). Because tests for statistical significance frequently result in the rejection of the null hypothesis when sample sizes are large, an effect size measure was also estimated to determine whether or not there was any *practical* significance of the differences in means estimated (Cohen, 1992). An iteration of what is frequently referred to as "Cohen's D" was employed to avoid over estimation of any differences between groups that could in fact be the result of a statistical artifact. Calculations of effect size were done in excel software using the formula:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\frac{(n_1 - 1)\sigma_1 + (n_2 - 1)\sigma_2}{n_1 + n_2 - 2}}}$$

Where:

μ = the mean of the samples, 1 and 2

n = the number in the samples 1 and 2

σ = variance of the samples 1 and 2

Cohen's D is a standardized measure of effect size that can be applied to weighted samples and yet reports on a standard, recognizable scale (in practice, it reports the distance two means are from each other in standard deviation terms). The effect size values (determined by Cohen) are:

.2 (small effect)

.5 (medium effect)

.8 (large effect)

If the null hypothesis of no difference was retained, there was no need to calculate the effect size measure. Because many analyses were run to interpret the subtask data, model conditions for each analysis below will not be repeated in the text: The null hypothesis in all

the statistical tests that follow was that the means were equal across groups (male and female, urban and rural). Letter Name Recognition (Grade 2)

The *Letter Name Recognition* (LNR) subtask was developed to assess knowledge of the alphabet which is the foundation of learning to read. The Kyrgyz alphabet has 36 letters, and in its current iteration is considered to be an "adapted Cyrillic alphabet" due to the commonality of most letters with the Russian Cyrillic alphabet.[12] Kyrgyz, however, has several additional letters that are not found in most Slavic-language alphabets. Although few, if any, LNR studies on Kyrgyz-speaking populations are available to English-speaking audiences, studies of letter name recognition in the English and French languages have indicated that early letter name recognition is a strong predictor of long-term reading development (Chiappe, Siegel, & Wade-Woolley, 2002).

The LNR task in the Kyrgyz Republic determined: (1) whether pupils could correctly identify and read aloud both capital and small case letters, and (2) the pace of reading letters, or letters per minute. The identification of both and upper and lower case letters was employed because research in other languages has suggested that reading skills progress only after 80% of letters in both cases are mastered (Seymour et al., 2003). For this subtask in both the Kyrgyz and Russian languages, each pupil received a booklet with all letters in the alphabet. Upper and lower case letters were presented in random order in to prevent recall and recitation from memory. The Kyrgyz subtask had 69 items (letters) and the Russian subtask had 64 items (letters).

The number of letters successfully recognized was calculated at both the 1-minute and 2-minute marks as well as at a per minute rate (total time, 120 seconds). Providing 120 seconds allowed pupils with low- and high-ability levels adequate time to demonstrate what they knew and were able to do. The score calculations required the administrator to determine how many letters were attempted, how many were read correctly, and in what amount of time.

RESULTS

The mean score on the Kyrgyz LNR subtask (n = 658) was 69.9 letters per minute (SD 17.74). Chronbach's alpha, a coefficient of scale reliability, was estimated to provide an estimate of internal consistency of the items. Internal consistency refers to the extent to which the items in the test are consistently measuring the same construct at the cohort level. As the alpha coefficient increases, the portion of a score that can attributed to error will decrease: Hence higher values are desirable (generally above .80), though a high alpha coefficient does not always indicate quality as alpha can be increased simply by adding test items or by other random factors. Further, very high alpha coefficients can indicate that there is redundancy and perhaps unnecessary items in the scale. The alpha coefficient for the Kyrgyz subtask LNR was .76, not high, but acceptable for this type of subtask. The total mean score on the Russian LNR subtask (n = 324) was 57.08 letters per minute (SD 19.75). Although alpha

[12] See Appendix 1 for the full Kyrgyz alphabet. Kyrgyz has not always used the Cyrillic script; both the Arabic and Latin alphabets have been used in recent times (Hu & Imart, 1989).

coefficients for both the Kyrgyz and Russian groups were satisfactory, there was a noted difference between the two subtasks as alpha was .94 for the Russian subtask.

RESULTS BY GENDER

The 2014 baseline report noted that Kyrgyz Grade 2 pupils read at 65 letters per minute in 2014, 4.9 letters fewer than in 2015. Recall that the subtasks are not comparable across language because the tests are language-dependent and completely different. However, it is interesting to note that rate of learning across languages appears to be different. Kyrgyz readers were identifying letters at a faster rate than their Russian peers in their respective languages. In Table 10 below we present the results for this subtask by gender in both languages. Scores are presented along with the standard deviations of each subtask in parentheses under the score. In the table below, and in all tables going forward, in addition to presenting scores, we also present the "difference" in the two scores, hence "diff." We also note with asterisks whether or not those differences were statistically significant and in the next column present the effect size value, Cohen's D, as described above. This combination of indicators enables the reader to get a sense for the cores information about the item under study

TABLE 10: LNR RESULTS BY GENDER

Subtask	Kyrgyz Language					Russian Language				
	Total n = 658	Male n = 333	Fem. n = 325	Diff.	Cohen's d	Total n = 324	Male n = 152	Fem. n =172	Diff.	Cohen's d
Letter Name Recognition	69.93 (17.74)	67.04 (17.31)	73.1 (17.68)	6.1**	0.349	57.08 (19.79)	53.0 (18.74)	60.4 (20.06)	7.41	0.381

* Significant at the .05 level
** Significant at the .01 level

Females identified 73.1 letters per minute on the Kyrgyz Grade 2 task, 6.1 letters more per minute than males for this task on average. This difference on the Kyrgyz test was statistically significant at the .01 level. The effect size of .34 can be considered a relatively small effect. Russian females also outscored their male peers, reading at more than 7 letters per minute more, but these results were not statistically significant (Table 11 below).

TABLE 11: LNR RESULTS BY DEMOGRAPHICS

Subtask	Kyrgyz Language				Russian Language			
	Urban n = 119	Rural n = 523	Diff.	Cohen's d	Urban n = 54	Rural n = 213	Diff.	Cohen's d
Letter Name Recognition	63.57 (18.69)	72.30 (16.68)	8.73**	0.511	58.17 (14.70)	56.5 (22.32)	1.65	0.275

* Significant at the .05 level
** Significant at the .01 level

For the LNR subtask Kyrgyz Grade 2, the 8.7 letter difference between urban and rural cohorts was a statistically significant difference at the .01 level. Cohen's d of .51 indicates a

moderate effect size. For the Russian Grade 2, there were no significant differences between rural and urban cohorts with a total 1.65 letter per minute differential.

INITIAL LETTER SOUND (GRADE 2)

The *Initial Letter Sound* (ILS) subtask was an assessment of phonemic awareness. A phoneme is the smallest linguistically distinctive unit of sound allowing for differentiation of two words in a language. The 2000 *National Reading Panel* meta-analysis of the literacy research (conducted primarily on literacy in the English language) determined that skill in phoneme identification and phonological awareness is strongly associated with good reading comprehension. Phonemic awareness is the foundation for learning phonological awareness, a domain that includes skills in hearing and manipulating onsets, rimes, and syllables (Snow et al., 1998; NIHCD, 2006).

As the Kyrgyz alphabet observes a strict *vowel harmony* and has rules about how letters can be combined into various phonemes, the successful manipulation of phonemes is essential for young learners. Although Kyrgyz pronunciation is regular and consistent, certain vowels can be used only in combination with other vowels in a certain letter order. For example, when adding suffixes to words to create new meaning (*agglutination*), the vowel in the suffix must take on the same letter as the last vowel in the word. The exceptions to this vowel harmony rules apply only to words of foreign origin (Oruzbaeva, 1997). There are also some letters that are used only in Russian loan words: For example, "*ж" (pron.* "zh"*)* as in the word *Журнал* (journal). In terms of pronunciation, Kyrgyz also has *"voiced"* sounds (vocal cords used) and *"voiceless"* sounds (without vibration of vocal cords) (Oruzbaeva, 1997).

For the ILS task the pupil subtask booklet included a list of the 10 most frequently used letters in the Kyrgyz or Russian alphabets, randomly arranged. The frequency of letters in everyday use was determined during development of the assessment by text analysis and calculations of word count frequencies. The administrator read each word two times and asked the pupils to make the first sound of the word. If a pupil did not answer within 3 seconds, a response "no answer" was recorded. The maximum score for this section was 10 points, 1 point for each correct answer (Table 12 below).

TABLE 12: ILS SUBTASK BY GENDER

Subtask	Kyrgyz Language					Russian Language				
	Total n = 658	Male n = 333	Female n = 320	Diff.	Cohen's d	Total n = 324	Male n = 152	Female n = 172	Diff.	Cohen's d
Initial Letter Sound	95.24% (10.05)	94.30% (11.71)	96.28% (7.70)	1.99	0.203	94.37% (14.03)	94.39% (13.53)	94.36% (14.43)	0.03	0.002

* Significant at the .05 level
** Significant at the .01 level

As in 2014, results on the *Initial Letter Sound* task in 2015 indicated high performance across gender and language groups with mean scores well over 90% correct. There were no significant differences in the means by gender on the *Initial Letter Sound* subtask. The scores on this subtask across demographic groups indicate that urban cohorts scored numerically a

bit higher than their rural counterparts, but that these differences were not statistically significant for either language group (Table 13 below).

TABLE 13: ILS BY SCHOOL DEMOGRAPHICS (GRADE 2)

Subtask	Kyrgyz Language				Russian Language			
	Urban n = 119	Rural n = 523	Diff.	Cohen's d	Urban n = 54	Rural n = 213	Diff.	Cohen's d
Initial Letter Sound	96.26% (9.49)	96.02% (10.15)	2.64	0.078	96.30% (12.54)	93.60% (14.92)	2.64	0.182

* Significant at the .05 level
**Significant at the .01 level

FAMILIAR WORD RECOGNITION (GRADES 2 AND 4)

Familiar Word Recognition (FWR) was an assessment of the ability to recognize and read frequently occurring words. For this task, EGRA administrators were able to attain a measure of decontextualized pupil recognition and decoding skills that is a distinct skill from reading comprehension from text (Gove, 2009). The FWR subtask was selected for the EGRA in the Kyrgyz Republic because word formation and decoding in the Kyrgyz language can be grammatically challenging. Like other Altaic (Turkic) languages, Kyrgyz is an *agglutinative* language. In agglutinative languages, word meaning is conveyed through the attachment of affixes (primarily suffixes in Kyrgyz) to nouns. These suffixes in Kyrgyz can determine possession, number, location, direction, as well as other cases (Oruzbaeva, 1997). For example, consider the word *kyzdarga* in the sentence below:

Men kyzdarga bara jatamin.

In English the sentence reads "I am going to the girls." Word for word in English it can be understood as *Men* (I)—*kyz/dar/ga* (kyz/girl +dar/plural form +ga/to)—*bara jatamin* (go, present tense). Note that the single word *kyzdarga* indicates direct object, number, and direction (three different cases) and is constituted by three separate morphemes. If you wanted to say "I am going to *my* girls" the word becomes *kyz/**ym**/dar/ga*—note the additional case of ownership added by the use of the second morpheme *"ym."* Negation is conveyed by a morpheme added to verbs. Kyrgyz words can consist of up to six or seven morphemes, and the removal or addition of one morpheme results in a different meaning, idea, or construct. An implication for language learners is that the acquisition of Kyrgyz words requires attention to morpheme development and how they combine in myriad ways to create different meaning in single words.

The *Familiar Word Recognition* subtask examined whether pupils in Grades 2 and 4 were able to read aloud 40 familiar words at grade level. Recall that unlike the *Reading Passage*, this subtask presents a list of unrelated words that are not presented as a story or complete text. Frequency of words at both grades was determined through a word count analysis of the most commonly used words in textbooks of appropriate level. A selection of 40 words was then randomly arranged in the pupil stimulus. The FWR tasks were scored on a *words per minute* calculation that called for the administrator to determine how many words were attempted,

how many were read correctly, and in what time over the course of 120 seconds. Data were collected on words per minute at both the 1-minute and 2-minute points of the task.

The total mean score on the Kyrgyz Grade 2 subtask (n = 658) was 57.13 letters per minute (SD 26.89). Chronbach's alpha was estimated to be .76 for this subtask, not high but adequate for this type of word identification task. The total mean score on the Russian Grade 2 subtask (n = 334) was 65.23 letters per minute (SD 23.1). The alpha coefficient for the Russian subtask was also not high, .79. Recall from above that this subtask was one of the comparable subtasks with the 2014 baseline data. After scores were put on a common scale, the Kyrgyz Grade 2 pupils increased their scores by 6.5 words per minute in 2015, while the Russian Grade 2 pupils increased their scores by 16 words per minute from 2014.

The total mean score on the Kyrgyz Grade 4 subtask (n = 677) was 89.03 letters per minute (SD 33.42). The alpha coefficient for the Kyrgyz subtask was .72, not high but reasonable for this type of subtask. The total mean score on the Russian Grade 4 subtask (n = 324) was 94.42 letters per minute (SD 34.02). The Grade 4 subtask was also comparable to year 2014, and both the Kyrgyz and the Russian cohorts increased their words per minute from last year: Kyrgyz by 17.4 words per minute; Russian by 27.22 words per minute. See Table 14 below for the full data panel for 2015.

TABLE 14: RESULTS BY GENDER AND DEMOGRAPICS (GRADES 2 AND 4)

Subtask	Kyrgyz Language					Russian Language				
Gender	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**
Grade 2	n=658	n=333	n=325			n=324	n=152	n=172		
FWR	57.13 (26.89)	52.38 (25.85)	62.38 (27.08)	10.01**	0.378	65.23 (23.3)	62.73 (25.87)	67.27 (20.82)	4.54	0.195
Grade 4	n=677	n=346	n=331			n=312	n=152	n=160		
FWR	89.03 (33.42)	80.25 (33.59)	98.01 (30.81)	17.76**	0.551	94.42 (34.02)	85.13 (33.86)	103.07 (31.91)	17.94**	0.546
Demographics		**Urban**	**Rural**	**Diff.**	**Cohen's d**		**Urban**	**Rural**	**Diff.**	**Cohen's d**
Grade 2		n=119	n=523				n=54	n=213		
FWR		54.99 (24.29)	57.74 (27.71)	2.77	0.102		72.91 (19.90)	61.59 (23.88)	11.32*	0.489
Grade 4		n=121	n=539				n=55	n=201		
FWR		90.95	87.61	3.34	0.1		110.05 (30.22)	88.47 (34.11)	21.58**	0.648

* Significant at the .05 level
**Significant at the .01 level

On the *Familiar Word Recognition* subtasks, females outperformed males in both languages at both grade levels. These differences were statistically significant in all groups at the .01 level except at Russian Grade 2. Note the relatively high d values indicating firm, "moderate" effect sizes. Most importantly, the male-female score gap almost doubles by Grade 4 for the Kyrgyz cohort and *more than doubles* for the Russian group. In 2014 the differences in *Familiar Words Recognition* reading rates across demographic groups were statistically significant in almost all cases. In 2015, the differences are only significant for the Russian

cohorts, not the Kyrgyz cohorts: Note the large score difference of 21.58, significant at the .01 level with a moderate effect size of .648 for Russian Grade 4.

NONSENSE WORD RECOGNITION (GRADES 2 AND 4)

Nonsense Word Recognition (NWR) assessed the ability of pupils to decode one- and two-syllable nonwords that could plausibly exist in the language in question due to similar structure and composition, but in fact are not actual words. The NWR task provided a measure related to that of the *Familiar Word Recognition* task but had the advantage of not allowing respondents to *sight read* words. To achieve in reading, pupils need to acquire both sight reading and decoding skills. According to Hirsch (2003), there is significant evidence that an over reliance on "sight word vocabulary" often leads to regression in reading development by age 9 or 10. Forty 40 nonwords were randomly arranged on a list for both grades in the pupil booklets and participants were asked to read as many as they could. Data were collected on words per minute at both the 1-minute and 2-minute points of the exercise. The NWR task was graded on a *words per minute* calculation that called for administrator to determine how many words were attempted, how many were read correctly, and in what time frame on this 120 second task.

The total mean score on the Kyrgyz Grade 2 *Unfamiliar Words* (UWR) subtask (n = 658) was 25.99 nonwords per minute (SD 11.10). Chronbach's alpha was estimated to provide an estimate of internal consistency. The alpha level for the Kyrgyz subtask UWR was .83, adequate for this type of subtask. The total mean score on the Russian Grade 2 subtask (n = 324) was 30.57 letters per minute (SD 11.21). The alpha coefficient for the Russian subtask was .77, not high for a test with this many items. The total mean score on the Kyrgyz Grade 4 subtask (n = 677) was 33.95 letters per minute (SD 13.06). The alpha level for the Kyrgyz subtask UWR was .80, adequate for this type of subtask. The total mean score on the Russian UWR Grade 4 subtask (n = 312) was 32.68 letters per minute (SD 11.90), and the coefficient alpha was .79.

TABLE 2: RESULTS BY GENDER AND DEMOGRAPHICS (GRADES 2 AND 4)

Subtask	Kyrgyz Language					Russian Language				
Gender	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**
Grade 2	n=658	n=333	n=325			n=324	n=152	n=172		
Nonsense Words	25.99 (11.10)	23.99 (10.70)	28.20 (11.13)	4.21**	0.385	30.57 (11.21)	28.22 (10.82)	32.50 (11.18)	4.28**	0.388
Grade 4	n=677	n=346	n=331			n=312	n=152	n=160		
Nonsense Words	33.95 (13.06)	30.37 (12.68)	37.61 (12.43)	7.24**	0.576	32.68 (11.90)	30.56 (11.46)	34.65 (11.99)	4.09	0.348
Demographics		**Urban**	**Rural**	**Diff.**	**Cohen's d**		**Urban**	**Rural**	**Diff.**	**Cohen's d**
Grade 2		n=119	n=523				n=54	n=213		
Nonsense Words		26.26 (10.96)	25.82 (11.16)	0.44	0.039		33.55 (11.15)	29.31 (11.20)	4.24	0.379
Grade 4		n=121	n=539				n=55	n=201		
Nonsense Words		35.52 (11.55)	33.,03 (13.55)	2.49	0.189		35.52 (11.55)	33.03 (13.55)	0.49*	0.516

"* Significant at the .05 level
**Significant at the .01 level"

A gender gap is evident at both grade levels and across language for the *Nonsense Word Recognition* tasks. Females outscored males in both languages and at both grade levels: Three of these four categories have statistically significant differences with only Russian Grade 4 females having higher numerical scores that were not statistically significant. For Russian Grade 4, the differences between urban and rural pupils were statistically significant (see Table 15 above).

ORAL VOCABULARY (GRADES 2 AND 4)

The *Oral Vocabulary* (OV) subtask examined whether pupils in Grades 2 and 4 were able to understand the meaning of spoken familiar words at grade level. This subtask is not part of the standard EGRA administered in many developing countries (Gove, 2009), but was considered to be worth developing in the Kyrgyz Republic. The OV task required listening skills as the administrator read aloud a list of 10 words, one word at a time. Pupils were presented with a set of four pictures for each word read and were asked to identify the picture that best matched the spoken word. Raw scores (1 point per correct answer) were calculated and converted to a percentage score for this subtask. Data tables for the subtasks by mean total scores at grade level, gender, and differences by school demographics (urban or rural) are in Table 16 below.

TABLE 16: ORAL VOCABULARY

Subtask	Kyrgyz Language					Russian Language				
Gender	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**
Grade 2	n=658	n=333	n=325			n=324	n=152	n=172		
Oral Vocabulary	88.51 (12.59)	88.19 (12.55)	88.86 (12.63)	0.67	0.053	83.99 (14.39)	84.02 (14.04)	83.97 (14.71)	0.05	0.003
Grade 4	n=677	n=346	n=331			n=312	n=152	n=160		

Oral Vocabulary	90.70 (11.12)	91.81 (10.39)	89.55 (11.73)	2.26*	0.204	86.94 (14.34)	88.56 (12.15)	85.44 (16.00)	3.12	0.219
Demographics		**Urban**	**Rural**	**Diff.**	**Cohen's d**		**Urban**	**Rural**	**Diff.**	**Cohen's d**
Grade 2		n=119	n=523				n=54	n=213		
Oral Vocabulary		87.25 (12.34)	89.05 (12.65)	1.8	0.142		84.66 (13.47)	83.87 (14.69)	0.79	0.055
Grade 4		n=121	n=539				n=55	n=201		
Oral Vocabulary		88.03 (12.13)	91.76 (10.51)	3.73	0.344		86.37 (14.89)	87.64 (14.15)	1.27	0.089

"* Significant at the .05 level
**Significant at the .01 level"

On the *Oral Vocabulary* subtask, females outperformed males numerically only in Kyrgyz Grade 2. In Grade 4, males outperformed females in both language groups, with males performing statistically better on the Kyrgyz Grade 4 with a small effect size at .20. Interestingly, rural pupils scored higher in Grade 4 (both languages), but these results were not statistically significant. For a visual representation of how the data patterns are merging with each of the subtasks and their relationships with each other, see Tables 21–25 in the findings section.

READING PASSAGE AND READING COMPREHENSION SUBTASKS (GRADES 2 AND 4)

Reading is an active process that engages the learner in all of the constituent parts of language examined thus far on the EGRA (phonemes, phonics, morphemes, vocabulary knowledge, etc.). Reading *fluency* (assessed here with the *Reading Passage* subtask) can be defined as the ability to read with speed, accuracy, and proper expression. The purpose of the timed *Reading Passage* subtask was to examine whether pupils in Grades 2 and 4 were able to read a passage with speed and accuracy with grade-appropriate words (familiar words) as presented in the pupil booklets. The *Reading Passage* task was an "oral reading" task in which pupils read the passage out loud. Oral reading was assessed because empirical studies in many contexts have demonstrated that there is a strong correlation between oral fluency and reading comprehension (Fuchs et al., 2001).

The *Reading Passage* task included paragraphs with 41 words (Kyrgyz Grade 2), 48 words (Russian Grade 2), 80 words (Kyrgyz Grade 4), and 93 words (Russian Grade 4). In subtask design, test developers conducted textbook reviews to determine what words could be considered grade appropriate. The subtask was scored on a *words per minute* calculation that called for the administrator to determine how many words were attempted, how many were read correctly over a 120-second period. Total number of words read at minute 1 and minute 2 was also collected.

Although reading fluency is considered an important precursor to reading comprehension, fluency alone is not an indicator of reading comprehension; nonetheless, it is an important foundational skill. The *Reading Comprehension* subtask, which relied on questions about the text read in the *Reading Passage* subtask, sought to determine understanding of the text and the ability of pupils to answer factual questions and make inferences based on what they read.

After a pupil completed the *Reading Passage* subtask, the administrator then moved to the *Reading Comprehension* task that was a series of questions about the passage just read. The number of questions was four questions (Kyrgyz Grade 2); five questions (Russian Grade 2); five questions (Kyrgyz Grade 4); and four questions (Russian Grade 4).

THE CONTEXT

In the Kyrgyz Republic formally learning to read at school is an integrated process, with reading tasks and activities incorporated into various school subjects during the day; reading is not taught as a stand-alone subject (Tvaruzkova & Shamatov, 2012). However, the school subject *Native Language and Literature* is taught from the earliest grades. As in other school subjects in the republic, a fundamental purpose of *Native Language and Literature* has historically been to provide a strong knowledge base and to transmit core cultural values and ideals (Shamatov & Niyozov, 2010). Instructional emphases on mastering (memorizing) core knowledge, accuracy in orthography and text and oral reproduction is manifest in the considerable time invested on pupil *reproduction of knowledge* during daily lessons. The oral reproduction of texts (memorization of key portions of literature, poems, essays, etc.) occupies a prominent place in classrooms as does orally *answering* questions posed by the teacher (Shamatov& Niyozov, 2010).

Pupils also spend time at the blackboard writing dictations and reproducing works of others, such as famous literary figures. The use of the short lecture on literary topics is also common, and knowledge acquisition is measured through performance on classroom dictations, answering questions about texts in class, and take-home reading and writing assignments. On in-class assessments and take-home work, neatness, accuracy, style, and form, are all assessed in addition to content knowledge. Historically, pupil interpretation of content was expected to be based on standard, officially accepted narratives often marked by a strong connection to accepted state ideologies (Shamatov& Niyozov, 2010).

The fact that Kyrgyz is an *agglutinative* language has implications for reading tasks. Agglutination of words results in short sentences (on average) and longer words (on average) than many Indo-European languages use to express complex ideas (Hu & Imart, 1989). For example, the word "камсыздандыруу" is an attempt to convey the concept of *private insurance* in the Kyrgyz language. Bilingual respondents in a recent study on cross-lingual test item adaptation in the republic noted that as there was no notion of private insurance during the Soviet period, there was no equivalent word in Kyrgyz today: камсыздандыруу had to be "created" in modern times. The interesting point, however, is not the lack of a culturally equivalent construct but rather that agglutination results in the lengthening of *a single word*, not the articulation of a longer series of words that seek to convey the complex meaning as one might encounter in a non agglutinative language (Drummond, 2011). According to one respondent in that 2011 study:

> "The challenge for test writers is that for some Kyrgyz texts it becomes complicated when we try to repeat the Russian syntax and constructs. It becomes complicated when translation is literal. The best way to keep the Kyrgyz intact is to break the

> Russian sentences into more sentences rather than trying to capture the Russian structure. In Kyrgyz, ideas are built not through one complex sentence, but through a series (many sentences) with simpler ideas that when compounded, express the same idea …." (Drummond, 2011, p.143).

Kyrgyz also observes a strict word order in sentences: Subject- object- verb. During the 2011 study (noted above) one of the respondents observed:

"We often start to translate from the end of the sentence because the main idea comes last. Word order is different in Kyrgyz and Russian which can also cause complexities. Because of the word order, sometimes, I translate the literal sentences first—then rearrange them in order..." (Drummond, 2011, p. 144).

The above challenges, and other such linguistic nuances, are especially important to consider when adapting EGRA subtasks across language groups (e.g., from English to Kyrgyz). The USAID Quality Reading Project seeks to provide linguistically and culturally relevant assessments as part of the EGRA research and will continue to investigate key issues in test item development in both the Russian and Kyrgyz languages so as to make reading tasks and texts high-quality instruments.

RESULTS

The total mean score on the Kyrgyz *Reading Passage* Grade 2 subtask (n = 758) was 40.14 words per minute (SD 20.43). Interestingly, as noted in the previous section, this mean score is also the national benchmark of 40 words necessary to be considered meeting the standard. Chronbach's alpha was estimated and the alpha coefficient for the Kyrgyz Grade 2 *Reading Passage* was .81, acceptable for this type of subtask. The total mean score on the Russian Grade 2 subtask (n = 324) was 43.89 words per minute (SD, 19.43), or just 3 words per minute over the national standard for minimum competency. Chronbach's alpha was estimated at .79 for the Russian second graders.

The total mean score on the Kyrgyz *Reading Passage* Grade 4 subtask (n = 677) was 60.24 words per minute (SD 23.78). The alpha level for the Kyrgyz subtask Grade 4 was .82. The total mean score on the Russian *Reading Passage* Grade 4 subtask (n = 312) was 62.35 words per minute (SD 26.84). Note that unlike Grade 2 pupils, the mean scores of pupils in Grade 4 in both languages are considerably lower than the national benchmark expectations of reading at 80 words per minute. If the Grade 2 pupils average mean was at the national benchmark, the Grade 4 pupils mean score was almost 20 words per minute lower than the expected result for pupils at that developmental stage. This explains why a much larger proportion of Grade 4 pupils are below proficiency than their Grade 2 counterparts.

As in 2014, numerical score differences by gender were large and these differences were statistically significant for three of four grades on *Reading Passage*, with scores favoring females with the one exception of Russian Grade 2. On the *Reading Comprehension* subtask, females outscored males numerically in three of the four groups, but not Russian Grade 4. However, the only difference that was statistically significant was Russian Grade 2, at the .01

level. Unlike earlier subtasks like FWR, the gaps in performance did not widen between Grades 2 and 4 for this subtask. In fact, the gaps between scores in Grade 4 are smaller for both language groups (see Tables 17 and 18 below).

TABLE 3: READING PASSAGE AND READING COMPREHENSION (GRADES 2 AND 4)

Subtask	Kyrgyz Language					Russian Language				
Gender	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**	**Total**	**Male**	**Female**	**Diff.**	**Cohen's d**
Grade 2	n=658	n=333	n=325			n=324	n=152	n=172		
Reading Passage	40.14 (20.43)	35.45 (18.79)	45.32 (20.95)	9.87**	0.496	43.89 (19.43)	41.34 (20.24)	45.98 (18.54)	4.64	0.24
Reading Comp	56.50 (32.32)	52.99 (32.46)	60.38 (31.76)	7.39	0.23	54.72 (29.83)	49.79 (29.36)	58.76 (29.68)	8.79**	0.304
Grade 4	n=677	n=346	n=331			n=312	n=152	n=160		
Reading Passage	60.24 (23.78)	52.97 (22.32)	67.67 (22.93)	14.7*	0.65	62.35 (26.84)	55.85 (25.34)	68.40 (26.86)	12.55*	0.48
Reading Comp	78.42 (25.37)	76.31 (25.96)	80.58 (24.60)	4.27	0.169	63.11 (32.71)	64.26 (35.04)	62.04 (30.45)	2.22	0.068

"* Significant at the .05 level
**Significant at the .01 level"

TABLE 4: READING RESULTS BY SCHOOL DEMOGRAPHICS

Subtask	Kyrgyz Language				Russian Language			
Gender	**Urban**	**Rural**	**Diff.**	**Cohen's d**	**Urban**	**Rural**	**Diff.**	**Cohen's d**
Grade 2	n=119	n=523			n=54	n=213		
Reading Passage	40.67 (19.54)	39.83 (20.71)	1.29	0.041	49.13 (16.50)	41.80 (20.65)	7.33	0.368
Reading Comp	64.51 (29.71)	53.67 (32.67)	10.84	0.337	65.91 (26.87)	49.06 (29.15)	16.85**	0.587
Grade 4	n=346	n=331			n=152	n=160		
Reading Passage	64.94 (21.51)	57.75 (24.19)	7.19*	0.303	77.11 (23.91)	57.05 (26.10)	20.06*	0.782
Reading Comp	85.06 (20.28)	75.48 (26.86)	9.58*	0.371	70.00 (30.48)	60.64 (33.55)	9.36	0.284

"* Significant at the .05 level
**Significant at the .01 level"

In *Reading Passage* urbanites outscored their rural counterparts on all tests in both languages. Differences for the *Reading Passage* subtasks and were statistically significant at Grade 4 for both groups. Note in particular the score gap on *Reading Passage* for the Kyrgyz Grade 4 pupils. The score gap is 20 words per minute, and the effect size value is the highest recorded from all the subtests, at .782. For Reading Comprehension, the differences were statistically significant for Russian Grade 2 and Kyrgyz Grade 4, with urbanites being favored both times.

LISTENING COMPREHENSION AND DICTATION (GRADES 2 AND 4)

Listening comprehension is a skill that involves hearing and interpreting stimuli to understand facts, construct meaning, and draw inferences. Successful listening comprehension is being able to understand and integrate new meaning with existing

knowledge. In this EGRA subtask, listening comprehension was demonstrated by the pupil by answering several questions from a simple oral story (series of sentences) read aloud by the administrator (an interactive situation). According to O'Maggio (1986), some of the core dimensions of listening are retaining parts of language in short-term memory, discriminating among distinctive sounds, detecting key ideas, and guessing meaning from context.

This subtask determined whether pupils could answer several questions based on a passage read by the administrator. The subtask included a paragraph of approximately 40 words for Grade 2 and 80 words for Grade 4. The test administrator read the passage aloud only once at a pace of about one word per second. When the text was completed, Kyrgyz medium Grade 2 pupils were asked four oral comprehension questions, and Russian medium pupils were asked four questions; for Grade 4 the number of questions was also four for both language groups.

Dictation is a commonly used pedagogical tool in the Kyrgyz Republic and throughout the former Soviet Union (Tvaruzkova & Shamatov, 2012). It is frequently employed to assess listening comprehension as well as writing (reproductive) ability. Pupil ability to hear sounds and correctly recreate the letters and words corresponding to what they hear indicates knowledge of the alphabet and skill in word formation. The Kyrgyz Republic subtask for this particular assessment was adopted from the EGRA main study, the specific design of which has been validated in other contexts (Denton, Ciancio, & Fletcher, 2006). Pupils were marked for spelling, size, symbols, capitalization, punctuation, spacing direction, and accuracy in vowel and consonant sounds. Each category had a total of 2 possible points for completeness, with 1 point for partial credit and 0 points for incorrect answers.

The *Dictation* subtask for Kyrgyz Grades 2 and 4 consisted of 8 and 11 items, respectively, and 9 and 11 items, respectively, in the Russian versions. In Kyrgyz, the maximum possible scores for Grades 2 and 4 were 16 and 22, respectively. In Russian, the maximum possible scores were 18 and 22 for Grades 2 and 4, respectively. Data tables for the subtasks *Listening Comprehension* and *Dictation* by mean total scores at grade level, gender, and differences by school demographics (urban or rural) are presented in Table 19 below.

TABLE 5: LISTENING COMPREHENSION AND DICTATION (GRADES 2 AND 4)

Subtask	Kyrgyz Language					Russian Language				
	Total	Male	Female	Diff.	Cohen's d	Total	Male	Female	Diff.	Cohen's d
Grade 2	n=658	n=333	n=325			n=324	n=152	n=172		
Listening Comp	69.48 (29.44)	70.39 (29.95)	68.48 (28.88)	6.91	0.065	80.94 (27.73)	80.01 (28.33)	81.69 (27.29)	1.68	0.06
Dictation	73.39 (24.93)	68.26 (26.39)	79.08 (21.87)	10.82**	0.446	79.44 (18.29)	78.22 (20.15)	80.44 (16.60)	2.22	0.121
Grade 4	n=677	n=346	n=331			n=312	n=152	n=160		
Listening Comp	66.93 (29.17)	63.98 (29.66)	69.94 (28.39)	5.96*	0.205	70.97 (28.22)	73.93 (27.89)	68.21 (28.34)	5.72	0.203
Dictation	80.34 (20.68)	74.53 (23.20)	86.27 (15.68)	11.74**	0.591	87.85 (15.49)	84.73 (19.24)	90.76 (10.11)	6.03*	0.396

* Significant at the .05 level
**Significant at the .01 level

For *Listening Comprehension* Kyrgyz Grade 2, across both the Russian and Kyrgyz groups, scores were quite similar at Grade 2. Although in Grade 4 some greater numerical differences existed (females better on the Kyrgyz task, males better on the Russian task), there were no statistically significant differences at either grade in *Listening Comprehension*. In Grade 4 males scored lower in the Kyrgyz group but higher in the Russian group, though no differences were statistically significant. Interestingly, in 2014 gender differences were statistically significant for Grade 2 pupils in both groups.

For the *Dictation* tasks, note that females scored at statistically significant higher levels in *three out of the four groups*, with an almost 12-point average difference for the Kyrgyz fourth graders. Only Russian Grade 2 had similarity in results across genders.

Demographic differences on *Listening Comprehension* were statistically significant for three of four subtests, with Russian Grade 4 being the only nonsignificant difference between urban and rural pupils. The effect sizes were large for both Kyrgyz Grade 4 differences and Russian Grade 2 differences. There was also a statistically significant difference between rural and urban pupils for *Dictation* on Kyrgyz Grade 4 (see Table 20).

TABLE 20: BY SCHOOL DEMOGRAPHICS

Subtask	Kyrgyz Language				Russian Language			
	Urban	Rural	Diff.	Cohen's d	Urban	Rural	Diff.	Cohen's d
Grade 2	**n = 119**	n = 523			n = 54	n = 213		
Listening Comprehension	71.83 (28.11)	68.97 (29.87)	2.86**	0.097	93.74 (17.40)	73.82 (29.74)	19.94**	0.719
Dictation	74.74 (25.45)	73.01 (24.79)	1.74	0.07	86.80 (12.53)	75.17 (19.78)	11.63**	0.627
Grade 4	**n = 121**	n = 539			**n = 55**	**n = 201**		
Listening Comprehension	76.59 (25.46)	62.76 (29.86)	13.83**	0.475	72.57 (27.03)	70.81 (26.83)	1.76	0.062
Dictation	84.20 (16.38)	78.53 (22.27)	5.67*	0.266	93.15 (7.85)	85.52 (17.95)	7.63*	0.468

* Significant at the .05 level
**Significant at the .01 level

VI. FINDINGS AND RECOMMENDATIONS

STUDY LIMITATIONS

Before proceeding with an analysis of findings and proposing possible recommendations, some commentary on the study's limitations is in order. The EGRA data in this report has a number of limitations. The first is related to the urbanity categorization of the randomly sampled schools. The disparities in educational achievement across the various demographic regions of the republic have well been studied. On a host of measures, there are large gaps between urban and rural students in educational access and achievement. The categorization of schools into rural, urban and semi urban was determined with data from the MOES. Though a sample weighting calculation was conducted to account for over or underrepresented groups (e.g. school sizes) and population distributions across demographic regions, overall interpretations of the EGRA demographic data should be made with caution, especially in regard to the category of "semi-urban."

The small number of pupils sampled in this category makes firm inferences tenuous at best, especially for the Kyrgyz Grade 2 cohort. While the data on semi-urban pupils was collected, analyses for statistically significance in outcomes by urban-semi-urban or rural-semi-urban were not conducted as results might be misinterpreted. Comparisons were made only between urban and rural schools.

Second, it should be reiterated that this midterm data cannot be used to draw inferences about national reading levels. The data are not nationally representative and present a much smaller sample size than the 2014 baseline data collection. 2015 data was collected only from the Cohort 1 regions of the country, Bishkek, Chui, Talas, and Jalal-Abad, where the intervention rollout began and school-level teacher training was almost complete at the time of data collection.

Third, the data in this report cannot be used to attribute change to the USAID Quality Reading Project intervention. The forthcoming impact evaluation report will provide a thorough analysis of the data that compares treatment schools to control schools. By making use of the randomized controlled trial design, we will be able to determine whether there is change in scores attributable to the effect of the intervention. That analysis is not present in this midterm report.

FINDINGS

There is evidence of improved performance on EGRA 2015 from the 2014 baseline. Scores on 3 of the 4 comparable subtasks for Kyrgyz language pupils in the 2nd Grade increased. The increases ranged from the smallest at +3.8 percentage points for *Initial Letter Sound* to the largest at a +8.66 letters per minute on *Letter Name Recognition*. On the *Dictation* subtask, scores decreased by 3.7 percentage points. For the Russian language (Grade 2) there were substantial increases in score on the *Familiar Word Recognition*, at + 16.03 words per

minute. In Grade 4 (both languages) there were substantial score increases on the Familiar Word Recognition subtasks, + 17.01 for the Kyrgyz cohort and 22.72 for the Russian cohort. It is important to reiterate that these scores presented above represent equated subtasks put on a comomon scale.

Yet, despite some of these increases, according to the data now available, the overall state of affairs with reading outcomes remains somewhat unchanged from previous years. The overall results of the Midterm EGRA 2015 are similar to previous EGRA iterations in fundamental ways (2012, 2014). Approximately half of Kyrgyz Grade 2 pupils are attaining the standard for reading fluency and just over half are attaining that level in Russian Grade 2. These numbers decline in Grade 4 to well below half. While some of the EGRA tasks remain relatively easy, with respondents mean score averages across gender and demographics at over 90%, in the key areas of letter and word identification, and word and passage reading, all subtasks related directly to reading fluency and comprehension, the results are not yet promising. There are both large (and increasing gaps) by gender and between urban and rural pupils. Note in particular the urban-rural score gap on *Reading Passage* for the Kyrgyz Grade 4 pupils. The score gap is *20 words per minute* and the effect size value is the highest recorded from all the subtests, at .782.

THE FOCUS ON GENDER

A clear pattern that emerges from these EGRAs results in relation to gender is that females outperformed males on a large number of tasks, and that *these gaps appear to be increasing over time*. On EGRA 2015 subtasks, females *numerically* outscored males on 25 of the subtasks across the 32 EGRA tasks: On 8 of 9 (Kyrgyz 2), 7 of 9 (Russian 2), 6 of 7 (Kyrgyz 4), and 4 of 7 (Russian 4). Independent sample t-tests of statistical significance of these differences revealed statistically significant differences in score by gender (null hypothesis of "no difference") on 22 of the 32 subtasks. Of these results, three had very small effect sizes, i.e. below .01 which makes the differences small in practical terms (Cohen, 1992). Of the remaining 19 subtasks, only 1 statistically significant gap favored males while 18 favored females.

While all of the effect sizes in EGRA were either negligible, small or in the moderate to large range, the largest effect values were for tasks related to reading, word recognition and dictations; in the middle of the "moderate effect range" at .55, .57, .59, and .65.[13] When examining the data by strength of effect size, four of the largest effect size values were associated with gender gaps *for the Kyrgyz cohorts*, with highest values concentrated in Grade 4. These subtasks were *Familiar Word Recognition*, *Nonsense Words*, *Reading Passage* and *Dictation.* These four subtasks also had the highest effect size values from all EGRA subtests. On the *Reading Comprehension* subtask, while only 2 of the 4 tests showed gender differences, both were in favor of females.

[13] Recall the interpretation of Cohen's D effect size coefficient range: small (.2 - .5), medium (.5 to .8), and anything above .8 is considered to be large.

TRENDS IN DATA (KYRGYZ)

Table 21 below vividly demonstrates three important findings in regard to the Kyrgyz cohorts from the EGRA 2015. First, note that females were overwhelmingly favored when there were differences in mean scores by gender. The green color-coded bars indicate subtasks where the differences between males and females were statistically significant, and had an effect size of at least .2. Note that in Grades 2 and 4 for the Kyrgyz cohorts, the same four subtasks were significant in both Grades 2 and 4 of EGRA.[14] Second, note the Cohen's d values in the columns on the right side of the charts (Grade 4). As noted above, these values indicate the strength of the significance. Note that in all four cases, the *strength of the relationship is stronger in Grade 4* with the largest change on the *Reading Passage*, from .39 in Grade 2 to .65 in Grade 4. If all these d values were in the "lower part of the small range" in Grade 2, by Grade 4 most were pushing the "moderate" effect size ranges. Again, the four subtasks with significance in both Grades 2 and 4 were: *Familiar Word Recognition*, *Unfamiliar Word Recognition* (2014)/*Nonsense Words* (2015), *Reading Passage* and *Dictation*, a set of subtasks that demands similar skills and capabilities.

TABLE 21: KYRGYZ LANGUAGE GRADES 2 AND 4 (GENDER)

Kyrgyz Grade 2						Kyrgyz Grade 4					
	Mean		Score Difference	Sig?	D		Mean		Score Difference	Sig?	D
	M	F					M	F			
LNR	67.00	73.10	-6.10	Yes	0.35	LNR					
ISL	94.30	96.30	-2.00		0.20	ISL					
FWR	52.38	62.38	-10.00	Yes	0.38	FWR	80.25	98.01	-17.76	Yes	0.55
UFR	23.99	28.20	-4.21	Yes	0.39	UFR	30.37	37.61	-7.24	Yes	0.58
OV	88.19	88.86	-0.67		0.05	OV	91.81	89.55	2.26	Yes	0.20
RPR	35.45	45.32	-9.87	Yes	0.39	RPR	52.97	67.67	-14.70	Yes	0.65
RPQ	52.99	60.38	-7.39		0.23	RPQ	76.31	80.58	-4.28		0.17
LCQ	70.39	68.48	1.91		0.07	LCQ	63.98	69.94	-5.95		0.21
DICT	68.26	79.08	-10.82	Yes	0.45	DICT	74.53	86.27	-11.74	Yes	0.59

Legend:

Girls Favored

Boys Favored

Significant with at least small or moderate effect size

Cohen's D		effect size
small	=	0.2
moderate		0.5
large	=	0.8

TRENDS IN DATA (RUSSIAN)

The trend over time is less clear for the Russian data (below) than it is for the Grade 2, Kyrgyz cohort. Unlike the trends of the Kyrgyz cohorts, there is less of a pattern in terms of which subtasks had statistically significant differences in gender means across grade levels. As with the Kyrgyz cohort however, the effect size *is larger in Grade 4 than in Grade 2*. What this could be indicating, though still a somewhat tentative inference due to the relative

[14] *Letter Recognition* (LNR) was conducted only for Grade 2, but it was significant in that year.

small level of the effect sizes overall, is that these gaps are growing over time. This finding would be consistent with "the learning drop off" that Hirsch (2003) and others have noted in their research in terms of a "slow down" in learning by that time, especially for males.

TABLE 22: RUSSIAN LANGUAGE GRADES 2 AND 4 (GENDER)

Russian Grade 2						Russian Grade 4					
	Mean		Score Difference	Sig?	D		Mean		Score Difference	Sig?	D
	M	F					M	F			
LNR	53.01	60.41	-7.41		0.38	LNR					
ISL	94.39	94.36	0.02		0.00	ISL					
FWR	62.73	67.27	-4.54		0.20	FWR	85.13	103.07	-17.94	Yes	**0.55**
UFR	28.22	32.50	-4.28	Yes	**0.39**	UFR	30.56	34.65	-4.09		0.35
OV	84.02	83.97	0.04		0.00	OV	88.56	85.44	3.12		0.22
RPR	41.34	45.98	-4.64		0.24	RPR	55.85	68.40	-12.55	Yes	**0.48**
RPQ	49.79	58.76	-8.96	Yes	**0.30**	RPQ	64.26	62.04	2.23	Yes	0.07
LCQ	80.01	81.69	-1.68		0.06	LCQ	73.93	68.21	5.72		0.20
DICT	78.22	80.44	-2.23		0.12	DICT	84.73	90.76	-6.04	Yes	**0.40**

Legend:

- Girls Favored
- Boys Favored
- Significant with at least small or moderate effect size

Cohen's D		effect size
small	=	0.2
moderate	=	0.5
large	=	0.8

In the Russian cohorts, girls were favored again in more cases, but not as exclusively as with the Kyrgyz cohort. Again, effect sizes for differences were higher in Grade 4 than they were in Grade 2.

Contrast the above findings for the tasks related to word identification, reading and dictation with both the *Oral Vocabulary* and *Listening Comprehension* subtasks: There was only one *statistically significant gender differences* (with practical significance, effect size at least .01) for either grade, in either language, on either of these subtasks. For the *Oral Vocabulary* subtask (where respondents listen and point to a correct picture from a choice of four pictures) males numerically outperformed (just slightly) females at both grades and in both languages, though these numerical differences were small and not statistically significant. These tasks required no reading and minimal writing, and perhaps this explains why males performed on par with females on these two subtasks.

The data indicates a growth in the gender gap from Grade 2 to Grade 4, and a gap that was larger for the Kyrgyz cohort. These overall findings on gender are consistent with both the 2012 EGRA in the Kyrgyz Republic and the baseline USAID Quality Reading Program EGRA in 2014 that found males falling behind in reading and writing (Tvaruzkova & Shamatov, 2012; EGRA Baseline Analytical Report, 2014). USAID and other international donors have focused on identifying and rectifying performance gaps by gender throughout the world in recent decades. The emphasis in many countries has appropriately been placed on closing access and achievement gaps that favor males over females. However, caution

should be taken in extrapolating results from neighboring regions, as country context and empirical trends in the in-country data should determine the trajectory of policy recommendations. In countries of the former Soviet Union, the data from a host of sources has consistently indicated that it is females, not males, who have been outperforming their counterparts in both achievement and educational access (Drummond & De Young, 2004; De Young et al., 2006; Bruner & Tillet, 2007; CEATM, 2009; CEATM, 2010; Drummond, 2011).

Important questions to consider at the policy level relate to how the republic is going to address the gaps and improve reading and other educational outcomes for males, starting in the early grades. It is essential to address the issue of poor reading early on in order to avoid pupils falling further and further behind as they progress throughout school. The data on these gaps and the results for Kyrgyz boys in particular would not necessarily warrant urgent measures if they only indicated a temporary lag from which recovery in later grades was possible. However, the available data on educational attainment by gender and demographics in the republic indicate the situation for Kyrgyz boys does not improve over a life of schooling. This question is interconnected with demographic issues which are presented below.

DEMOGRAPHICS

The data indicate that the gaps between urban and rural schools are wide, though not at all levels. In Kyrgyz Grade 2, for example, urban pupils outscored their peers on only 5 of the 9 subtasks. Further, effect size values for most of these statistical tests indicate that the practical differences were negligible in several cases. This fact contrasts with the Russian rural scores where effect sizes for statistically significant differences were much higher.

Below we can also follow the trend over time in terms of urban and rural divides for the Kyrgyz cohort. From left to right, the charts list the name of the subtask, mean scores of urban and rural pupils, the difference in those scores, whether those differences are statistically significant, and the effect size value, Cohen's D. Notice that more subtasks have significant differences in Grade 4 indicating growing cleavages in the later period. The number of statistically significant differences overall was lower and the effect size values were lower, however, than they were for the gender variable and there was more balance between which group was favored than for the gender factor for the Kyrgyz group.

TABLE 23: KYRGYZ LANGUAGE GRADES 2 AND 4 (DEMOGRAPHICS)

Kyrgyz Grade 2						Kyrgyz Grade 4					
	Mean		Score Difference	Sig?	D		Mean		Score Difference	Sig?	D
	Urban	Rural					Urban	Rural			
LNR	63.57	72.30	-8.73	Yes	**0.51**	LNR					
ISL	93.26	96.02	-2.76		0.28	ISL					
FWR	54.99	57.74	-2.75		0.10	FWR	90.95	87.61	3.34		0.10
UFR	26.26	25.82	0.43		0.04	UFR	35.52	33.03	2.49	Yes	0.19
OV	87.25	89.05	-1.79		0.14	OV	88.03	91.76	-3.72		0.34
RPR	40.67	39.83	0.84	Yes	0.04	RPR	64.94	57.75	7.19	Yes	**0.30**
RPQ	64.51	53.67	10.84	Yes	**0.34**	RPQ	85.06	75.48	9.58	Yes	**0.37**
LCQ	71.83	68.97	2.86	Yes	0.10	LCQ	76.59	62.76	13.83	Yes	**0.48**
DICT	74.74	73.01	1.73		0.07	DICT	84.20	78.53	5.67		0.27

Legend:

Urban Favored

Rural Favored

Significant with at least small or moderate effect size

Cohen's D		effect size
small	=	0.2
moderate	=	0.5
large	=	0.8

TABLE 24: RUSSIAN LANGUAGE GRADES 2 AND 4 (DEMOGRAPHICS)

Russian Grade 2						Russian Grade 4					
	Mean		Score Difference	Sig?	D		Mean		Score Difference	Sig?	D
	Urban	Rural					Urban	Rural			
LNR	58.17	56.52	1.65		0.078	LNR					
ISL	96.27	93.63	2.64		0.182	ISL					
FWR	72.91	61.59	11.32	Yes	0.489	FWR	110.05	88.47	21.58	Yes	0.65
UFR	33.55	29.31	4.24		0.379	UFR	37.04	31.02	6.01	Yes	0.52
OV	84.66	83.87	0.79		0.055	OV	86.37	87.64	-1.27		0.09
RPR	49.13	41.80	7.33		0.368	RPR	77.11	57.05	20.06	Yes	0.78
RPQ	65.91	49.06	16.85	Yes	0.587	RPQ	70.00	60.64	9.35		0.28
LCQ	93.74	73.82	19.92	Yes	0.719	LCQ	72.57	70.81	1.76		0.06
DICT	86.80	75.17	11.63	Yes	0.627	DICT	93.15	85.52	7.64	Yes	0.47

Legend:

Urban Favored

Rural Favored

Significant with at least small or moderate effect size

Cohen's D		effect size
small	=	0.2
moderate	=	0.5
large	=	0.8

On the Russian Grade 2 subtasks, urbanites outscored their rural counterparts on every subtask and by wide margins. Only on three subtasks were gaps not evident *Letter Name Recognition*, *Initial Letter Score*, and *Oral Voc*abulary. Some of the highest effect sizes were found on Russian Grade 2 with *Listening Comprehension* at .719. On the *Reading Passage* subtask for Russian Grade 4, the differences in words per minute were just over 20, effect size at .78. Urban and rural differences at Grade level 4 on the Russian subtasks were also statistically significant (in favor of urbanites) on all but the *Oral Vocabulary* subtask.

It has been noted throughout the midterm report that comparisons across the Kyrgyz and Russian language groups is not an appropriate use of the data: Learning to read in the Kyrgyz and Russian languages are two entirely different processes, and differences in rates of acquisition, knowledge accumulation, and effective use are very much a function of the language properties in question as well as the socio-cultural context of the learner. EGRA results by language should be considered independently of each other.

Yet, EGRA results do raise interesting questions for deeper consideration in regard to language of instruction in multi-lingual societies such as the Kyrgyz Republic. In the Kyrgyz Republic there are large achievement gaps by language of instruction in later years of schooling (favoring Russian cohorts, plausibly explained by demographic factors and selectivity bias as Russian tracks send a greater proportion of pupils on to higher education (CEATM, 2010). The apparent relative parity by language in terms of "where pupils start" or "how fast they acquire skills" as indicated by the EGRA data is interesting when one considers the difference in ultimate educational outcomes across language tracks. Considering what is known about the importance of the early years – parental effects, pre-school preparation, parents reading to children, one might expect to see greater disparities by language in some EGRA subtests by rates of acquisition (i.e. Russian language cohorts acquiring skills at faster rates). Perhaps *rate/time* spent on reading acquisition has much less of a relationship to long term educational outcomes than some might suppose.

FOR POLICY DISCUSSION

The USAID Quality Reading Project does not have a mandate to propose specific, suggested program interventions (in addition to those already ongoing in the republic) within the scope of this report. The purpose of the midterm report is to present and analyze the new 2015 EGRA data. However, based on the data presented there is a warrant for raising some related questions for policy makers to consider when thinking about how to improve reading outcomes. It seems logical to suppose that special attention should be paid to those areas of particular weakness (large gender gaps in reading fluency, word familiarity, reading comprehension) and at particular levels (Grade 4). These questions should be considered at the policy level as only strong instructional leadership in a highly centralized education system will likely result in progress over time.

The purpose of this section is not to suggest a list of policy or program interventions to the MOES. Simply suggesting "adding more programs" or interventions based on the data presented is not possible and would be counter-productive. Careful investigation is needed to avoid the ill-advised practice of "piling on new stuff" which is as likely to result in incoherence and burdening of teachers and instructional leaders as bringing about real change. However, as reading is essential to a child's growth and development in all areas, this third wave of EGRA results in the Kyrgyz Republic could serve as a means to mediate important conversations about the need for more focus on outcomes, especially in areas noted above. The questions framed here are in the areas of *focus and resource allocation*, *information and progress monitoring*, and *creating a reading culture*.

FOCUS AND RESOURCE ALLOCATION

Some key questions to consider in policy dialogue about focus and resource allocation include:

- How does the MOES currently signal to oblast and local educational administrative units that the focus on reading is a new national priority? Are there low cost opportunities to enhance the strength and frequency of such signaling?

- How does the MOES currently work with instructional leadership in schools to coordinate reading initiatives and programs that are a national priority? How could the current communications regime in this regard be improved so that there is more school-level ownership of national reading priorities?

- Are there currently ongoing discussions about improving instructional practices and adjusting time and resources allocated towards implementing change? If not, why not? How can obstacles to reform in this area be overcome?

- Has the MOES considered introducing instructional practices that might be new to the republic in order to focus more on reading outcomes, and specifically to differentiate in approach by gender? (E.g., differentiated instruction, more individual pupil reading time, allocation of resources and materials to students based on reading levels, use of reading circles, exposure to content "outside the curricula" that may increase motivation, and other creative classroom norms that promote interest in reading, especially for boys).

- What are the administrative barriers that hinder the reallocation of more time and resources towards supporting both increased dialogue about reading instruction and creation of enhanced teaching and learning materials? Can some of these barriers be reduced or removed? How?

- In a system where reading is not taught as a "stand alone subject" in schools, how might the MOES create additional space on the national curricula for inclusion of more themes, content, materials and activities with the intent of improving reading outcomes?

- How can professional development programming become more focused on increasing student reading outcomes? Can some, non-priority types of professional training be reduced in *lieu* of more concentration on the professional development of instructional leaders in reading?

- How does the MOES currently provide incentives for teachers to select professional development in the realm of reading instruction?
- Are there opportunities to create new incentives to get more buy in from teachers to focus on reading fundamentals at the early grades?
- Has the MOES considered systemic changes in the way reading is currently taught in schools? How are any plans being implemented?
- How can teacher preparation programs better ensure that all teachers learn the fundamentals of teaching pupils to read and write, based on the latest research?

PROGRESS MONITORING AND ALIGNMENT

Understanding the state of reading levels in the early grades in the republic is of course the first step towards improving reading instruction and outcomes. It is essential to understand what is happening in this area and to monitor progress at all levels and to calibrate and adjust interventions and supports as necessary. It also essential that any initiatives, reforms, or any proposed changes to the status quo in any one area in the system be tightly aligned with other parts of the system. For questions for discussion:

- Do regional and local authorities have the mandate and resources to collect data and information on the activities and initiatives in their communities in regard to improving reading instruction, including access to pupil level data?
- Does the MOES (or local authorities) monitor which teachers are engaged in professional development for reading improvement? Are there ways to leverage the strengths of well-trained *kadres* so that they may serve as a low cost resource for other teachers who might need mentoring or training support?
- Are school directors currently required to monitor progress on reading outcomes and submit information collected through the normal ministerial channels? Could they be? Can the MOES invest more in capacity building to be able to manage new information flows, communication flows, and data system development?
- How are school directors currently ensuring that the professional development needs of their teachers in regard to reading outcomes are being met?
- How is formative assessment in the early grades currently conducted to ensure continuous reading improvement? What are the training needs in this area?
- How are textbooks, which constitute "the curriculum" in many schools, addressing or not addressing skill in reading?
- How does the MOES ensure that best practices in reading instruction are maintained throughout the system?

FOSTERING A READING CULTURE

The USAID Quality Reading Project has taken vigorous steps to promote the "culture of reading" through public activities such as reading circles and other events and activities to encourage reading in the broader community. This holistic approach is important as reading outcomes can be dramatically improved if reading is an important home and community activity.

- How can MOES and local authorities scale up such grass roots activities so they remain sustainable after USAID Quality Reading Project is complete? Can structural incentives such as teacher education credits, honoraria, or other incentives be provided to encourage cost share of similar activities at the community level in the future?
- How can teachers devote more time towards instruction and assessment of reading outcomes without sacrificing too much the teaching of core "content knowledge" as currently defined?
- How can policy makers better ensure that what is currently being taught is tightly aligned with the new national expectations on reading?

APPENDIX 1. KYRGYZ ALPHABET (JUSAYEVA, 2004)

А а	Б б	(Вв)	Г г	Д д	Е е	Ё ё	Ж ж	З з
а	бе	ве	ге	де	е	ё	же	зе
a	b	v	g	d	e	yo	j	z
[ɑ]	[b/v/w]	[v]	[g~ɢ/ʁ]	[d]	[je/e]	[jo]	[ʤ]	[z]

И и	Й й	К к	Л л	М м	Н н	Ң ң	О о	Ө ө
и	ий	ка	эл	эм	эн	ың	о	ө
i	y	k	l	m	n	ñ	o	ö
[i]	[j]	[k~q/χ]	[l~ɬ]	[m]	[n]	[ŋ~ɴ]	[o]	[ø]

П п	Р р	С с	Т т	У у	Ү ү	(Фф)	Х х	(Цц)
пе	эр	эс	те	у	ү	эф	ха	це
p	r	s	t	u	ü	f	h	c
[p]	[r]	[s]	[t]	[u]	[y]	[φ]	[χ/q]	[ʦ]

Ч ч	Ш ш	(Щщ)	(Ъъ)	Ы ы	(Ьь)	Э э	Ю ю	Я я
че	ша	ща	ажыратуу	ы	ичкертүү	э	ю	я
ç	ş	şç	белгиси	ı	белгиси	é	yu	ya
[ʧ]	[ʃ]	[ʃʧ]	”	[ɯ]	’	[e]	[ju]	[ja]

Long vowels

аа	ээ	ии	оо	өө	уу	үү
aa	ee	ii	oo	öö	uu	üü
[ɑ:]	[e:]	[i:]	[o:]	[ø:]	[u:]	[y:]

APPENDIX 2. TRAINING FOR EGRA ADMINISTRATORS

In 2014 and 2015 the USAID Quality Reading Project trained supervisors and test administrators on one-to-one EGRA administration procedure, and how to record pupils' oral responses into scanned forms. The team conducted training of test administrators through a two-step cascading process: an international consultant conducted 1-day trainings for supervisors and project staff, and the supervisors conducted a 3-day training in their respective regions for test administrators. These training workshops trained a total of eight supervisors and 164 test administrators.

During the training, the test administrators practiced school-level sampling and test administration procedures. In preparation for various possible scenarios for school environments in Kyrgyzstan, test administrators practiced drawing the pupil sample by completing the sample selection forms and calculating the sample intervals to select the necessary 20 pupils for each grade. The final part of the test administrators' training audited the roles and responsibilities of the test administrators, team supervisors, and USAID Quality Reading Project office as explicitly described in the administration manual.

APPENDIX 3: 2015 EGRA ADMINISTRATION AND MONITORING

The 68 EGRA subtask administrators were deployed to collect data in the 60 schools (30 treatment and 30 control) in Cohort 1 regions: Talas, Jalal-Abad, Bishkek, and Chui. Each team of four administrators was responsible for administering the assessment in 3-4 schools. Data collection commenced on April 13 and was completed on 28, 2015. During the data collection, the administrators reviewed their work each evening before returning all document to their team leaders. The team leaders then checked all blanks again and submitted them to the regional supervisor; regional supervisors, in turn, were instructed to review the instruments thoroughly.

Many of the 2015 EGRA administrators gained experience during the baseline data collection in 2014. New administrators where allocated to different teams to ensure they had access to more experienced team members for support. All EGRA administrators were deployed following their respective rounds of training (4 days), resulting in a staged rollout of the EGRA administration to ensure timely completion. The National Testing Center (NTC) supervisors (assigned to regions) and project staff was mobilized to conduct monitoring visits to ensure proper administration of the assessment and to provide support with troubleshooting as necessary.

REFERENCES

American Institutes for Research. (October, 2014). *USAID Quality Reading Project: Kyrgyzstan: Early Grade Reading Assessment (EGRA) Baseline Data Report.* United States Agency for International Development.

Brunner, J. & Tillet, A. (2007). *Higher education in central asia: The challenges of modernization (case studies from Kazakhstan, Tajikistan, the Kyrgyz Republic and Uzbekistan)*. Washington, DC: The International Bank for Reconstruction and Development/TheWorld Bank.

Chiappe, P., Siegel, L.S., L Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific studies of reading* 6 (4). P. 369-400.

Cizek, G., & Bunch, M. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

CEATMa (2009). *PISA 2009: International survey of reading literacy for 15 years old,* retrieved from http://www.testing.kg/ru/projects/pisa2009/

CEATMb (2009). *Natsional'noye otsenivanie obrazovatel'nix dostizhenii uchashixsya.* [National Assessment of Educational Quality]. Bishkek: CEATM. Retrieved from http://www.testing.kg/ru/projects/NSBA2009/

CEATM (2010). *Rezul'tati obsherespublikanskova testirovaniya i zachisleniya na grantovie mesta vuzov goda Kirgizskoi Respubliki v 2010 godu*. [Results of national scholarship testing and enrollment in university grant places in the Kyrgyz Republic in 2010]. Bishkek: CEATM. www.testing.kg

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Denton, C. Ciancio, D..; Fletcher, J. (Jan -March 2006). Validity, Reliability, and Utility of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly*, v41. n1: p8-34.

De Young, A., Reeves, M., & Valyaeva, G. (2006). *Surviving the transition? Case studies and schooling in the kyrgyz republic since independence*. Greenwich, Connecticut: Information Age Publishing.

Drummond, T. & De Young, A. (2004). Perspectives and problems in education reform in kyrgyzstan: The case of national scholarship testing 2002. In S. Heyneman & A. De Young (Eds.), *The Challenge of Education in Central Asia* (pp. 225-242). Greenwich, CT: Information Age Publishing,

Drummond, T. (2011). *Predicting differential item functioning in cross-lingual assessments: The case of a high stakes admissions test in the Kyrgyz Republic*. Ph.D. dissertation, Michigan State University. East Lansing, MI.

Fuchs, L., Fuchs,D., Hosp, K., & Jenkins, J. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading.* Volume 5, Issue 3, 2001. pg. 239-256

Gove, A. (2009). *Early grade reading assessment toolkit*. RTI International.

Grenoble, L. (2003). *Language policy in the Soviet Union.* Boston, MA: Kluwer Academic Press.

Hirsch, E.D. (2003). Reading comprehension requires knowledge of words and the world: Scientific insights into the fourth-grade slump and the nation's stagnant comprehension scores. *American Educator, Spring.* Retrieved from http://www.aft.org/sites/default/files/periodicals/Hirsch.pdf

Hu, Z. & Imart, G. (1989). *A kirghiz reader*. Bloomington, IN: Research Institute for Inner Asian Studies.

Huskey, E. (1995). The politics of language in kyrgyzstan. *Nationalities Papers, 23(1),* 549-572.

Jusayeva, V. (2004). *Kyrgyz Grammar Supplement*. U.S. Peace Corps Kyrgyz Republic. Bishkek, Kyrgyz Republic.

Korth, B. (2004). Education and linguistic division in kyrgyzstan. In S. Heyneman and A. De Young (Eds.), *The Challenge of Education in Central Asia* (pp. 97-112). Greenwich, CT: Information Age Publishing.

Korth, B. (2005). *Language attitudes towards kyrgyz and russian: Discourse, education and policy in post-soviet Kyrgyzstan.* Bern: Peter Lang.

Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational occupational tests.* Princeton, NJ: Educational Testing Service.

Loomis, S. C., & Bourque, M. L. (2001). From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 175–217). Mahwah, NJ: Erlbaum.

National Institute of Health (2006). *Report on the findings of the national reading panel*. Retrieved from: https://www.nichd.nih.gov/publications/pubs/nrp/Pages/findings.aspx

National Statistics Committee (2000). *Obrazovanie v kyrgyzskoi respublike.* {Education in the Kyrgyz Republic}. Bishkek: Kyrgyzstan.

National Statistical Committee of the Kyrgyz Republic (2010). Population and Housing Census of the Kyrgyz Republic of 2009. Bishkek: Kyrgyzstan.

O'Maggio, A. (1986). A proficiency-oriented approach to listening and reading. In A. O'Maggio (Ed.), Teaching Language in Context, (pp. 121-174). Boston, MA: Henile & Heinle.

Perie, M. (2008). A guide to understanding and developing performance-level descriptors. *Educational Measurement: Issues and Practices 27*(4), 15–29.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.; pp. 221–262). New York, NY: Macmillan.

Plake, B. S., & Ferdous, A. A. (2005). *Setting multiple performance standards using the yes/no method: An alternative item mapping method*. Paper presented to the meeting of the National Council on Measurement in Education, Montreal, Canada.

The Organization for Economic Cooperation and Development. (2010). *Kyrgyz Republic 2010: Lessons from PISA, OECD Publishing.* Retrieved from http://www.keepeek.com/Digital-Asset-Management/oecd/education/reviews-of-national-policies-for-education-kyrgyz-republic-2010_9789264088757-en#page1

United Nations Children's Fund (UNICEF).(2005). *Monitoring learning achievement: Nationwide study of the quality of education in primary schools*.

Oruzbaeva, B. (1997). *Kirgizskii Yazyk: Yazyki Mira, Turkskie Yazyki.* [Kyrgyz Language: Languages of the World, Turkic Languages]. Bishkek: Izdatel'skii Dom.

Shamatov, D. & Niyozov, S. (2010). Teachers surviving to teach: Implications for post-soviet education and society in tajikistan and kyrgyzstan. In J. Zajda (Ed.), *Globalization,*

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. British Journal of Psychology, 94, 143–174.

Tvaruzkova, M. & Shamtov. D. (2012). *Review of early grade teaching and skills: The Kyrgyz Republic and Tajikistan*. Aguirre Division of JBS International, Inc. with support from the United States Agency for International Development (USAID).

Wright, S. (1999). Kyrgyzstan: The political and linguistic context. *Current Issues in Language & Society, 6(1),* 85-91.

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, DC 20523
Tel: (202) 712-0000
Fax: (202) 216-3524
www.usaid.gov