AFGHAN CHILDREN READ

EARLY GRADE READING ASSESSMENT IN DARI LANGUAGE CONDUCTED IN THE PROVINCE OF HERAT APRIL 2018

REVISED MIDLINE REPORT

MAY 31ST, 2019

Submitted to:

Creative Associates International

COR: Abdul Alim Ghafary

Contract No. 306-TO-16-00003

Submitted by:

Creative Associates International

Submission Date: May 31, 2019

TABLE OF CONTENTS

List of Figures

List of Tables

Acronyms

ACR	Afghan Children Read
AIR	American Institutes for Research
BRAC	Building Resources Across Communities
CARE	Cooperative for Assistance and Relief Everywhere
CBE	Community Based Education
CRS	Catholic Relief Services
DED	District Education Directorate
EGR	Early Grade Reading
EGRA	Early Grade Reading Assessment
EMIS	Education Management Information System
FWR	Familiar Word Reading
GPS	Global Positioning System
IWR	Invented Word Reading
LC	Listening Comprehension
MoE	Ministry of Education
NGO	Non-Governmental Organization
ORF	Oral Reading Fluency
PA	Phonemic awareness
PED	Provincial Education Directorate
PPS	Probability Proportionate Sampling
RC	Reading Comprehension
SSI	Syllable Sound Identification
UNICEF	United National International Children's Emergency Fund
USAID	United States Agency for International Development

INTRODUCTION

This document provides the final report of the midline Early Grade Reading Assessment (EGRA) for Dari language conducted in the Province of Herat by Creative Associates' *Afghanistan Children Read* (ACR) project in April 2018. The midline EGRA was administered to Grade 2 students, while the baseline conducted in April 2017 targeted Grades 2 and 3 students. This means that comparisons of achievement from 2017 to 2018, baseline to midline, are limited to Grade 2. Note that the design of the ACR study is not student cohort-based but grade-based. Thus, comparisons from baseline to midline discussed in this report reflect how a targeted grade (Grade 2), composed of different randomly selected students, has progressed from one year to the next, in the context of teachers who have received professional development interventions under ACR's study during the same period.

Creative had expected that this final report would have included results of the midline EGRA corresponding to Pashto language in the Provinces of Nangarhar and Laghman. This would have provided a comparison of progress in student achievement from baseline to midline relative to two key national languages, the target of ACR's early grade reading work in Afghanistan. Unfortunately, due to circumstances beyond the control of Creative and the ACR project and related to an inability to collect midline EGRA and context survey data in the targeted Pashto-speaking locations for reasons of security, this report will be limited to the following elements:

1. EGRA Grade 2 midline results and interpretation of results for Dari in the Province of Herat with comparisons from baseline to midline (for Grade 2 only). These results are presented in 2 forms: 1) Mean scores per EGRA sub-task by whole group and sub-group; and 2) By reference to the performance standards and benchmarks established during the standard-setting and benchmarking workshop held in Kabul in March 2018, i.e., as percentages of students per performance level.
2. Results of analyses of associated factors derived from student, teacher, and school principal surveys crossed with student performance data from the midline Dari Grade 2 EGRA; comparisons of results of associated factors analyses are provided from baseline to midline for Dari.
3. Results of multiple regression analyses for the midline EGRA conducted for Dari Grade 2 together with comparisons of baseline and midline results (for Grade 2 only).

This second version of the midline EGRA report was requested by USAID after discussions had taken place during the dissemination of the midline results at a workshop and follow-up meetings with clients in Kabul during the month of February 2019. During this meeting it was seen that there was a significant

discrepancy between the results obtained in public schools contrasted with those obtained in community-based education (CBE) centers to the point where gains made by public schools in student reading outcomes were obscured in the aggregated results. CBE centers are located uniquely in rural areas of Afghanistan and are exposed to greater social and educational instability. In addition, sample numbers for CBEs were significantly lower than those for public schools (see Table 3, Student Sample Sizes per Grade per Baseline and Midline) rendering their results difficult to interpret.

In this report, we focus on reporting student reading performance for the province of Herat (Dari language) based on public schools only, with a small section at the end addressing basic comparisons between public school results and those of CBEs (please see the chapter titled: Baseline vs. Midline EGRA results in Public Schools for the Province of Herat). This revised section on student performance also addresses the results of multiple regression analysis, correlating student performance with background factors. Those results addressing background factors posited as having a relationship with student performance achieved through univariate analysis can be found in the Appendices 3 and 4. These have been significantly simplified from those presented in Version 1 of this report, focusing on only the most relevant findings to accounting for reading achievement.

For those readers who wish to refer to the original midline report comparing April 2017 baseline with April 2018 midline results for Dari language Grade 2, please refer to the "Early Grade Reading Assessment in Dari Language Conducted in the Province of Herat, April 2018: Final draft midline report, submitted to the clients on December 4th, 2018.

Afghanistan Country Context

Afghanistan has made significant progress in creating sustainable governing institutions over the past 15 years thanks to the efforts of the Afghan people working in collaboration with their international partners. However, while much has been accomplished, many challenges remain. According to the World Bank, almost 40 percent of Afghans still live in poverty. Nearly one-third of the population is food insecure and many families lack access to basic services according to data from the United Nations World Food Program. Thousands of children suffer from chronic malnutrition and lack access to quality education. The United States Agency for International Development (USAID) seeks to promote stable and inclusive institutions and an economically prosperous country. Going forward, USAID's strategy for meeting development goals calls for sustaining achievements in education, health, and for women; continuing to

stimulate agriculture-led economic growth and fiscal sustainability; and supporting legitimate and effective Afghan governance (USAID, https://www.usaid.gov/afghanistan/education).

Afghanistan Education Context

Improving the quality of education and increasing basic literacy remains a priority for the Afghan government. Overall, literacy rates (defined as those aged 15 and above who can read and write) is just 38.2%, with males at 52% and females at 24.2% (2015 estimations, CIA Fact Book). Complicating the challenge to improving literacy rates is the fact that some students at primary levels are studying in languages of instruction that are not their home language. To address these challenges, USAID is working with the Ministry of Education (MoE) to improve access to education and reading in the early grades through Afghan Children Read (ACR).

Objectives

Funded by USAID, the objectives of Afghan Children Read are to improve access to education and to improve reading outcomes for primary students in Afghanistan. Afghan Children Read supports schools in the formal education system as well as Community Based Education (CBE) schools. The project will achieve its objectives by bolstering the capacity of the government to provide and sustain an evidence-based early grade reading (EGR) curriculum in Dari and Pashto for Grades 1 to 3. An analysis of the results of Afghan Children Read's EGR model for interventions will enable sound policy and targeted support to sustain growth in literacy levels in both public schools and communities where CBE predominates.

ACR conducted an Early Grade Reading Assessment (EGRA) baseline in three pilot provinces (Herat for Dari language, and Nangarhar and Laghman for Pashto language) in order to generate actionable data on the state of reading outcomes. The report that follows presents an analysis of the results of the EGRA Grade 2 midline conducted in April 2018 in Herat Province together with comparisons between baseline and midline results for Dari. Note that the baseline and midline data collected are not representative of all Dari speakers of Afghanistan. As indicated in the Introduction to this report, ACR was unable to collect midline EGRA data for Pashto in the Provinces of Nangarhar and Laghman due to security reasons. Comparisons between student achievement between the two languages for the midline are thus not possible although one may impute, with certain caution, that any differences found from baseline to midline in the Dari results may apply to the Pashto results since we know the relationship between the Dari and Pashto results on the baseline.

Rationale

To strengthen teacher and administrator skills, teacher preparation systems, and reading materials, reforms and policy initiatives must be undertaken on the basis of quality data on reading outcomes. *Afghan Children Read's* EGRA results will provide essential information to the ministry and donor community that will enable them to refine and focus their endeavors to promote better reading instruction and materials. The EGRA also provides an opportunity for policy makers and educators to measure and monitor learning progress over the next five years as several EGRA administrations are currently planned.

This report, prepared by AIR in collaboration with Creative Associates, presents learner results obtained from the midline data collected in Herat. The purpose of the analyses reported on in this document is to gain insight into student performance in critical reading skills and understand differences between certain student groups within the province derived from the midline EGRA administration. It is also the purpose of the report to present a comparison of student achievement in critical reading skills from baseline to midline in order to determine the degree to which students have progressed from baseline to midline. It is important to emphasize, as explained above, that baseline and midline administrations of EGRA were conducted on students of the same grade (Grade 2) rather than the same cohort of students (which would have meant comparing students from Grade 2 in the baseline with the same cohort of students in Grade 3 in the midline). Rather than follow cohorts of the same students, the design of ACR's study is grade-focused, evaluating progress of the same grade from one year to the next, and progress in teachers' abilities to improve their instructional skills from one year to the next having received reading-based interventions. The report presents EGRA results by sub-task, results by key groups of interest (e.g. gender, school type, etc.), and an analysis of school and home background factors related to reading outcomes.

Limitations

Afghan Children Read intended to collect midline reading outcomes data that allowed plausible inferences that reflect regional realities on the state of literacy in the two languages. However, due to security concerns and limited mobility, midline data collection in the districts of Nangarhar and Laghman was not possible, so the midline data were collected only in Herat.

Sampling for the EGRA has certain limitations, especially in terms of CBEs. The same sampling approach could not be applied for both public schools and CBEs because the number and locations of CBEs frequently change due to factors which include their funding status from the CBE providers.

Comparison of results between schools and CBEs is another limitation. They are not comparable because of the sampling limitations mentioned above, thus rendering any discussion of which schooling type (public school or CBEs) is performing better than the other as unsubstantiated.

There were also limitations in terms of what analytic methods could be used to examine the data. Because of the positive skew in EGRA results (primarily due to the prevalence of a large number of "zero scores"), some types of parametric methods have questionable utility as a primary means to analyze the data. More on the relationship between the reading outcomes and the analytic methods employed to analyze the results can be found in the methods section that follows.

An important caution about the use of EGRA results is also in order. These EGRA results should not be used as a high stakes evaluation of individual schools or students. The purpose of the *Afghan Children Read* data collection was to understand the situation in early grade reading and to employ the results to find ways to improve early grade reading outcomes in Afghanistan.

The Dari baseline and midline results are generalizable only at the provincial level in Herat because the sampling was not intended to give district level statistical significance. *Afghan Children Read*'s original intent was to establish the baseline and midline for the Dari language based on the entire province. However, due to security concerns and inaccessibility to some parts of certain districts as advised by the Provincial Education Directorate (PED), baseline EGRA was administered in 10 accessible districts of Herat (Herat City, Injil, Guzara, Ghoryan, Kohsan, Karukh, Zenda Jan, Kushk Rubar Sangi, Pashtoon Zarghoon and Obe), and for midline EGRA data were collected only in 8 accessible districts (Herat City, Injil, Guzara, Ghoryan, Kohsan, Karukh, Zenda Jan, and Kushk Rubar Sangi),

And as we have already indicated, data collection for the midline in Nangarhar and Laghman had to be abandoned completely for security reasons.

THEORETICAL PERSPECTIVES

Early Grade Reading

Research demonstrates that in languages with alphabetic scripts, most children do not learn to read simply from exposure to written words and their pronunciation (Gough et al., 1996). Instead, children need to be explicitly taught the phonological structure of words (the sounds that letters make in words) and then map the sounds they learn to words in print (Hattie, 2014). Most children need to go through an explicit process where they learn to read words first by sounding them out, grapheme by grapheme, or by grapheme cluster, and then learning to recognize patterns of graphemes within words so that recognition becomes automatic. By teaching children to sound out words, grapheme by grapheme, or by clusters of graphemes – the decoding process required before automaticity is acquired - they can begin to figure out recently learned and new words independently.

This focus on learning to read words and understanding what the words mean underpins a set of hypotheses that form the *Simple View of Reading* (Hoover & Gough, 1990). In this view, reading is comprised of two constituent parts: learning to read words and understanding the meaning of the words read. In other words, reading consists of the two constructs of decoding skill and language comprehension skill. Hoover and Gough (1990) define decoding as "efficient word recognition; the ability to rapidly derive a representation from printed input that allows access to the appropriate entry in the mental lexicon and, thus, the retrieval at the word level" (ibid, p. 130). They define language comprehension as "the ability to take the lexical information (i.e., semantic information at the word level) and derive sentence and discourse interpretations" (ibid, p. 131). To become a proficient reader, both decoding and language comprehension skills are necessary.

The *Simple View of Reading* has been validated in several languages and writing systems (Joshi, Tao, Aaron, & Quiroz, 2012) and in second language (L2) reading as well (Lervåg & Aukrust, 2010; Proctor, Carlo, August, & Snow, 2005; Verhoeven, van Leeuwe, & Vermeer, 2011). The *Simple View of Reading* with its focus on both decoding skills and language comprehension thus provides a foundational framework to view reading proficiency regardless of the language. In a report produced by the National Reading Panel and the National Institute of Child Health and Human Development (2000) the *Simple View of Reading* identifies 5 core components of reading considered as important for the development of decoding and comprehension skills; they are: Phonological awareness; Alphabetic principle, Fluency, Vocabulary, and Comprehension, the first 3 of which relate to decoding skills, the latter 2 to comprehension. It is important

that all reading programs in Afghanistan ensure that children learn about the sounds of the letters in words, how to map the sounds to print, and to learn the meaning of new words (Gove & Cvelich, 2011).

Factors That Influence Early Grade Reading

Many factors influence achievement in reading. The *Afghan Children Read* project focuses on the five fundamental factors that are amenable to control by system administrators, teachers, teacher trainers, and curriculum and assessment specialists as they develop grade appropriate programming to successfully teach reading. These factors, which are slightly adapted to capture ACR's focus, include (Gove & Cvelich, 2011):

- The allocation of sufficient **time** to teach reading
- Access to quality **texts** at the appropriate reading level
- Provision of effective **teaching**
- Appropriate classroom-based **testing**
- Use of the child's mother **tongue**

Early Grade Reading in Multilingual Settings

Research supports the idea that for alphabetic scripts there is somewhat of a "universal" framework for teaching children to read (Geva & Genesee, 2006). Instruction in decoding skills and spelling, for example, can be similar across languages and contexts (Lesaux et al., 2006). There is also evidence of a positive relationship between the native language and a second language (frequently a language of instruction) in learning to decode, read sight words, and spell (Comeau et al., 1999). Nonetheless, the EGRA instruments used in this study were designed to meet the needs of mother tongue speakers of specific languages. The Dari instruments differ in some ways from the Pashto EGRA instruments as linguistic nuance, curricula, and culture were taken into consideration during instrument development. As will be highlighted in later sections, it should also be noted that some students sitting for the EGRA in Dari and Pashto were not taking the assessment in their mother tongue.

Why EGRA?

The EGRA is a battery of early grade reading skills assessments (sub-tasks) that assess learning outcomes in the core skills essential for the development of reading fluency with comprehension. The EGRA sub-tasks include assessments of the five core skills: Phonological awareness, alphabetic principle, fluency,

vocabulary, and comprehension (Vaughn & Linan-Thompson, 2004; Gove & Cvelich, 2011). We provide a brief definition of each of these 5 core skills in Table 1:

Table 1: The 5 Core Skills of Reading

Core skill	Definition
Phonological awareness	• Ability to hear and manipulate the sounds in spoken words • Prepares students to link sounds with print • Precedes decoding (sounding out words) and encoding (spelling) General acquisition progression (from beginning to more advanced): Hear – Identify – Manipulate (from phoneme (sound) to syllable to word)
Alphabetic principle	• Ability to match letters (symbols) to phonemes (sounds) • Basis for an approach (phonics) to teaching reading that stresses letter-sound correspondence and its use in reading and spelling words • Enables students to decode, which is an essential means for reading new words General acquisition progression (from beginning to more advanced): Print awareness – Letter/grapheme awareness – Letter/grapheme-sound correspondence – Decoding – Spelling – Word attack skills
Fluency	• Ability to read words with no noticeable cognitive effort • Must be developed to an 'automatic' level • Automaticity allows student to focus attention on understanding and on content • Research shows that students need to reach a fluency threshold before they can comprehend reliably • Fluency thresholds vary by language General acquisition progression (from beginning to more advanced): Decode accurately – Increase speed – Self-correct – Read with expression
Vocabulary	• Recognition and acquisition of knowledge of words and the meaning of words • Students will only comprehend text if they understand the words they are reading • Familiarity/recognition of words aids in decoding
Comprehension	• Active process of constructing meaning from text • Ability to actively engage with, and derive meaning from, text being read • The goal of reading

The appropriateness of the EGRA for the Afghanistan context stems from three primary rationales:

1. More information is needed about what reading and pre-reading skills need to be enhanced in Afghanistan;
2. EGRA is research-based and has been proven effective in similar contexts (RTI International, 2016). In its initial analyses of the baseline data (see the baseline report produced in 2017), AIR conducted what are known as "factor analyses" to show that EGRA data are indeed based on a theoretically valid foundation. We describe the results of these validity analyses in the section "Construct Validity of EGRA Measures" to be found on pages 20-23 of the baseline report. It was not necessary to repeat factor analyses based on the midline data because the instruments used in the midline were the same as those used in the baseline.
3. EGRA is efficient to administer, analyze, and score.

While the EGRA does not serve as a diagnostic tool for individual students or as a high-stakes examination for teachers, schools, or regions, it does provide an efficient means to check overall reading progress in early grades and determine where improvement is necessary. A thorough analysis of EGRA results can provide evidence to support improvements in ministerial and local instructional policy, teacher training programs, and classroom-level approaches to improving reading outcomes. Wider dissemination of EGRA results can also encourage greater community and parental involvement in early childhood education and literacy initiatives (Dubeck & Gove, 2015).

RESEARCH METHODOLOGY

Research Design

Over the course of the project, *Afghan Children Read* is piloting an early grade reading program in Dari and Pashto languages in the Herat, Nangarhar, and Laghman provinces of Afghanistan. The project is also working in two districts of Kabul to showcase the implementation of the reading program for stakeholders. To establish the reading outcomes baseline for the pilot program, *Afghan Children Read* conducted the EGRA in Herat Province for Dari language (April 2017), and in Nangarhar and Laghman Provinces for Pashto language (September 2017). The baseline was conducted for completers of Grade 2 and Grade 3[1] in public schools and CBE settings. CBEs are community-based classes established and managed by NGOs/CBE providers such as UNICEF, BRAC, CRS, CARE, and the Swedish Committee Afghanistan. CBEs are associated with the nearest public school that acts as hub school for CBE classes.

A midline EGRA was conducted in April 2018 for students who had completed Grade 2 (but started Grade 3 – see footnote 1 on the previous page). The midline was limited to Dari language in the Province of Herat and included students from both public schools and CBEs, as was the case with the baseline. We have explained in earlier sections why Pashto was not part of the midline. The midline was conducted in the same districts as the baseline but in different schools. Our demographic data indicate that approximately 33% of the schools that participated in the midline had also participated in the baseline, while the remaining 67% of the midline schools were new to the EGRA assessment.

Research Questions

Data from Afghan Children Read's 2018 midline EGRA will be used to answer the following research questions:

(1) What is the overall level of reading in public schools at the end of Grade 2 in Dari language in *Afghan Children Read*'s pilot provinces? In what sub-skill areas are the students performing well? Where is improvement needed?

(2) How does the level of achievement in reading in public schools at the end of Grade 2 midline in Dari compare with that of the baseline?

[1] Due to logistical reasons, the EGRA was administered to these students at the start of Grade 3 (for Grade 2 completers) and at the start of Grade 4 (for Grade 3 completers).

(3) What schools and non-school background factors are associated with reading outcomes and how do these associated factors compare across baseline and midline?

EGRA Instrument Adaption

Afghan Children Read engaged the MoE in adaptation of the EGRA instruments. Its team, comprised of personnel from the Departments of Academic Supervision, Teacher Education, General Education and early grade teachers, participated in the adaptation process. The EGRA instruments developed under USAID's national EGRA (2016) were adapted for the *Afghan Children Read* baseline. EGRA instrument adaptation took place in a five-day workshop for each of Dari and Pashto languages and was facilitated by international reading experts. Adapted instruments were field tested and finalized based on results from the field test. The following instruments were adapted in both languages:

- Student EGRA Test (described in next section)
- Student Questionnaire (covering student characteristics, activities by their teachers, their absenteeism, their home support and resources, and their evaluation of school environment),
- Teacher Questionnaire (covering teacher characteristics, their instructional behavior, activities of their supervisors, and their vision of school environment),
- Principal Questionnaire (covering principal characteristics, their pedagogical approach, activities by school inspectors, and their vision of school environment and resources).

Following the finalization, the instruments were digitized in Tangerine software for use on tablets.

Selection of Sub-Tasks

The MoE team, facilitated by international experts, adapted the sub-tasks for the country context using a protocol for their localization. After reviewing primary grade reading standards, textbooks, and other related materials, item writers participated in item and sub-task development workshops in which sub-tasks were developed for all three administrations at the same time (baseline, midline, and endline). The team adapted EGRA sub-tasks to two main national languages and cultural contexts – Dari in the Herat province and Pashto in the Nangarhar and Laghman provinces – for the 2017 baseline. The baseline consisted of a total of seven sub-tasks for both grades (2 and 3) and languages (Dari and Pashto). Parallel test forms for Grade 2 Dari were used in the midline. The sub-tasks measured skills in: Listening Comprehension (LC); Phonemic Awareness (PA); Syllable Sound Identification (SSI); Invented Word Reading (IWR); Familiar Word Reading (FWR); Oral Reading Fluency (ORF); and Reading Comprehension

(RC). All EGRA sub-tasks were administered orally in one-on-one sessions with a test administrator and a single student over a 20 to 25-minute period. During this time, a student questionnaire was also administered. A full description of each sub-task can be found in Appendix 1. Below is a brief description of each of the sub-tasks with a description of what the sub-task demands.

Table 2: Summary of EGRA Baseline Sub-tasks

Sub-task	Pre/Reading Skills	Pupils were asked to:
1. Listening Comprehension (LC)	Oral language comprehension, vocabulary knowledge	Demonstrate listening comprehension of grade-appropriate text by answering 5[2] questions about the text
2. Phonemic Awareness (PA)	Phonemic awareness	Identify the first phoneme from 10 commonly used words by isolating and sounding out just the first sound (phoneme) from a whole word read by the administrator
3. Syllable Sound Identification (SSI)	Decoding, produce sound	Produce the sounds of the written form of a series of syllables (TIMED)
4. Invented Word Reading (IWR)	Letter–sound correspondence, decoding	Read aloud 50 grade-appropriate one to three syllable pseudo words (TIMED)
5. Familiar Word Reading (FWR)	Word recognition and decoding	Read aloud 50 familiar, grade-appropriate unrelated words (TIMED)
6. Oral Reading Fluency (ORF)	Reading with fluency, accuracy, and speed	Demonstrate oral reading of grade-appropriate passage with 66 (TIMED)
7. Reading Comprehension (RC)	Comprehension of text read	Demonstrate comprehension of a passage by answering 5 oral questions, including two inferential questions

Scoring Process

Timed tasks in EGRA are syllable sound identification, invented word reading, familiar word reading, and oral reading fluency. The scores for these tasks were calculated as the number of correct syllables, invented words, or words, read per minute. Three data points are needed to calculate the total score with the following formula applied:

[2] Numbers of items for each of the subtasks have been recommended in the EGRA toolkit (RTI, 2016) and were observed in the construction and localization of the instruments developed for this baseline.

$$Syllables/words\ \ per\ minute = \frac{total\ read - total\ incorrect}{(60 - time\ remaining)/60}$$

EGRA implements an early stop rule where, if the student does not provide a correct response for the first 10 items, the sub-task is discontinued, and the student gets a zero score on the task. Untimed tasks in EGRA are listening comprehension, phonemic awareness, and reading comprehension. The scores for these tasks were calculated as the proportion of correct responses based on the number of questions attempted by the learner.

$$subtask\ score = \frac{total\ correct}{total\ attempted} * 100$$

After administration, each item was reviewed and analyzed to ensure fairness and balance based on gender. Item p-values (a p-value is an estimate of the difficulty level of an item obtained from the number of students who got the item right over the total number of students who answered the item) were evaluated along with item-test correlation coefficients (this is a measure of how well an item discriminates between good and bad students by correlating item score with test score) and differential item functioning (analyses that are used to determine if an item favors a particular sub-group of students) by gender. The full results of these item level analyses can be found in Appendix 3.

Sampling Method

The purpose of this section is to explain the sampling approach and methods used for collecting baseline and midline data. The student population was divided into two major population groups by school type, i.e. public schools and CBEs for each of the three provinces targeted.

The MoE's Education Management Information System (EMIS) data was used for sampling of schools. Since CBEs were not part of EMIS, CBE data was therefore obtained from the PEDs and CBE providers.

The student population was divided into two major population groups by type of schooling - public or CBE. A multi-stage sampling approach was employed for the EGRA Dari midline (the same as for baseline). Cluster sampling was applied to the total number of students in each relevant grade from the public schools. The cluster samples of schools in the province were proportionately allocated to the districts based on the district's proportional representation in the total population. This allowed for appropriate representation

in each of the districts in the sample and reduced skewing towards a district or set of districts. Schools were selected using a probability proportional technique in each district as per proportionate sample size. Ten students per grade per school was the sample size assessed in public schools.

CBEs are viewed as offering alternate, temporary arrangements for education and usually after some time these classes are merged into hub schools. CBEs are unlike schools and typically consist of isolated classes for one grade. In Herat, CBEs with required grades were fewer compared to public schools. Thus, for CBEs in Herat, most CBEs with grade 2 in the accessible districts were assessed. However, for already mentioned security reasons, midline data collection in Nangarhar and Laghman was not carried out.

Probability proportionate sampling (PPS) was used for random selection of the schools using the cumulative school population. Ten students per grade per school was the sample size assessed for the EGRA. A complete explanation of each step in the derivation of the sample can be found in Appendix 2.

The corresponding numbers of students, teachers and principals for the provinces are provided in

Table ***3*** and Table 4 (counts obtained after data cleaning).

Table 3: Student Sample Sizes by Grade per Baseline and Midline

	Baseline		Midline	
	School	**CBE**	**School**	**CBE**
Grade 2 completers - Herat	1,440	390	1414	484
Grade 3 completers - Herat	1,413	312		
Total (Herat)	2,853	702		
Grade 2 completers - Nangarhar & Laghman	1,331	988		
Grade 3 completers - Nangarhar & Laghman	1,316	102		
Total Nangarhar & Laghman	2,647	1,090		
Grand Total	5,500	1,921	1414	484

Table 4: Teacher and Principal Sample Sizes per Baseline and Midline

	Baseline		Midline	
	School	**CBE**	**School**	**CBE**
Principals - Herat	139		142	
Teachers - Herat	285	46	142	34
Principals - Nangarhar & Laghman	131			
Teachers - Nangarhar & Laghman	266	72		

Selection and Training of Assessors

Afghan Children Read, in consultation with the MoE and USAID, selected the assessors from within the MoE for administering the EGRA based on criteria that included: ability to read and speak the necessary language (only Dari was assessed in midline 2018); previous experience administering assessments or with data collection; experience working with primary-age children; availability during the data collection phase; ability to work in target areas; and, experience and proficiency using a computer or hand-held electronic device (tablet or smartphone). *Afghan Children Read* trained 46 assessors in a 5-day workshop to administer EGRA in paper and electronic formats from the April 1-5, 2018. Training included in-class, as well as, field-based test administration exercises to observe the assessors that were trained to administer a quality test. To ensure quality of data, 44 assessors scoring high in the Gold Standard were selected for administration of EGRA for Midline 2018 in Herat.

Data Collection and Cleaning

EGRA data were collected using tablets (100% digitized data collection) by the trained and qualified MoE assessors. A senior official of the MoE Academic Supervision department, with the Afghan Children Read M&E team, led the data collection process. To ensure data quality, the data collection process was spot checked 100% by a team of PEDs, District Education Directorates (DEDs) and Afghan Children Read. The quality was also monitored through Global Positioning System (GPS) enabled digitized instruments and location data of the sample schools was used to ensure that data collection was done properly in the field. Using EGRA guidelines, *Afghan Children Read* performed thorough data cleaning before beginning the analysis. Various filters were used to clean the data keeping intact, inter alia, the validity, accuracy, relevance, completeness and reliability of the data.

Data Analyses and Interpretation

AIR conducted the following analyses on the Dari Grade 2 midline data:

(1) Mean scores per EGRA sub-task by whole group and sub-group, including zero scores;
(2) Percentages of students per performance level per EGRA sub-task by whole group and sub-group;
(3) Context factors derived from the midline student, teacher, and school principal surveys associated with midline student reading performance on the EGRA;
(4) Multiple regressions on midline EGRA performance data associated with survey data.

In each case, comparisons of results were made between the midline and the baseline.

For the 2017 baseline, AIR analyzed Dari and Pashto EGRA data at the item level to determine the reliability of EGRA sub-tasks, sought evidence to ensure the validity of our inferences about reading skills based on those sub-tasks, and examined the empirical evidence that there is a relationship between Oral Reading Fluency and Reading Comprehension. In our 2017 baseline report, we presented coefficients of reliability, the results of factor analyses to provide some construct validity evidence, including correlation coefficients, to demonstrate the convergent validity between Oral Reading Fluency and Reading Comprehension. These results confirmed that there were indeed two major latent traits that determine performance on the EGRA sub-tasks – oral reading fluency and comprehension – and that oral reading fluency represented a strong prerequisite for comprehension.

Although the baseline and midline instruments were parallel forms, we have analyzed midline data to reconfirm the validity of and relationship between the two constructs of oral reading fluency and comprehension. The reader can also refer to the baseline report for the similar evidence presented.

We also calculated the reliability, or internal consistency, of each of the EGRA sub-tasks using the midline data. The term internal consistency refers to the extent to which the items in the instrument measure the same construct consistently. The results of these analyses are presented in the Overall Findings section below, compared with their corresponding baseline results.

All midline statistical analyses were carried out with sampling weights applied[3]. When comparing performance by groups (e.g. gender, school type), we have provided the results of tests of the statistical significance of mean score differences. The purpose of testing for the statistical significance of differences in outcomes is to determine whether the differences were an effect of chance due to sampling error or were the result of some systematic factor present in the population.

Table 5 shows the student sample sizes for the baseline and midline, Table 6 shows the sample sizes for teachers and principals, and

Table **7** contains unweighted and weighted sample totals.

[3] Sampled districts and schools should be proportionally represented in analyses because they are of different sizes and different enrolment rates. For this reason, weights are applied to account for relative differences in sampled units.

Table 5: Baseline and Midline Student Sample Sizes for Dari Language Grade 2

	Grade 3 (Grade 2 completers)					
	Baseline			Midline		
	Male	Female	Total	Male	Female	Total
School	710	730	1440	520	894	1414
CBE	218	172	390	312	172	484
Total	928	902	1830	832	1066	1898

Table 6: Baseline and Midline Teacher and Principal Sample Sizes for Dari Language

	Baseline						Midline					
	School			CBE			School			CBE		
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
Principals	105	34	139	-	-	-	98	44	142	-	-	-
Teachers	108	177	285	22	24	46	38	104	142	19	15	34
Total	213	211	424	22	24	46	136	148	284	19	15	34

Table 7: Comparison of Unweighted and Weighted Sample Sizes for the Midline

	Public Schools				CBEs			
	Male		Female		Male		Female	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
G2	520	708	894	1071	312	75	172	44

GENERAL FINDINGS

Reliability of Sub-Tasks

AIR first assessed the reliability, or internal consistency, of the EGRA sub-tasks. The term internal consistency refers to the extent to which the items in the instrument measure the same construct consistently. As the reliability coefficient increases, the portion of a score that can be attributed to error will decrease; thus, higher values (generally above .80) are desirable. Several of the sub-tasks (e.g., Listening Comprehension, Reading Comprehension, and Phonemic Awareness) have a very low number of items, which can result in low reliability coefficients.

For the midline EGRA administration, the Listening Comprehension sub-task (5 items) showed weak reliability, while Phonemic Awareness (10 items) and Reading Comprehension (5 items) showed acceptable reliability levels (above 0.80; see tables 8-11 below), levels which were similar to those found on the baseline. As mentioned in the baseline reports, for LC the low reliability index may be attributed to one of three possible problems: the LC items not well constructed; the items or the listening text may contain language that is confusing to students; or students guess their responses.

As standard reliability estimation approaches (such as Chronbach's alpha) are not appropriate for the reliability estimation of timed tests, summary scores of the timed sub-tasks were used to calculate the overall Alpha for those sub-tasks (RTI International, 2016). Table 8 contains the reliability coefficients for all 4 combined sub-tasks that are timed, viz., FWR, IWR ORF, and SSI, providing an overall estimate of their reliability. We see that the reliability for timed tests is high at 0.88. We provide a comparison of reliability scores for the Dari midline compared with the baseline:

Table 8: Reliability Coefficients for EGRA Sub-tasks for Dari

Sub-task	N of Items	Baseline	Midline
		Cronbach's Alpha	Cronbach's Alpha
Untimed tasks			
Listening comprehension	5	.585	.539
Phonemic awareness	10	.926	.908
Reading comprehension	5	.883	.801
Timed tasks (4)		.827	.880

Item Quality on the EGRA Administration

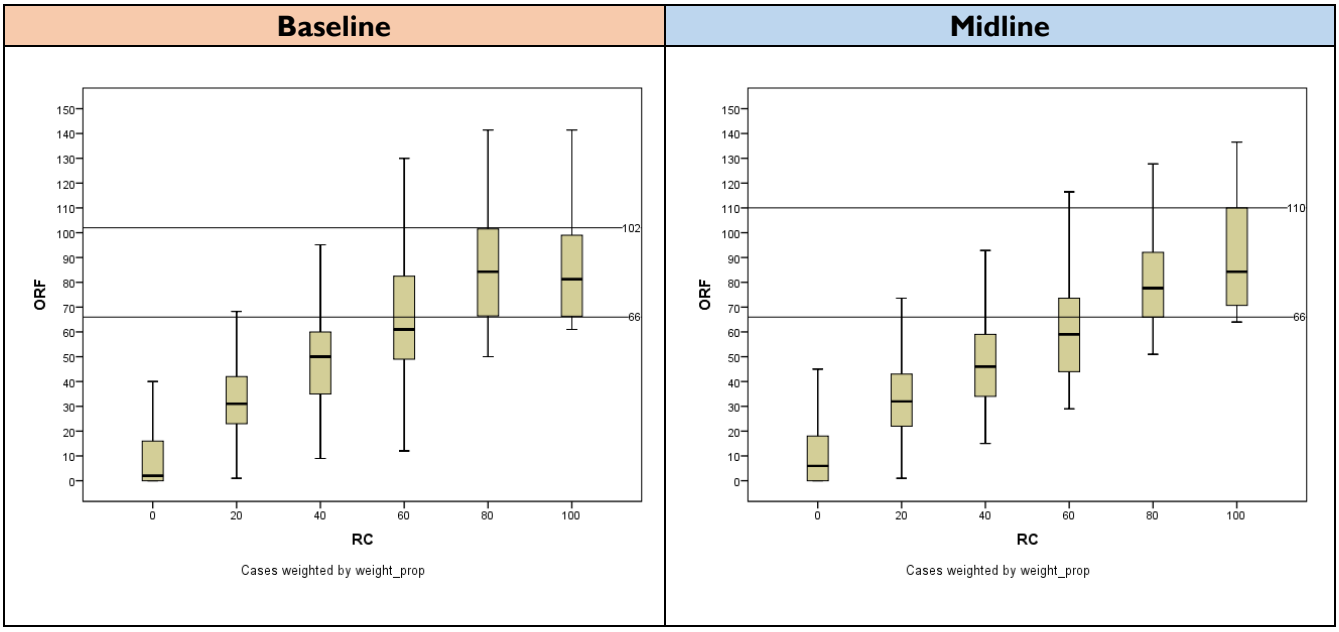
In the baseline, item level evaluation of the EGRA items on each of the sub-tasks suggested a sound design with a homogeneous distribution of item difficulty and discrimination indices. Given that the same items were used for the midline, no item quality analyses were conducted as we expect that items show the same behavior in the midline as they did in the baseline.

Relationship Between Oral Reading Fluency and Reading Comprehension

Before presenting the EGRA results by sub-task, an analysis of the relationship between Oral Reading Fluency and Reading Comprehension is in order. That is because it is possible that an examinee (in some instances) can simply read a list of words (or coherent text as in the ORF sub-task) and not fully comprehend what he or she has read. As attention to EGRA results frequently focuses on the number of words-per-minute read (ORF sub-task results), the relationship between this important sub-task and Reading Comprehension needs to be established. Many studies indicate that ORF is positively correlated with Reading Comprehension (AIR, 2016; Dubeck & Gove, 2015; RTI International, 2015) and, as we have explained on page 19 in the Data Analyses and Interpretation section, our own factor analyses confirm the positive relationship between these two constructs.

On the 2018 *Afghan Children Read* midline EGRA, students responded to up to five reading comprehension questions upon completion of the ORF sub-task to ensure they are not simply reciting words in a list they do not understand (comprehension questions were only asked for the portion of the text read by each examinee). The box plots shown in Figure 1 below, Oral Reading Fluency and Reading Comprehension Grade 2 (Dari) for Midline and Baseline, indicate the range of ORF scores obtained at each level of comprehension for midline compared with baseline. There are 6 levels of comprehension, each corresponding to the number of questions answered correctly on the reading comprehension sub-task (from 0-5). Each box shows the range of ORF scores at one of the levels of comprehension. The major conclusion drawn from these box plots, as we pointed out for the baseline study, is the strong relationship or correlation between ORF and RC – the higher the ORF the greater the level of reading comprehension. This is borne out with the midline data and, as shown in the box plots, the correlation coefficient between the two sub-tasks for Dari was 0.848 (note that for the baseline, the correlation was calculated on the basis of Grade 2 and Grade 3 data combined and was positive but somewhat weaker than for the midline at 0.57).

Figure 1: Oral Reading Fluency and Reading Comprehension



EGRA Benchmarking

Setting performance standards or benchmarks is a process of developing conceptual and operational definitions of knowledge and skills that students should demonstrate to be considered as achieving at a specific level of mastery. Conceptual and operational definitions of performance standards or benchmarks represent a necessary component of standards-based assessments which are used for the evaluation of student performance relative to specified learning criteria. Performance standards are used in national assessment systems as descriptions of national learning targets; they substantially improve the interpretability of assessment results, as well as provide a meaningful framework for monitoring educational progress within institutional, regional, and national contexts.

In Afghanistan, the workshop on Setting Performance Standards for Early Grade Reading was conducted in Kabul on March 6-7, 2018. The panel of participants was composed of about 60 educators, curriculum experts, teachers, and policy makers. The panelists developed conceptual definitions of performance levels, including their labels, and specific descriptors. The general conceptualization of performance levels is applicable across grades and subjects and will serve as a foundation for establishing performance standards for other national learning assessments.

Specific performance level descriptors are developed to describe levels of reading achievement in grades 2 and 3, with emphasis on performance in reading fluency and reading comprehension. These specifications of reading achievement levels convey the following specification of reading fluency and comprehension. Table 9 below shows the performance levels as defined in the benchmarking workshop held in Kabul in early March 2018.

Table 9: Descriptions of Performance Levels

Level label	ORF score range	Conceptual definition of level
Good reader	65+	Students at this level read short stories accurately and use a phrasal pace when reading connected text. They read with phrasing, expression and inflection. They comprehend almost all the text and retell the main idea of a text and draw conclusions. They use prior knowledge to clarify understanding. Full or almost full comprehension was defined by reading comprehension scores of 80% or above. This level represents a reading benchmark that will be used as a national target.
Average reader	40-64	Students at this level read short stories accurately and mostly use a word level phrasing. They use appropriate tone and stress at the word level. They comprehend most of the text and retell the text. They have difficulty expressing the main idea of a text and drawing conclusions. They sometimes use prior knowledge to clarify understanding. Partial comprehension was considered at 40% or above, but below 80% points on the reading comprehension scale.
Weak reader	1-39	Students at this level make mistakes when reading words. They read sentences at the word level, but have difficulty reading connected text. They do not always use appropriate tone and stress at the word level. They comprehend only explicitly stated information. They have great difficulty expressing the main idea of a text and drawing conclusions. Low comprehension was defined by less than 40% points on the reading comprehension scale. This level includes zero scores that will be reported as a separate category that requires special attention.
Non-reader	0	No reading ability.

The panel of participating experts proposed reading benchmarks expressed in Oral Reading Fluency scores (correct words per minute) required to achieve the levels of reading comprehension described above. Their judgements were data-driven based on the association between reading comprehension and reading oral fluency derived from the EGRA 2017 baseline data. These proposed benchmarks are due to official review by MOE, which may include additional moderation. They are presented in Table 10 below:

Table 10: Early Grade Reading Benchmarks in Oral Reading Fluency Scores

	Dari			Pashto		
Grade	**Weak**	**Average**	**Good**	**Weak**	**Average**	**Good**
1*	0-19	20-40	41 & above	0-14	15-37	38 & above
2	0-39	40-64	65 & above	0-29	30-59	60 & above
3	0-52	53-72	73 & above	0-39	40-74	75 & above

* The benchmarks for Grade 1 are proposed based on extrapolation.

Based on these proposed benchmarks, the results of EGRA can be presented as percentages of students that reach each performance level. This kind of information is helpful for the MOE to support development of capacity of teachers to monitor and evaluate student learning progress, as well as for policy makers to better support the implementation of the reading curriculum.

BASELINE VS. MIDLINE EGRA RESULTS IN PUBLIC SCHOOLS FOR THE PROVINCE OF HERAT

Overall Baseline vs. Midline EGRA Results for Grade 2 Dari

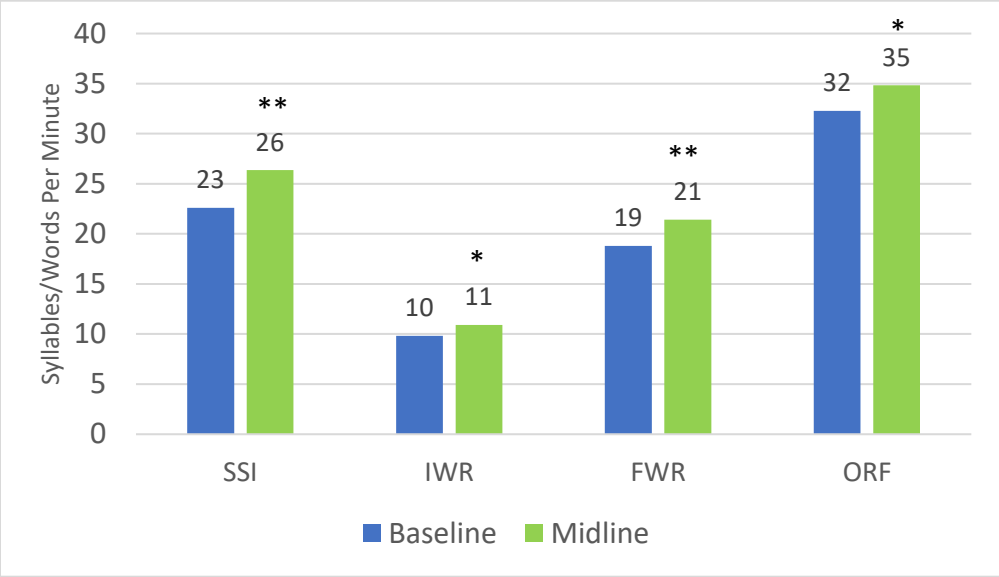
EGRA results by sub-task are presented in this section comparing the results of the baseline and midline test administrations in public schools for the Province of Herat in Grade 2 Dari language. We present the mean scores for the sub-tasks in graphs to simplify understanding of the main outcomes and trends. We also present the results as percentages of students reaching each of the benchmarks. Sub-tasks are labeled with their sub-task abbreviations where appropriate. The timed and untimed sub-tasks are presented separately due to the nature of the scale in which the results are presented: words/syllables per minute for timed tasks and percent-correct scores for untimed tasks.

In Figure 2 below it can be observed that the mean scores in all four timed tasks increased from the baseline to the midline. The gains on SSI, IWR, FWR, and ORF were all statistically significant. On the ORF sub-task, midline students are reading an average of 35 words-per-minute while baseline students read at an average of 32 words-per-minute, a difference that should be acknowledged as statistically significant, but the practical size of this gain is still very small.

It should be also noted that ORF values between 32 and 35 words-per-minute are associated with reading comprehension values of about 20%, as can be seen in the box plots shown in Figure 1 (page 25). Although these ORF scores fall within a "weak" level on the reading performance standards scale, which strongly calls for further improvement for students to reach an acceptable reading comprehension of at least 80%, it is encouraging to observe the positive trend in student reading performance.

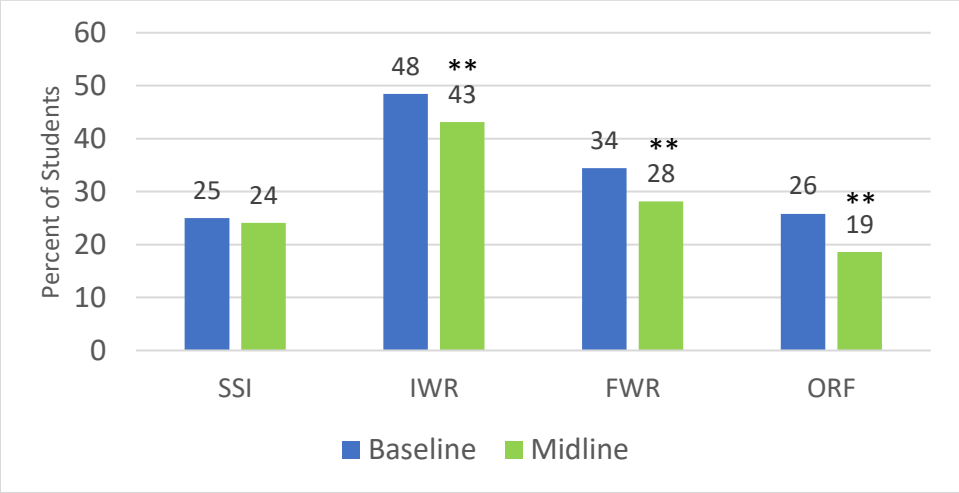
Another way to look at the results and the trends across baseline and midline administrations is to examine the "zero score" rates. Figure 3 below presents the percentage of students who could not answer a single item correctly for timed tasks. The Invented Word Reading sub-task had the highest rate of zero scores at 43% in the midline, while the sub-task with the lowest rate was Oral Reading Fluency at 19%. Similar results could be observed in the baseline with statistically significant improvements (i.e. reductions in zero scores) from baseline to midline on 3 of the 4 sub-tasks. Again, the trend of decreasing percentage of non-readers is an encouraging finding.

Figure 2: Average Syllables/Words Per Minute for Timed Tasks



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Figure 3: Percentage of Zero Scores for Timed Tasks

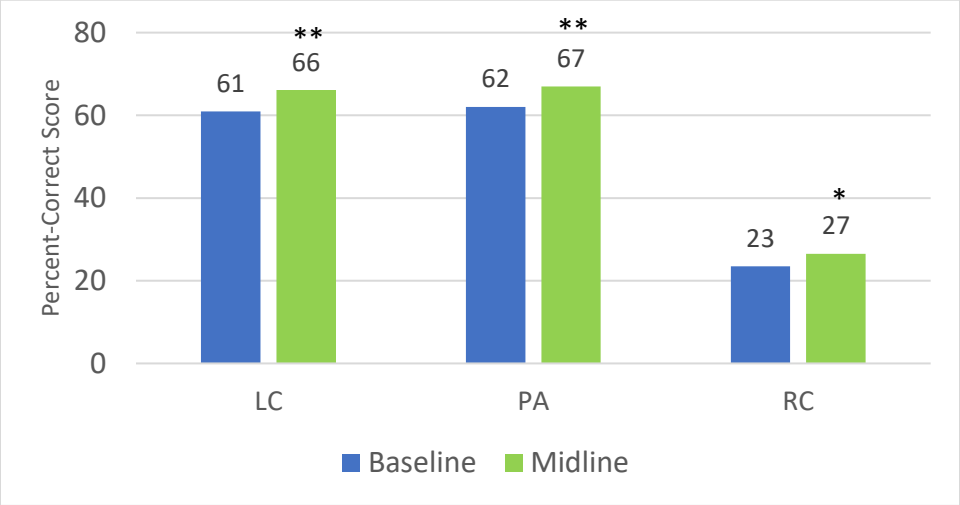


Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Figure 4 below shows an increase in performance on untimed task scores when compared to the baseline. The increases on the Listening Comprehension and Phonemic Awareness sub-tasks were about 5 percentage-correct points, while in Reading Comprehension the increase was 4 percentage-correct points. The differences between the baseline and midline were statistically significant for all three sub-

tasks. Although the sizes of these differences are relatively small, they contribute to the plausibility of the notion that the ACR intervention yielded a positive effect in relatively short time.

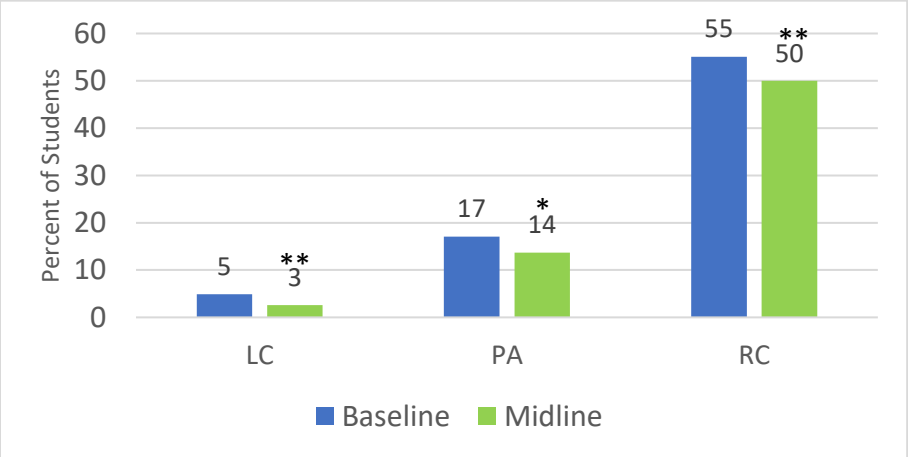
Figure 4: Average Percent-Correct Scores for Untimed Tasks



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

As was the case with timed tasks, the percentage of zero scores on the untimed tasks also decreased from the baseline to the midline as shown in Figure 5 below. The sub-tasks with the largest decrease was Reading Comprehension with 5% fewer students with zero scores, however the zero score rates are still high for the midline at 50%.

Figure 5: Percentage of Zero Scores for Untimed Tasks



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Figure 6 below shows midline results compared with baseline results in terms of percentage of students reaching each performance level. The target levels that teachers should expect their students to reach – the desirable levels – are the "average" and "good" levels shown in two shades of green. These levels are separated with a dark horizontal line from the 2 levels that can be considered as levels where students are not achieving desired standards – "non-readers" and "weak" readers shown in red and orange, respectively. The percentage of students showing proficiency increased from 34% in the baseline to 36% in the midline, a 2% increase, although not yet statistically significant. The percentage of non-readers significantly decreased by 7% from the baseline to the midline.

Figure 6: Percentage of Students in Performance Levels

Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Highlights and implications of overall 2018 EGRA scores for Dari

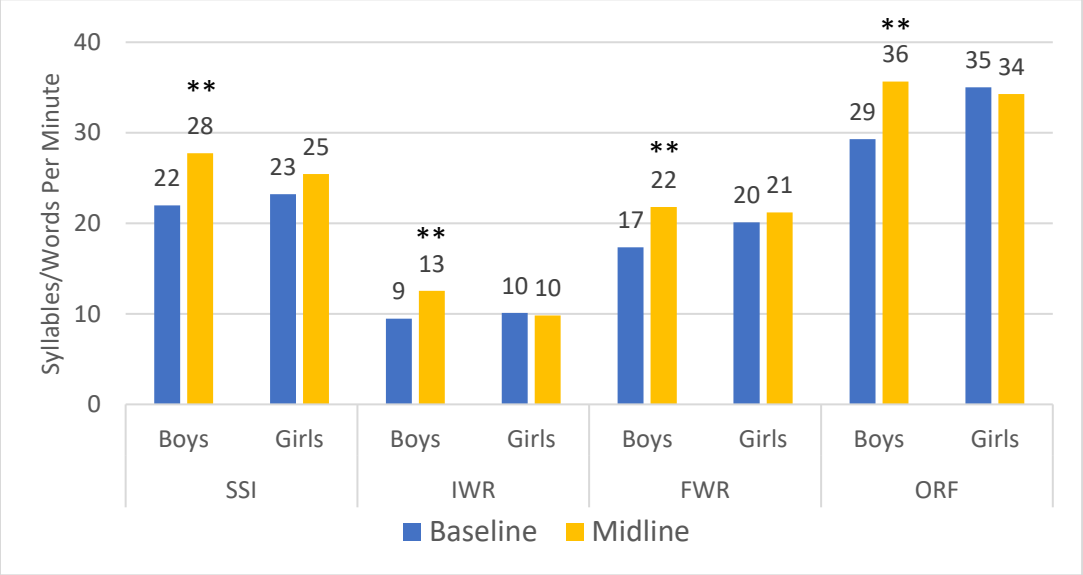
- Scores on all 4 timed tasks (SSI, IWR, FWR, and ORF) increased from baseline to midline, yielding statistical significance on all 4 tasks.
- Corresponding zero scores significantly decreased on 3 of the 4 timed tasks (except SSI).

- Scores on all 3 untimed tasks (LC, PA, RC) increased from baseline to midline, with statistically significant differences on all 3 tasks, with corresponding significant decreases on zero scores.
- 50% of Grade 2 Dari students still scored zero on RC (as compared to 55% on baseline) while on ORF 19% of students scored zero (as compared to 26% on the baseline).
- Average ORF is at 35 words-per-minute on the midline (as compared to 32 on the baseline) while RC is at 27 percent-correct points (as compared to 23 on the baseline).
- Overall, gains from baseline to midline are very modest – there is an increase of 2% of students achieving desired proficiency levels ("average" and "good") – but 64% of students are still insufficient readers, either at "non-reader (19%) or "weak" (45%) levels.

Baseline vs. Midline EGRA Results by Gender

As shown in Figure 7, the scores for Dari Grade 2 students on timed tasks have increased from baseline to midline at different rates by gender. For boys, the differences between the baseline and the midline were statistically significant for all four timed tasks, while for girls the differences between baseline and midline were not statistically significant.

Figure 7: Average Syllables/Words Per Minute for Timed Tasks by Gender

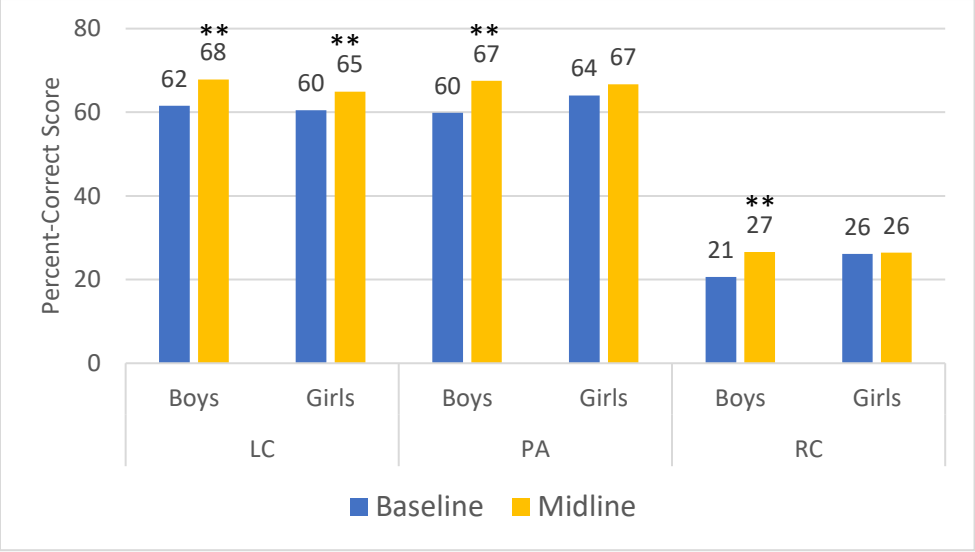


Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

There was an increase in the performance on untimed tasks for both boys and girls from baseline to midline (see Figure 8 below). For boys, the differences were statistically significant for all three untimed sub-tasks – LC, PA, and RC – while for girls only the LC task shows a difference that is statistically significant. These findings raise an interesting and concerning question – why the improvement from baseline to midline was much more evident for boys than for girls.

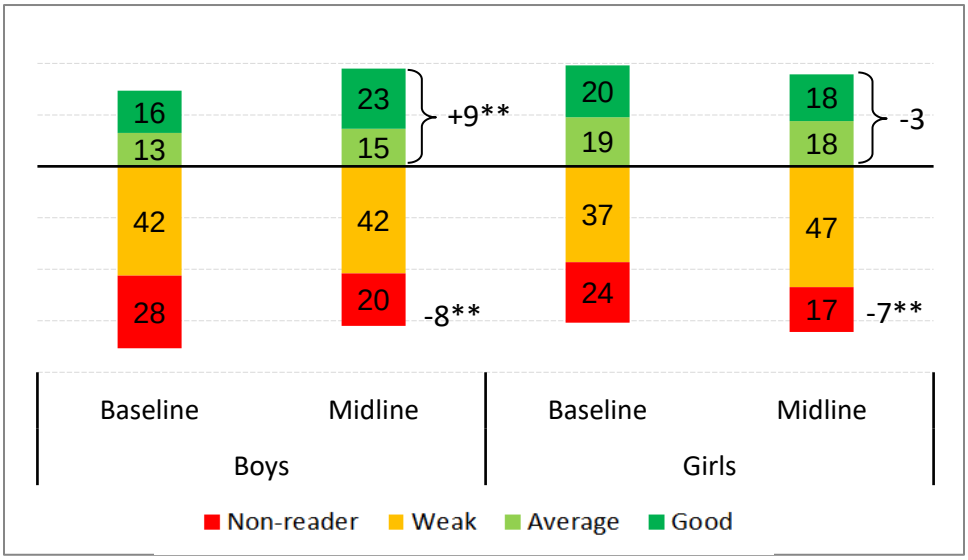
In Figure 9 below, there appear to be clear signs that boys have improved from baseline to midline while girls have not demonstrated such a change. At the 2 targeted levels of "good" and "average", the percentage of boys has increased from 29% to 38% (a 9% increase), which is statistically significant, while for girls the percentage has dropped from 39% to 36% (a 3% decrease); however this decrease is not statistically significant. The percentage of non-readers for boys has fallen by 8% and for girls by 7%, which still indicates that girls have improved considering the evidence of decreased percentage of non-readers.

Figure 8: Average Percent-Correct Scores for Untimed Tasks by Gender



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Figure 9: Percentage of Students in Performance Levels by Gender



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Highlights and implications of 2018 EGRA scores by Gender for Dari

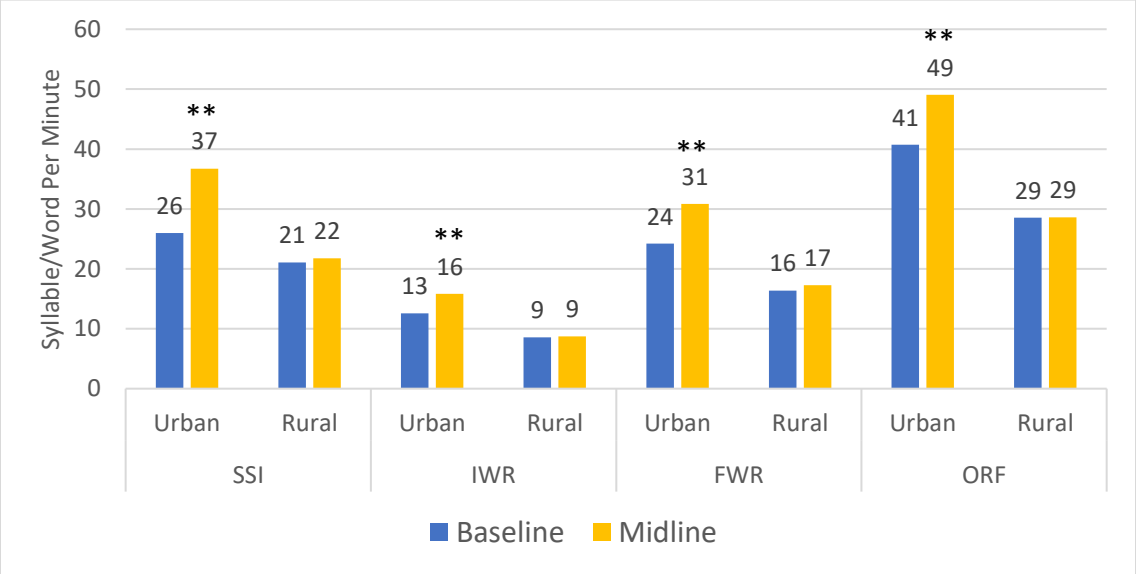
- Boys have made significant gains from baseline to midline on all 7 EGRA sub-tasks while girls have made significant gains on only 1 sub-task (LC).
- There is no significant difference in the scores on the midline between boys and girls, however, on the baseline boys performed significantly lower than girls.
- On ORF, boys' midline score is 36 words-per-minute, very close to girls' score (34 wpm), however, girls outperformed boys on baseline (35 and 29 wpm, respectively).
- On RC, boys' midline average is 27 pct-correct points, and girls' midline average is virtually the same (26 points), whereas on baseline girls scored significantly higher than boys (26 vs. 21 pct-correct points, respectively).
- In terms of the performance standards, there is a significant increase in the percentage of boys achieving at the targeted levels of "average" and "good" (29% on the baseline increased to 38% on the midline); on the other hand, the percentage of girls scoring at the targeted levels did not change significantly (from 39% to 36%).
- Both boys and girls significantly decreased a percentage of non-readers from baseline to midline (8% for boys and 7% for girls).

- This evidence suggests that the progress for boys was uniform across the performance scale, whereas for girls the progress was evident only at the bottom part of the scale.

Baseline vs. Midline EGRA Results by School Location

Figure 10 below shows increased performance on the midline for urban schools on all timed sub-tasks, all measures with statistically significant differences. There was an increase of 8 words-per-minute for ORF. Results for rural schools on all timed sub-tasks remained essentially the same from baseline to midline. Urban schools outperformed rural schools on the midline on all four timed sub-tasks. On ORF, urban students outperformed rural students by 20 words-per-minute (49 versus 29).

Figure 10: Average Syllables/Words Per Minute for Timed Tasks by School Location



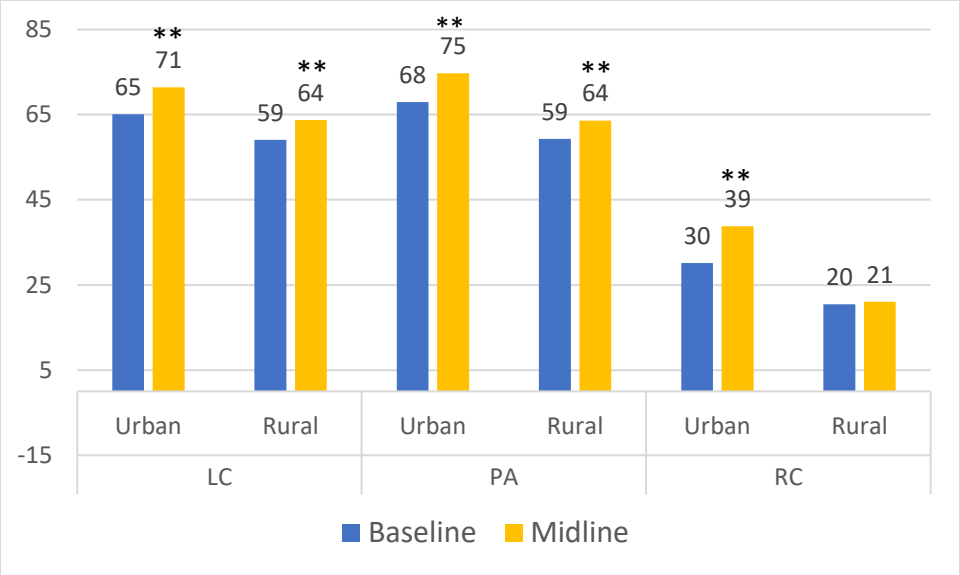
Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

The same trend is observed for untimed tasks, as shown in Figure 11 below. Urban school results on all 3 untimed sub-tasks were significantly higher on the midline than the baseline. The largest increase was on RC, a 9% increase from the baseline (at 30%) to the midline (at 39%). Rural students show increased performance on the midline compared to the baseline on two untimed subtasks, LC and PA, both of which do not involve reading.

On the midline, urban students outperformed rural students on all three untimed tasks and these differences were found to be statistically significant, with the largest difference in RC, 18 percent-correct points favoring urban students.

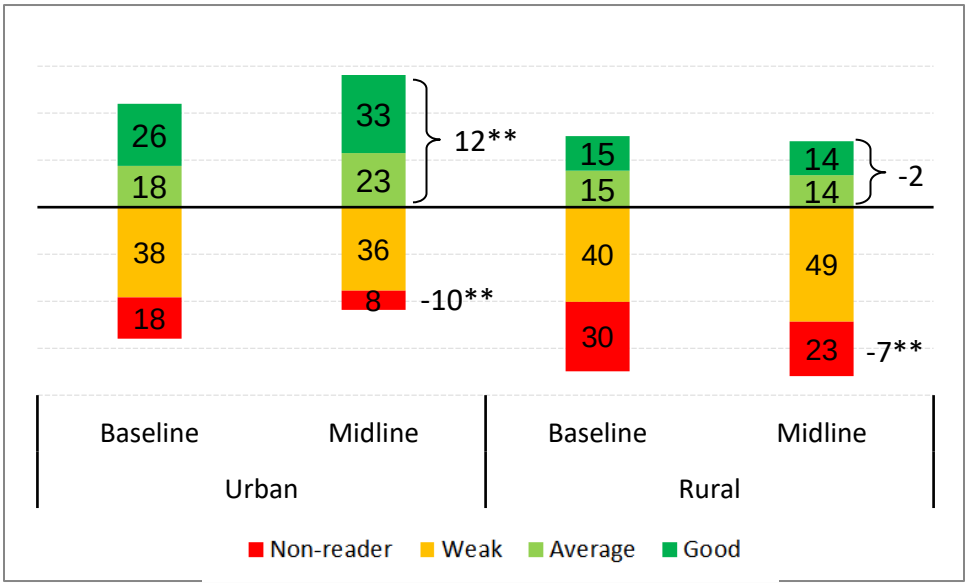
In Figure 12 further below we see that the percentage of urban students achieving at the targeted levels of "average" and "good" increased from 44% to 56%, a statistically significant increase of 12% from baseline to midline, while the percentage of rural students in the top 2 levels did not change significantly from baseline to midline (from 30% to 28%, respectively). The percentage of "Non-readers" decreased significantly in both urban and rural areas: 10% for urban schools and 7% for rural schools.

Figure 11: Average Percent-Correct Scores for Untimed Tasks by School Location



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Figure 12: Percentage of Students in Performance Levels by School Location



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Highlights and implications of 2018 EGRA scores by School Location

- Urban schools improved their performance from baseline to midline on all 7 EGRA sub-tasks, with statistically significant differences on all measures.
- Rural schools did not improve scores on timed tasks, but they did improve on 2 of the 3 untimed tasks (LC and PA).
- Urban schools outperformed rural schools on all 7 measures on the midline at statistically significant levels.
- On ORF, urban schools achieved 49 words-per-minute, an increase of 8 words-per-minute from baseline to midline; rural schools remained the same at 29 words-per-minute from baseline to midline.
- On RC, urban schools scored at 39 percent-correct points on the midline (an increase for 9 points from the baseline), while rural schools did not show a significant change from baseline to midline.

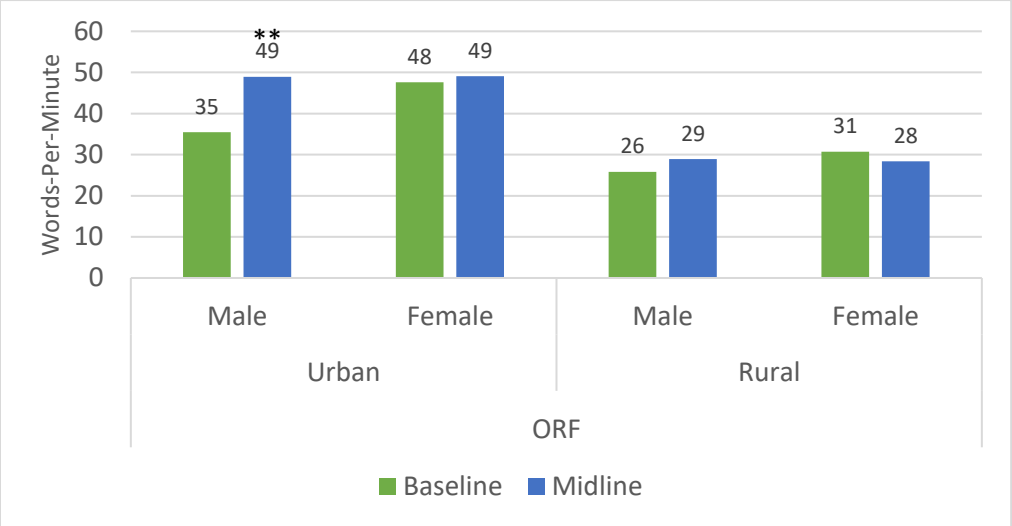
- In terms of performance standards, 56% of urban students achieved scores in the top 2 targeted levels (an increase of 12% from the baseline), while for rural schools the percentage on the midline stood at 28% (a non-significant difference from baseline).
- Considering the bottom part of reading scale, both urban and rural schools showed improvement by significantly decreasing the percentage of non-readers (for urban schools this decrease was from 18% to 8%, and for rural schools from 30% to 23%).

Baseline vs. Midline EGRA Results by Gender and Location

To better understand the trend from baseline to midline EGRA for grade 2 Dari students in Herat, the results were hierarchically disaggregated first by school location, and then by gender within each school location. We present here the results of such a disaggregation for the two targeted EGRA subtasks: ORF and RC.

Figure 10 below shows the average baseline and midline scores for ORF (words per minute) by gender within school locations for grade 2 Dari. A large statistically significant increase in performance on ORF from the midline to baseline can be observed for boys in urban schools (gain of 14 wpm), whereas there was no significant change in ORF for girls in urban schools, neither for girls and boys in rural schools.

Figure 13: Average WPM for ORF by Gender and School Location



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

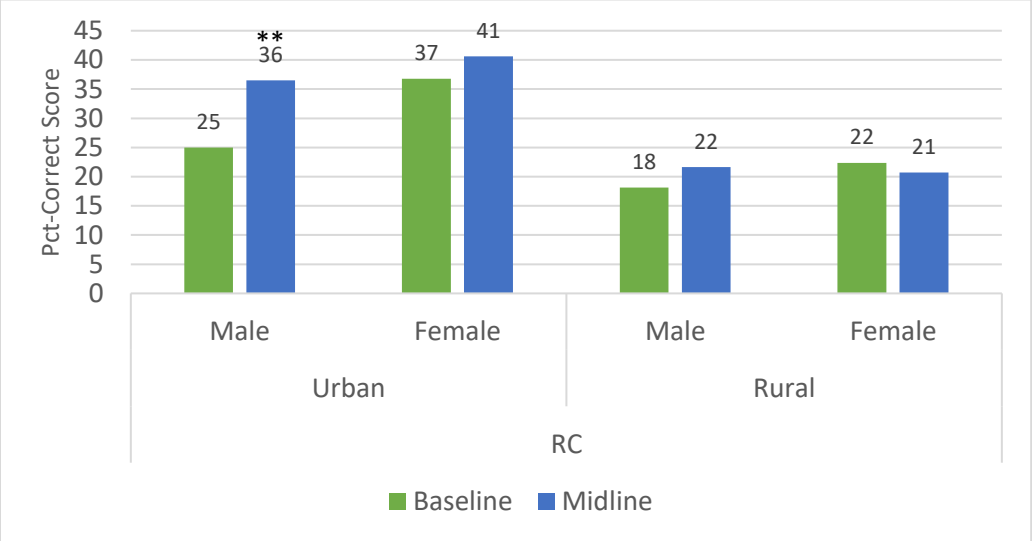
The same trend of results from baseline to midline is observed for RC, as shown in Figure 14 below. The results for boys in urban schools were significantly higher on the midline than the baseline, whereas the differences in other compared categories were not statistically significant. The increase on RC for urban boys was 11 percent-correct points (from the baseline at 25 to the midline at 36 percent-correct points). Urban girls, as well as both rural boys and girls, did not show a significant change from baseline to midline.

In Figure 12 further below we see that the percentage of urban boys achieving at the targeted levels of "average" and "good" increased dramatically from 35% to 58%, a large statistically significant increase of 23% from baseline to midline. The gains in percentage of students in the top two reading levels were not significant for urban girls and rural boys; however, the percentage of rural girls in the top two levels dropped significantly by 5%.

On the other hand, considering the percentage of non-readers, significant improvements can be observed in all combinations of gender and school locations. In urban schools the percentage of non-readers decreased by 5% for boys and 12% for girls, whereas in rural schools the percentage of non-readers went down 12% for boys and 3% for girls.

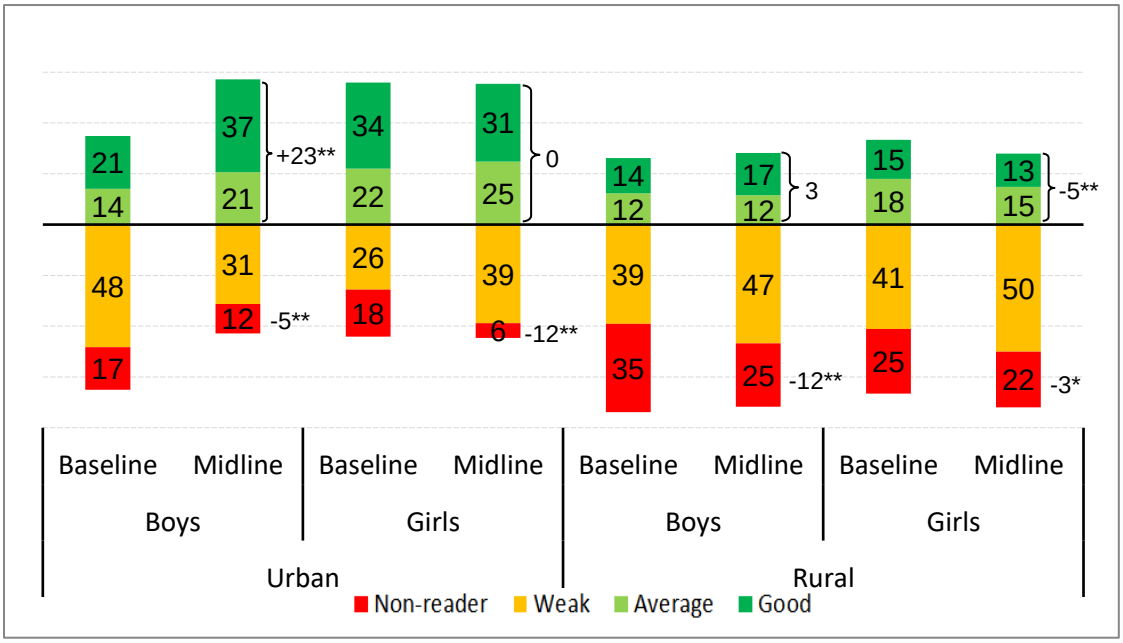
It should be noted that for both genders in rural schools, and for girls in urban schools, the percentage of "weak" readers increased because a significant portion of non-readers moved up to that category, but the move from "weak" up to "average" or "good" was not significant for those categories of students.

Figure 14: Average Scores for RC by Gender Within School Location



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Figure 15: Percentage of Students in Performance Levels by Gender & School Location



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Highlights and implications of 2018 EGRA scores by Gender and School Location

- Considering the average ORF and RC scores, large significant gains from midline to baseline in grade 2 Dari were observed for boys in urban schools (14 wpm for ORF and 11 pct-correct points for RC), whereas differences in other groups (urban girls, rural boys, and rural girls) were not significant.
- Similarly, based on the percentage of students in top two performance levels, a large significant increase (35% to 58%) was observed only for boys in urban schools.
- On the other hand, considering a decrease in percentage of non-readers as a measure of improvement, significant cut-downs of non-readers were observed in all combinations of gender and school locations, ranging from 3% to 12%.

Baseline vs. Midline EGRA Results by School Type

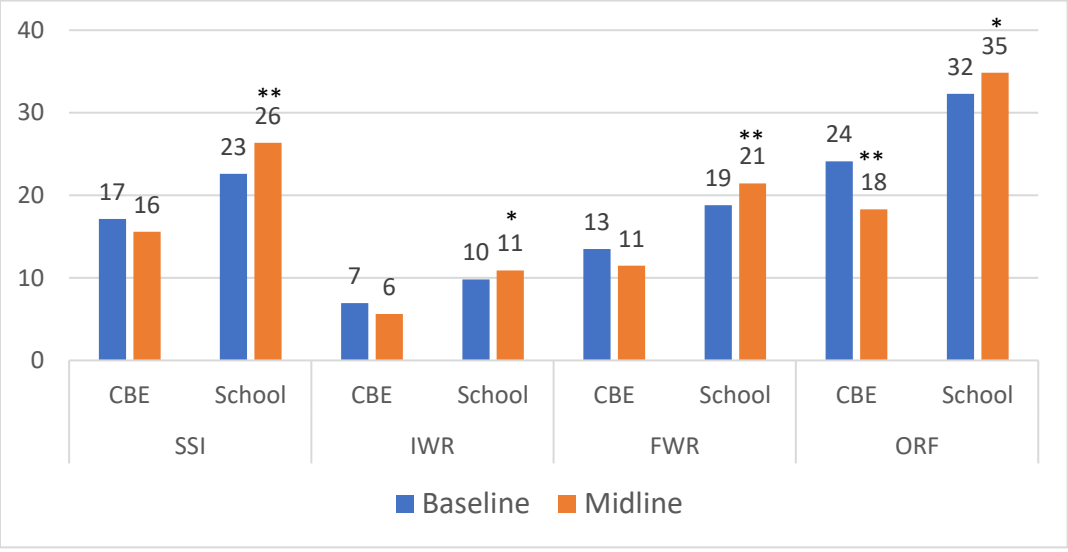
Although this report focuses on public schools, in this section we present the baseline and midline results for CBEs comparatively with the results for public schools. In such a way it is possible to draw conclusions about differential gains between baseline and midline for these two types of schools. Different settings and conditions within which these two types of schools operate was the main reason for presenting the results for public schools separately from the results for CBEs.

Figure 16 below presents the results for the timed EGRA sub-tasks by school type: public school and CBEs. Public school students performed better on the midline than on the baseline in all four timed sub-tasks, with the largest increases on SSI and ORF, about 3 syllables and 3 words-per-minute respectively.

When comparing baseline with midline results, we see a decreasing trend in the performance of CBE students on all four timed sub-tasks, but only a difference in ORF reached statistical significance where the measure fell from 24 to 18 words-per-minute.

Within the midline, it can be noted that public schools outperform CBEs, especially on ORF where public schools have an advantage of 17 words-per-minute over CBEs. However, the increases achieved by public schools still do not reach levels of decoding (fluency) and comprehension that would be considered desirable.

Figure 16: Average WPM for Timed Tasks by School Type

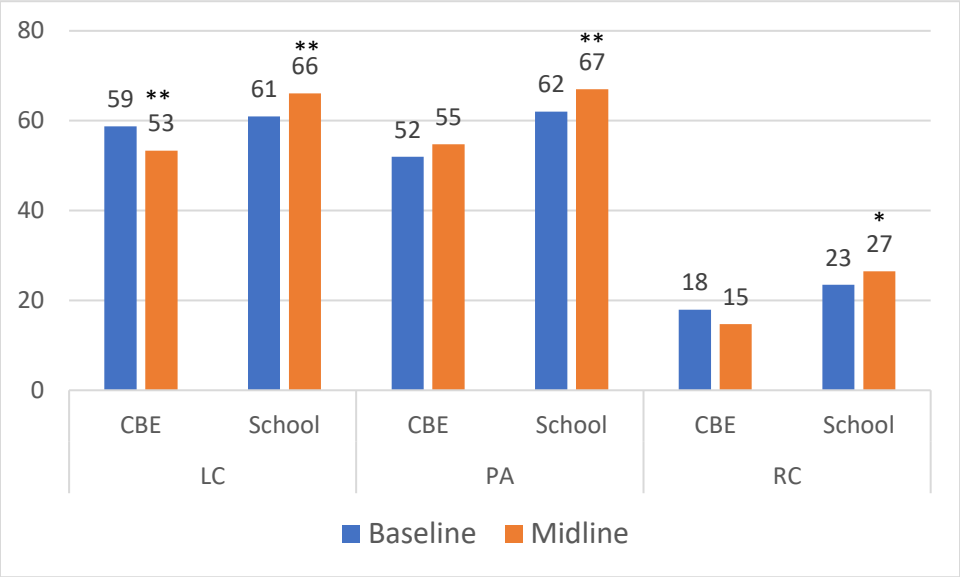


Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

As with timed tasks, public school results on untimed tasks increased in respect to the baseline, with the differences in all three untimed tasks (LC, PA, and RC) being statistically significant. For CBEs, it can be noted that mean scores on LC are significantly lower than on the baseline, while differences on PA and RC were not statistically significant between the two administrations.

Within the midline, public school students outperformed CBE students with large gains in LC, PA, and RC: 13%, 12%, and 12% point respectively. All differences were statistically significant.

Figure 17: Average Percent-Correct Scores for Untimed Tasks by School Type

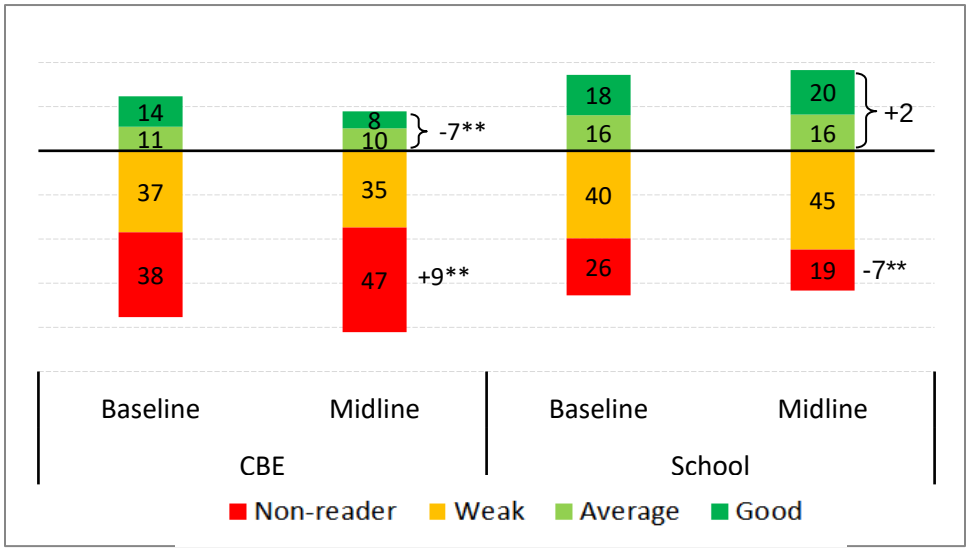


Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

In **Error! Reference source not found.** further below we see that public school students slightly improve from baseline to midline at the top 2 desired levels of "average" and "good" (from 34% to 36%), while the percentage of CBE students reaching the top 2 levels significantly dropped from 25% to 18% a 7% decrease.

It should also be noted that the percentage of CBE "Non-readers" significantly increased from 38% in the baseline to 47% in the midline (a 9% increase), while for public schools this percentage significantly fell from 26% to 19% (a 7% decrease).

Figure 18: Percentage of Students in Performance Levels by School Type



Note: * Difference from Baseline to Midline statistically significant at p < 0.05
** Difference from Baseline to Midline statistically significant at p < 0.01

Based on this evidence it can be concluded that public schools functioned more effectively than CBEs, showing gains from baseline to midline administration that suggest, with certain limitations, possible beneficial effects of the ACR project interventions. These findings represent a challenge for interpretation and require additional analysis of the context within which CBEs function. Possible directions for additional analysis of the CBE context include:

1) Since the population of CBEs is very volatile, there is possibility that the observed differences in outcomes for the CBEs are due to changes in the populations of the schools. Thus, it would be good to investigate how the enrolment in CBEs changes over the years to evaluate whether more needy kids were enrolled in 2018 vs. 2017.
2) Another possible factor that may have spurred differences in outcomes could be a change in management of CBEs (simply due to shifts in administration). Thus, it is recommended to look up the dates and procedures associated with that type of change.
3) Based on multiple research evidence (Hatsaandh, 2019), when the home language of students is the same as language of instruction "they develop certain types of perceptive and metalinguistic awareness sooner and better than those who are enrolled in schools where language of instruction is different from their mother tongue". Thus, using both baseline and midline data, we

conducted additional checks of how the language spoken at home relates to student performance and found the following:

a. Based on both baseline and midline data, the percentage of students whose home language is different than the language of instruction is significantly higher for CBEs than for public schools. In baseline data, there were 19% of CBE students with mother tongue different than Dari compared to 9% of students in public schools. In midline administration the ratio is virtually the same confirming there is significantly more home-language disadvantaged students in CBEs than in public schools: there were 21% of CBE students, compared to 9% of public-school students that don't speak language of instruction (Dari) at home.

b. In the context of the literature mentioned above, this finding raises the hypothesis that the language spoken at home could be one of the factors that contributes to significantly lower reading performance in CBE compared to public schools. Thus, in next step we compared the ORF performance between the students speaking Dari at home and those students whose mother tongue is other than Dari. It could be noted in both baseline and midline administrations that the CBE students whose language at home is other than language of instruction (Dari) tend to have lower ORF mean scores than the students whose language at home is Dari (18 vs. 25 in baseline and 13 vs. 20 for midline administration). However, these differences did not reach statistical significance (p equaled 0.46 in baseline and 0.23 in midline), most likely due to low counts of students with non-Dari mother tongue and overall low counts of students in the CBE sample. On the other hand, in public schools there was no such a tendency that the students whose mother tongue is different from language of instruction would perform lower on ORF than the students who speak at home the same as language of instruction. In baseline data the difference was virtually zero, and in midline data observed difference was 3 points in favor of students with non-Dari mother tongue, but it was statistically insignificant.

To summarize about the association between home and school language – our results are not conclusive as our data were not designed to target this specific problem, which is reflected in low statistical power in regard this issue (small sample of CBE schools and generally low proportion of students speaking at home language other than language of instruction). However, there are some indications that in CBEs the factor of home language not being equivalent with language of

instruction might adversely affect reading performance in early grades, the clue that deserves further attention.

Highlights and implications of 2018 EGRA scores by School Type

- Public schools have made clear gains on all 7 EGRA sub-tasks from baseline to midline, with ORF at 35 words-per-minute (from 32 on the baseline) and RC at 27% (from 23% on the baseline).
- When results for public schools are viewed by reference to performance standards, 36% of students achieve "average" and "good" levels (up from 34%), meaning that 64% of students are still not achieving targeted levels. The rate of non-readers significantly decreased from 26% to 19%.
- The evidence for CBEs is in the opposite direction from public schools – scores decreased on all EGRA sub-tasks, except Phonemic Awareness. ORF dropped from 24 words-per-minute to 18; and RC fell from 18% to 15%.
- Only 18% of CBE students achieved at the targeted levels on the midline (versus 25% on the baseline), meaning that 82% of CBE students are not reaching targeted levels. 47% of CBE students are non-readers (versus 38% on the baseline).

BACKGROUND FACTORS ASSOCIATED WITH STUDENT READING PERFORMANCE

This section presents the findings on the associations between contextual variables from the ***principal***, ***teacher***, and ***student*** questionnaires on one side, and performance on the Oral Reading Fluency (ORF) sub-task of EGRA for Grade 2 Dari, on the other side. The identification of factors significantly associated with reading performance can help the Ministry of Education and the Afghan Children Read program target interventions to those factors that can improve the reading proficiency of Afghan children.

The analyses of associations between background factors and student reading performance was carried out using two methodological approaches: 1) multivariate approach utilizing stepwise multiple regression analysis, and 2) univariate approach based on differences on student ORF scores associated with responses in student, teacher, and principal questionnaires. The advantage of the multivariate approach is the identification of background variables that are uniquely associated with reading performance avoiding redundancies caused by collinearity between background variables. The advantage of the univariate approach is its simplicity and transparency, yielding results that are easier to understand for policy makers and that are more easily transferable to policy actions.

In this section we present the results of the multiple regression analyses for the background factors associated with student reading performance, whereas the results of the univariate analyses are presented in Appendices 3 and 4.

The results of the multiple regression analyses in this section are presented comparatively for the baseline and midline data. The comparison between baseline and midline multiple regressions is carried out through a two-step procedure:

1. The background variables that were identified as significant in separate multiple regression runs for baseline and midline data, using a 'STEPWISE' strategy, were listed side by side. They were matched to identify those that were significant in both analyses and those that were uniquely significant either in baseline or midline analysis.
2. All the predictor variables that were significant in either baseline or midline analysis were included together in the additional analyses using the multiple regression modeling strategy 'ENTER'. In this model, all the predictors are included into the analysis simultaneously and there is no iterative procedure for inclusion or exclusion of predictors to or from the multiple regression equation. Rather, all the predictors are concurrently evaluated for the significance of their contributions to the explanation of the criterion variance.

The results of midline and baseline multiple regression analyses for student, teacher, and principal background factors are presented comparatively in Table 11 through Table 13. The labels for predictor variables are listed in the first column. Note that in the midline administration the question numbers were slightly different to those used in the baseline, but the formulations of the questions were virtually identical. We also took care that the preparation of data for regression analyses (e.g., recoding) was done in a consistent way.

The columns on the right side of Table 11 through Table 13 contain standardized regression coefficients and corresponding significance levels for baseline and midline results, respectively. The significant coefficients are highlighted in yellow for easier identification of contextual variables that were significant in both baseline and midline analyses, or uniquely significant in just one of them.

Student Background Factors – Results Baseline and Midline Multiple Regressions

The comparative results of baseline-midline multiple regression modeling for student contextual variables are presented in Table 11. There were 26 student questionnaire variables that showed significant associations (at $p<=0.05$) with student reading performance in the baseline analysis and 18 variables that were significant in the midline analysis, and among them, there were 9 variables that appeared significant in both analyses. These factors that exhibit consistent associations with student reading performance should be considered as priority for planning actions by policy makers.

The contextual variables showing consistently positive associations with student reading performance in both baseline and midline data are found in the category of home and school support to student learning expressed in the form of encouragement for success and constructive critique for failure.

For example, consistently significant variables from the category of family support were expressed as *Last time I had good marks family congratulated and encouraged me*, … *family hugged and kissed me*, and … *family gave me a gift*. In the midline analysis, *the last time got good marks family/guardian knew* was also positively significant but *last time had not good marks family showed no reaction* the association with student performance was significantly negative.

On the side of teacher support, consistently significant positive association was found in both analyses for *When lessons well learned teacher admires me*, and additionally in midline analysis, significant positive associations were found with *When lessons well learned teacher asks my classmates to applaud for me* and with *When cannot answer question the teacher repeats the question to me*. Significant negative association with

student performance was found when a teacher uses adverse reactions to failure, such as *When cannot answer question the teacher forces me to stand in the class*.

Another set of contextual variables that showed consistently significant associations with student reading performance come from the category of home resources that typically constitute the elements of socio-economic status. A consistent positive association was found with the questions such as *Besides the school books, do you have other books (story books, magazine, newspaper, etc.) at home for reading?* and *Have you eaten something before coming to school today?* Also, in the midline analysis, a significant positive association was found with *attending kindergarten before school*.

On the other hand, it is interesting that the variable *Can your mother read and write?* showed a negative association in both baseline and midline analyses. A deeper look into the univariate results for this variable suggests it is likely that mother's literacy behaves within multiple regression as a special variable type called 'variable suppressor'. This label denotes a variable that has low or no direct correlation with criterion but is significantly correlated with other good predictors (thus identifying the unpredictive part of their variance), which can make it a significant contributor to the prediction, in which case it gets a negative regression coefficient.

Other consistently significant associations were found between student motivation for school, such as *Why I like school: In order to have a better future*, and consistent school attendance, expressed through the question *Absent last week: No, I wasn't absent*. Interestingly, in the midline analyses, if the students reported reasons for absence as *I was ill* or *had work to do outside of home*, the associations with their reading performance were positive, tentatively suggesting that these reasons for absence might be more common among better performing students.

Table 11: Results of Multiple Regression Analyses for Student Background Factors

Predictor Variables	Baseline		Midline	
	Stand. Coeff.	Sig.	Stand. Coeff.	Sig.
Like school: Because I learn something new	.078	.000	-.012	.644
Like school: In order to have a better future	.107	.000	.118	.000
Like school: In order to become a good human	.118	.000	-.019	.418
Do you still attend lessons in the mosque after you are admitted to school?	.047	.004	.039	.079
Have you attended kindergarten before school?	.022	.167	.049	.031

Predictor Variables	Baseline		Midline	
	Stand. Coeff.	Sig.	Stand. Coeff.	Sig.
Does the teacher check your homework notebook (Dari subject)?	.053	.001	-.017	.444
When lessons well learned teacher: He admires me	.050	.005	.149	.000
When lessons well learned teacher: He asks my classmates to applaud for me	-.020	.219	.078	.004
When lessons well learned teacher: He awards me (gives me a gift)	.090	.000	.031	.274
When cannot answer question the teacher: The teacher asks the question in another way to explain in better	.083	.000	.028	.220
When cannot answer question the teacher: The teacher encourages me to try it once more	.042	.022	-.011	.670
When cannot answer question the teacher: The teacher repeats the question to me	-.004	.828	.084	.000
When cannot answer question the teacher: The teacher beats me	-.070	.000	-.006	.807
When cannot answer question the teacher: The teacher forces me to stand in the class	.014	.379	-.059	.009
Absent last week: No, I wasn't absent	.117	.000	.116	.000
Absent last week: Yes! Because I was ill	.009	.644	.073	.007
Absent last week: Yes! Because I had work to do outside home	-.008	.630	.070	.002
Absent last week: Yes! Because the weather was bad	.055	.001	-.001	.947
Absent last week: I had been to a wedding	.099	.000	.021	.379
Can your mother read and write?	-.044	.008	-.093	.000
The last time you had got good marks, did your family/guardian know it?	-.001	.960	.082	.001
Last time had good marks family: They congratulated and encouraged me	.136	.000	.119	.000
Last time had good marks family: They hugged and kissed me	.092	.000	.141	.000
Last time had good marks family: They took me for sightseeing	.047	.004	.026	.259
Last time had good marks family: They gave me a gift	.118	.000	.196	.000
The last time you had not got good marks, did your family/guardian know it?	-.048	.020	.035	.169
Last time had not good marks family: They showed no reaction	.013	.482	-.053	.030
Last time had not good marks family: They encouraged me to work hard	.091	.000	.032	.468
Does anyone help you with your homework?	-.065	.000	.047	.283
Does anyone at home ask you to read a book aloud? (It is about Dari book)	.003	.850	.041	.107

Predictor Variables	Baseline		Midline	
	Stand. Coeff.	Sig.	Stand. Coeff.	Sig.
Besides the school books, do you have other books (story books, magazine, newspaper, etc.) at home for reading?	.074	.000	.126	.000
Have you eaten something before coming to school today?	.042	.009	.057	.012
Do you have electricity at home?	.050	.002	.038	.093
Afraid of anything at school: Yes. I am afraid of students of other classes.	-.044	.007	-.039	.077
Afraid of anything at school: War, Fighting	-.049	.007	-.036	.104
Were you happy/have you laughed yesterday?	.036	.024	.001	.963

Note: Stand. Coeff. = Standardized Coefficient; Sig. = Significance

Teacher Background Factors – Results of Baseline and Midline Multiple Regressions

The comparative baseline-midline results of multiple regression analysis for teacher contextual variables are presented in Table 12. There were 18 teacher questionnaire variables that showed significant associations (at p<=0.05) with student reading performance in the baseline analysis, 19 variables were significant in the midline analysis, and among them, there were 8 variables that appeared significant in both analyses. Although all the significant background factors should be considered by policy makers, those that appear consistently significant in multiple studies should be prioritized for planning actions.

Consistent positive associations with student reading performance in both baseline and midline analyses are found when teachers report high expectations about their students (indicating expectations that students should write and read simple texts in earlier grades), and when they report using the following practices: *arranging meetings with other teachers when facing problems in teaching*; *frequency by which a principal, teaching deputy, headmaster or department head check their teaching process*; when they *prepare for new lesson by arranging study aid items*; when they report *writing the student's book text as method for teaching reading*, and when they report that *story books used in the class were prepared by government agency*.

Other background factors significantly associated with student performance in both baseline and midline analyses are teacher gender (being male is associated with lower performance of students, likely indicating that male teachers predominantly work in remote low-performing rural areas), and teacher age (indicating that older teachers are associated with higher performing students). Additionally, in the midline analysis, significant positive association was found with teacher level of education and teacher's length of experience of working as a teacher (which is consistent with the finding for their age).

Interesting negative associations with reading performance in the midline analysis were observed when teachers report that they encourage students verbally to *make further effort, give [him/her] a gift*, or *get in touch with [his/her] family*. It is plausible to suppose that these types of student encouragements are more typical in contexts where students have low reading performance. Teachers reporting that actions taken for absent students were *getting in touch with the student's family* was also negatively associated with student performance, which suggests these actions are taken in the areas of prevalent absenteeism.

And finally, uniquely for the midline analysis, positive associations were found with *teachers handling misbehaving students by providing counseling and advice*, and *with frequency they use the teacher guide* (note that the regression coefficient for the latter variable is negative, but since the frequency scale is inversed it denotes a positive association).

Table 12: Results of Multiple Regression Analyses for Teacher Background Factors

Predictor Variables	Baseline		Midline	
	Stand. Coeff.	Sig.	Stand. Coeff.	Sig.
School type	-.039	.035	-.012	.670
Are you a permanent teacher for this class at this school?	.040	.027	-.025	.334
Gender	-.078	.000	-.068	.008
How old are you?	.058	.005	.063	.021
What is your level of education?	.020	.317	.116	.000
How long have you been working as a teacher?	.024	.251	.053	.052
Do you receive your monthly salary on regular basis?	.023	.221	.049	.078
Training on children rights: Yes	.050	.006	-.022	.356
Action taken absent students: I get in touch with the student's family	.004	.846	-.054	.024
What to do misbehaving students: Providing counseling and advice	-.005	.824	.076	.006
What to do misbehaving students: Getting in touch with the family	.061	.004	-.008	.771
What to do misbehaving students: Other treatment	.047	.040	-.039	.148
Encourage students: I write encouraging remarks in [his/her] notebook	.067	.001	-.043	.204
Encourage students: I encourage [him/her] verbally, and encourage him to make further effort	-.031	.119	-.090	.003
Encourage students: I give [him/her] a gift	.014	.525	-.137	.000
Encourage students: I get in touch with [his/her] family	-.006	.759	-.071	.013
Consults when facing problems teaching: I arrange a meeting with other teachers	.076	.000	.083	.001

Predictor Variables	Baseline		Midline	
	Stand. Coeff.	Sig.	Stand. Coeff.	Sig.
During the past one month, how many times did the principal, teaching deputy, headmaster or department head check your teaching process?	.041	.023	-.102	.000
Advice received from supervisors: I have received no help whatsoever	.003	.879	.189	.000
How prepares for new lesson: I arrange study aid items	.042	.043	.057	.023
Items used as teaching aid: I use materials that I make made myself	.059	.002	.006	.825
Methods used teaching reading: Writing the student's book text.	.054	.002	.111	.000
Methods used teaching reading: Reading aloud individually	-.025	.193	-.033	.232
Methods used teaching reading: Reading silently	.064	.001	-.027	.270
Teacher guide book problems: Guide [book] has spelling and writing problems	-.052	.009	.004	.882
Do you have stories' book in the class?	-.012	.617	.090	.002
Who prepared story books: School's administration	.049	.020	.025	.434
Who prepared story books: Government agency	.038	.066	.078	.003
Use of students results: I do not use them at all	-.046	.011	-.009	.721
What grade students do you expect to be able to write a simple text?	-.088	.000	.015	.624
What grade students do you expect to be able to write a simple passage?	.045	.059	-.102	.001
How frequently have you been using the teacher guide?	n/a		-.094	.000

Note: Stand. Coeff. = Standardized Coefficient; Sig. = Significance

Principal Background Factors – Results of Baseline and Midline Multiple Regressions

The comparative baseline-midline results of multiple regression analysis for the principal context variables are presented in Table 13. There were 11 principal questionnaire variables that showed significant associations (at p<=0.05) with student reading performance in the baseline analysis, 14 variables were significant in the midline analysis, and among them, there were just 3 variables that appeared significant in both analyses. Although all the significant background factors should be considered by policy makers, those that appear consistently significant in multiple studies should be prioritized for planning actions.

Consistently in both baseline and midline analyses the principal's gender was significantly associated with student performance indicating that male principals are associated with lower performing students. This may suggest that male principals are more likely to work in remote rural areas with typically lower performing students.

The other two variables that were significant in both baseline and midline analyses were not consistent in terms of the association direction. When principals reported that they *ensure that students learn by monitoring the examination results of students*, it was positively associated with student reading performance in the baseline, and negatively in the midline regression analysis. Similarly, the *actions during inspections of the members of the monitoring team: advice on using different teaching methods* was positively associated with student reading performance in the baseline, and negatively in the midline regression analysis. This inconsistency between the midline and baseline finding calls for further exploration in the field.

Other significant positive associations between principals' background variables and student performance in the midline only include: *in case of teacher absence principal takes the students to other classes or distributes the students among other classes*; *when the school has education discipline/order regulations*; *where the Ministry of Education distributed books to all students*; and when they report that *students in grades 2-5 use the school library*.

Finally, negative associations found in the midline only include: *average hours a day students study at school* (paradoxical finding); *ensuring that students learn by monitoring the class and by checking students' homework* (likely principals in low performing areas need to do that more than principals in higher performing schools); when they report that *council actions on maintenance have not helped at all*; and when *during inspections advices were received on the preparation of a teaching plan*.

Table 13: Results of Multiple Regression Analyses for Principal Background Factors

Predictor Variables	Baseline		Midline	
	Stand. Coeff.	Sig.	Stand. Coeff.	Sig.
Gender	-.132	.000	-.110	.001
Does your school have education discipline/order regulations?	.009	.714	.086	.003
Average hours a day do students study at school	-.018	.418	-.269	.000
Average number of teachers in grades 2-5	.093	.001	.018	.574
If teacher is absent: I take the students to other classes	.032	.151	.083	.009
If teacher is absent: I bring them a backup teacher	-.059	.035	-.018	.552
If teacher is absent: I distribute the students among other classes	.031	.184	.073	.053
Ensure students learn: I monitor the class	-.014	.567	-.181	.000
Ensure students learn: I monitor the examination results of students	.066	.003	-.094	.007
Ensure students learn: I check the students homework	.022	.350	-.177	.000

Predictor Variables	Baseline		Midline	
	Stand. Coeff.	Sig.	Stand. Coeff.	Sig.
In the beginning of the current academic year, did the Ministry of Education distribute books to all your students?	-.002	.942	.107	.000
Use of library in grades 2-5	.005	.860	.107	.000
What grade do you expect students to be able to read a simple text fluently?	-.099	.000	.014	.678
When did you hold the last council meeting?	.091	.000	.054	.062
Council actions on maintenance: it has not helped at all	-.006	.824	-.118	.000
Council actions on maintenance: construction of the classes	.049	.049	.050	.110
Council actions on administrative: follow up on absences of teachers	.056	.017	-.024	.434
Actions during inspections: advice on using different teaching methods	.083	.001	-.151	.000
Actions during inspections: advice on different methods of student assessment	-.103	.000	-.007	.822
Actions during inspections: advice on the preparation of a teaching plan	-.029	.297	-.100	.001
Does your school have a safe environment?	.038	.148	-.054	.079
Have the students received any training on how to keep safe on their way to school?	-.048	.034	.001	.980

Note: Stand. Coeff. = Standardized Coefficient; Sig. = Significance

GENERAL CONCLUSIONS AND IMPLICATIONS

There are important features of the EGRA 2018 midline in Afghanistan that impose limitations on the scope of conclusions and implications that can be drawn from the results of the administration. For a start, data were only collected in the Province of Herat for the Dari language, with the socio-political situation in Nangarhar and Langham impeding the collection of data for Pashto language. This was a significant weakening of the power of the *Afghan Children Read* project's study of early grade reading acquisition in the two major national languages of Afghanistan. The second limitation is caused by the fact that the midline focused only on Grade 2 students, rather than on the two consecutive grades 2 and 3 of the baseline. Grades 2 and 3 are important years in the acquisition of early grade reading skills. We summarized the importance of these two grades in our baseline report in the following way:

> One would hope to see results in Grade 2 that would suggest that students are well on their way to acquiring solid reading skills – and by reading we refer to the acquisition of skills to comprehend key information contained in a grade-level text and on a topic that students are expected to be able to relate to. These expectations should be clearly consolidated by the end of Grade 3, a grade by which it is important for students to be able to read in order to ensure their ongoing involvement and success in the education system. By Grade 4 we expect students to be able to "read to learn", whereas the emphasis in earlier grades has been more on "learning to read". Where students fail to acquire effective reading skills by the end of Grade 3, we typically see students abandoning the school system in the following grades.

Another major difference with the baseline administration that needs to be taken into consideration is the fact that the midline was carried out after a year of implementation of a program of pedagogical interventions given to the schools participating in the midline that were designed to improve reading skills. The baseline scores clearly represented levels of reading ability of students at the outset of the program of interventions whereas the midline should reflect the value-added of such a program of support.

Having said that, let us now proceed with a summary of the key results of the 2018 midline for public schools. As in the report itself, we will provide results in terms of mean scores as well as percentages of students on each of the performance levels. We will also provide results with reference to zero scores on key variables as this offers another perspective on levels of achievement.

In Table 14 below, we provide a compilation of the summary and conclusions drawn for each of the groups for which midline test scores were produced.

Table 14: Summary and Conclusions for Herat

Summary and conclusions regarding overall 2018 midline scores
• Scores on all 4 timed tasks (SSI, IWR, FWR, and ORF) significantly improved from baseline to midline. • Corresponding zero scores decreased on 3 of the 4 timed tasks (except on SSI). • Scores on all 3 untimed tasks (LC, PA, RC) increased from baseline to midline, with significant differences on all 3 tasks, with corresponding significant decreases on zero scores. • 50% of Grade 2 Dari students still scored zero on RC (as compared to 55% on baseline) while on ORF 19% of students scored zero (as compared to 26% on the baseline). • ORF is at 35 words-per-minute on the midline (as compared to 32 on the baseline) while RC is at 27% (as compared to 23% on the baseline). • Overall, gains from baseline to midline are very modest – there is an increase of 2% of students achieving proficient levels ("average" and "good") – but 64% of students still do not meet expected levels, i.e. are either at "non-reader (19%) or "weak" (45%) levels.
Summary and conclusions regarding EGRA 2018 midline scores by Gender
• Boys have made significant gains from baseline to midline on all 7 EGRA sub-tasks while girls have made significant gains on only one sub-task (LC). • There is no significant difference in the scores on the midline between boys and girls, in other words scores are much the same. On the baseline, however, there were significant differences in favor of girls, which disappeared on the midline as boys increased their scores much more than girls did. • In terms of the performance standards, there is a significant increase in the percentage of boys achieving at the targeted levels of "average" and "good" (38% on the midline compared to 29% on the baseline); the percentage of girls scoring at the targeted levels decreased slightly (from 39% to 36%) not reaching statistical significance. • Girls' achievement on the midline on RC and ORF is the same as that of the boys, but boys have significantly under-performed on the baseline.
Summary and conclusions regarding EGRA 2018 midline scores by School Location
• Urban schools improved their performance from baseline to midline on all 7 EGRA sub-tasks, with statistically significant differences on all measures. • Rural schools did not improve scores on timed tasks, but they did improve on 2 of the 3 untimed tasks (LC and PA).

- Urban schools outperformed rural schools on all 7 measures on the midline at statistically significant levels.
- On ORF, urban schools achieved 49 words-per-minute, an increase of 8 words-per-minute from baseline to midline; rural schools remained the same at 29 words-per-minute from baseline to midline. Similar significant differences were found on RC in favor of urban schools, with urban schools comprehending at 39% on the midline (an increase from 30% from the baseline), while rural schools did not show significant change.
- In terms of performance standards, 56% of urban students achieved scores in the top 2 targeted levels (an increase of 12% from the baseline), while for rural schools the percentage of students at top 2 levels stood at 28% (insignificant change from 30% on the baseline).

Summary and conclusions regarding EGRA 2018 midline scores by School Type

- Public schools have made clear gains on all 7 EGRA sub-tasks from baseline to midline, with ORF at 35 words-per-minute (from 32 on the baseline) and RC at 27% (from 23% on the baseline).
- When results for public schools are viewed by reference to performance standards, on the midline 36% of students achieve "average" and "good" levels (up from 34% on the baseline), meaning that 64% of students are still not achieving targeted levels. Non-readers fell from 26% to 19%.
- The evidence for CBEs is mostly in the opposite direction from public schools – scores significantly decreased on ORF and LC, on other sub-tasks the changes were not statistically significant although a downward trend could be observed on most of them.
- Only 18% of CBE students achieved at the targeted levels on the midline (versus 26% on the baseline), meaning that 82% of CBE students are not reaching targeted levels. 47% of CBE students are non-readers, an increase from 38% on the baseline.

On the midline, ORF increased to 35 words-per-minute (compared with 32 on the baseline). RC is at 27%, an increase of 4% on the baseline. 50% of students scored zero on RC (a decrease of 5% from the baseline which was at 55%), and on ORF 19% of students scored zero compared with 26% on the baseline. On the midline 36% of students are performing at the "average" and "good" levels, the remainder (64%) at the "weak" (45%) and "non-reader" (19%) levels. There is an improvement of 2% of students at the upper 2 levels on the performance scale.

These are indeed modest gains. Given that these scores represent end-of-grade 2 achievement, where the benchmarks for ORF have been set for Dari at 40-64 words-per-minute at the "average" level, students will need to double the gains on ORF to move out of the "weak" level and break into the "average" level.

When we examine midline results at the sub-group level, given the overall results indicated above, there is nothing really surprising about results at any of the sub-group levels except for students in urban schools. Urban school students are performing on the midline at 49 words-per-minute (8 words higher than on the baseline), with comprehension at 39%, an increase from 30% on the baseline. These scores place urban students well into the "average" level on the performance scale. In fact, 56% of urban students are performing in the top 2 levels of the performance scale compared to 36% overall at these 2 levels.

Clearly achievements made by urban students on the midline represent the sort of levels of achievement that one would have hoped for from all students. There is a significant difference between achievement of urban students versus rural students – for example, ORF for rural students is at 29 words-per-minute, comprehension at 21%; 28% of rural students perform at the top 2 levels on the performance scale, a decline of 2% from the baseline).

The results demonstrated by urban students (who quite clearly have significantly better conditions under which learning can take place) and the professional development interventions that they receive need further analysis so that these can be transferred to rural contexts. Similarly, the gender difference in gains from baseline to midline suggest that boys reaped more benefits of the ongoing intervention and demonstrated stronger gains than girls, a finding that calls for further insights from local experts to understand why girls did not show the same type of improvement as boys.

REFERENCES

USAID/Afghanistan. (n.d.). Afghanistan Education. Retrieved from United States Agency for International Development: https://www.usaid.gov/afghanistan/education

Central Intelligence Agency. (n.d.). The World Factbook/Afghanistan. Retrieved from Central Intelligence Agency: https://www.cia.gov/library/publications/the-world-factbook/geos/af.html

Cohen, Jacob. (1988). Statistical power analysis for the Behavioral sciences. SERBIULA (Sistema Librum 2.0). 2nd.

Comeau, Liane; Cormier, Pierre; Grandmaison, Éric; Lacroix, Diane. (1999). A Longitudinal Study of Phonological Processing Skills in Children Learning to Read in a Second Language. Journal of Educational Psychology.

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. International Journal of Educational Development, 40, 315–322.

Geva, E., & Genesee, F. (2006). First-Language Oral Proficiency and Second-Language Literacy. In D. August & T. Shanahan (Eds.), Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth (pp. 185-195). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Gove, A. and P. Cvelich. 2011. Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice. Revised Edition. Research Triangle Park, NC: Research Triangle Institute.

Hatsaandh, A. H. (2019). *The Need for Bilingual Education in Afghanistan – OpEd*. Retrieved from EurAsia review: https://www.eurasiareview.com/23042019-the-need-for-bilingual-education-in-afghanistan-analysis/.

Hoover, W., & Gough, B. (1990). The simple view of reading. Reading and Writing: An Interdisciplinary Journal 2, 127–160. The Netherlands: Kluwer Academic Publishers.

Joshi, R. Malatesha; Tao, Sha; Aaron, P. G.; Quiroz, Blanca. (2012). Cognitive Component of Componential Model of Reading Applied to Different Orthographies. Journal of learning disabilities. 45. 480-6.

Kim, J., & Mueller, C. (1978). Introduction to factor analysis: What it is and how to do it. Quantitative Applications in Social Sciences. Sage University Press.

Lervåg, Arne & Grøver Aukrust, Vibeke. (2009). Vocabulary knowledge is a critical determinant of the difference in reading comprehension between first and second language learners. Journal of child psychology and psychiatry, and allied disciplines.

Lesaux, Nonie K., Koda, Keiko; Siegel, Linda; Shanahan, Timothy. (2006). Development of Literacy, Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth (pp. 75-122). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

National Reading Panel. (2000). Teaching Children To Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction. Bethesda, MD: National Reading Panel.

Proctor, C. P., August, D., Carlo, M., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, *97*(2), 246-256.

RTI International (2016). Early Grade Reading Assessment (EGRA) Toolkit, Second Edition. Washington, DC: United States Agency for International Development.

Vaughn, S., & Linan-Thompson, S. (2004). Research-based methods of reading instruction grades K-3. Alexandria, VA: Association for Supervision and Curriculum Development.

Verhoeven, Ludo & van Leeuwe, Jan & Vermeer, Anne. (2011). Vocabulary Growth and Reading Development across the Elementary School Years. Scientific Studies of Reading - SCI STUD READ. 15. 8-25.

Appendix 1. The EGRA Subtasks

The Phonemic Awareness (PA) subtask determined the ability of students to identify and produce the beginning sound of each of 10 words. The test administrator read each word twice and asked the students to make the first sound of the word. If a learner did not answer within 3 seconds, a "no answer" response was recorded. The maximum score for this section was 10 points, 1 point for each correct answer. This was an untimed activity.

The Syllable Sound Identification (SSI) subtask was designed to measure the ability of students to produce the sounds of the written form of a series of syllables. The test administrator asked the students to make the sound of each of the syllables shown to them. If a learner did not answer within 3 seconds, a "no answer" response was recorded. The SSI tasks were scored on a syllables-per-minute calculation to determine how many syllables were attempted, how many were read correctly, and in what time over the course of 60 seconds. The SSI task included 100 syllables.

The Invented Word Reading (IWR) subtask assessed the ability of students to decode one- to two-syllable non-words that could plausibly exist in the target language. The IWR task provided a measure of decoding related to that of the Familiar Word Reading task but had the advantage of not allowing respondents to sight-read words, that is, the task provides a measure of a learner's ability to apply the phonological rules of the target language in new unseen contexts. The IWR task was scored on a words-per-minute calculation to determine how many words were attempted, how many were read correctly, and in what time over the course of 60 seconds. The IWR task included 50 words.

The Familiar Word Reading (FWR) subtask assessed a learner's ability to recognize and read high-frequency age and grade appropriate words. This subtask presents a list of unrelated words that are not presented as a story or complete text; the words were randomly arranged on a card that formed part of the learner stimulus. The FWR 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, and in what time over the course of 60 seconds. The FWR task included 50 words.

The Oral Reading Fluency (ORF) subtask can be best understood as a measure of a learner's ability to read with speed and accuracy a passage made up of grade-appropriate words (familiar words) and presented in the learner stimulus booklet. The ORF task is "oral" in that students read the passage aloud. The ORF 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, and in what time over the course of 60 seconds. The ORF subtask included a paragraph with 66 words.

The Reading Comprehension (RC) subtask, which relied on questions about the text read in the Oral Reading Fluency subtask, measured a learner's understanding of the text he or she has just read through a series of factual and inferential questions. The RC subtask included five untimed questions related to the ORF subtask, each with a value of 1 score point for a correct answer.

The Listening Comprehension (LC) subtask measured students' understanding of a short text, read by the test administrator, through a set of questions based on what was read to them. The LC subtask included five untimed questions, each with a value of 1 score point for a correct answer.

Appendix 2. Sampling Method

The following steps guided the random selection of students in each grade:

Stage 1 – Cluster Sampling (CS) for selection of schools

Cluster sampling was applied to the total number of students in each relevant grade (in Herat for Dari and in Nangarhar and Laghman for Pashto) for the public schools. Schools were considered clusters for efficiency purposes. The necessary CS sample size was calculated at a 95% confidence level and 5% confidence interval. The design effect (DEFF) for schools was 3, due to the large and scattered population to get a significant sample size. To avoid loss of sample due to security or other reasons and to keep intact the significance of sample and confidence level, the sample was topped up by 25%.

Stage 2 - Proportionate allocation of sampled clusters to districts

The cluster samples of schools in the province was proportionately allocated to the districts based on the district's proportional representation in the total population. This allowed appropriate representation of each districts in sample and reduce skewing towards a district or set of districts. Schools were selected using a probability proportional technique in each district as per proportionate sample size.

Stage 3 – Probability proportionate sampling for selection of random selection of schools

For the random selection of schools, EMIS data was used. The Grade 2 and Grade 3 enrollment was summed up for each school to get the total population in each school. Probability proportionate to size sampling was used for random selection of the schools using the cumulative school population. The following steps were followed to draw probability proportionate samples:

1. Cleaned the EMIS data and filter all schools where G2 & G3 population doesn't exist in the target districts. Schools with less than 10 G2 or G3 students were also be filtered as 10 students per grade were required for assessment.
2. Added cumulative population column to the filtered schools list.
3. Calculated the sampling interval for each district sampling frame. i.e.

   Sampling Interval (SI) = Population total for the district / sample size of clusters (S)

4. Chose a random starting point (RSP) for the district (using Excel function *rand between*). The RSP was taken between 1 and SI for each district.
5. The first cluster (school; sample point) selected was the one corresponding (greater than or equal) to the row in cumulative population column containing the obtained RSP number
6. Picked the rest of sample clusters from the corresponding rows of cumulative population column to the series calculated as given below:

   For 2nd Sample = cumulative population row corresponding to [*RSP + 1xSI*]

   For 3rd Sample = cumulative population row corresponding to [*RSP + 2xSI*]

   And so on and so forth until Last Sample = cumulative population row corresponding to [*RSP + (S-1) x SI*].

7. Repeated this process for each district.

Stage 4 – Random selection of students in sampled schools (clusters)

Ten students per grade per school was the sample size assessed as for the EGRA. The following steps guided the random selection of students in each grade:

1. Students were randomly selected one class for each grade (Grade 2 & Grade 3) where the number of classes for a grade was more than one. This was done through a draw method. The names of all classes were written on equal size small sheets, folded alike and mixed. Then one sheet was drawn randomly and was the class where students would be assessed.
2. For random selection of students in the selected class, average interval method was used on class attendance register. The total number of students in the attendance registered were divided by 10 to get average interval (*m*, rounded off to the nearest lower whole number). Thus, from the attendance register, every m^{th} student was picked for assessment starting from the 1st student in the attendance list.

Table 15: Final Sample Herat

No.	District	Schools		Proportionate Sample	Students to be Assessed		
		Total	% age		G2 Completers	G3 Completers	Total
1	Hirat	80	13%	19	188	188	376
2	Injil	134	21%	31	315	315	630
3	Guzara	86	14%	20	202	202	404
4	Pashtoon Zarghoon	51	8%	12	120	120	240
5	Karokh	55	9%	13	129	129	258
6	Kuhsan	38	6%	9	89	89	179
7	Ghoreyan	50	8%	12	117	117	235
8	Zenda Jan	29	5%	7	68	68	136
9	Oba	41	7%	10	96	96	193
10	Kushk (Rubatak i Sangi)	66	10%	16	155	155	310
	Total	630		148	1,480	1,480	2,960

Appendix 3. Factors Associated with Student Reading Achievement for Dari

This section presents the findings on the associations between contextual variables from the principal, teacher, and student questionnaires and performance on the Oral Reading Fluency (ORF) sub-task of EGRA. The identification of factors associated with reading performance can help the Ministry of Education and the Afghan Children Read program target interventions to those factors that can improve the reading proficiency of Afghan children.

The factors included in the analyses comprised the following: *availability of teaching materials; school monitoring activities*; *teacher and principal pedagogical activities*; *school environment*; *home environment factors*; and *safety at school*. This is not a comprehensive list; each section in this report describes the factors that were taken into consideration for each survey.

The contextual variables for principals and teachers were grouped into five categories:

1. Teachers' (or principals') personal characteristics
2. Responses about activities by others (their supervisors)
3. Responses reflecting their own pedagogical activities
4. Responses reflecting school environment factors
5. Responses about training provided by *Afghan Children Read*

The contextual variables for students were grouped into six categories:

1. Student characteristics
2. Activities by teachers
3. Student absenteeism
4. Home support
5. Home environment
6. School environment

An important note for the reader on the interpretation of the analyses is that these analyses report the associations between reading performance and contextual variables, which does not provide sufficient information to attribute cause. Thus, these contextual variables should not be viewed as factors that influence reading achievement. The associations between two variables, "Variable X" and "Variable Y", may be due to "Variable X" causing "Variable Y", or vice versa, or even some third factor Z affecting both variables leading to their correlation. Causal interpretation of the associations between contextual

variables and reading performance requires additional scrutiny and research and understanding of many circumstances that could contribute to associations between the different variables.

The analyses of the contextual factors were carried out by comparing ORF scores based on the respondents' answers to the questions in the surveys. For example, for the question "Do you enjoy reading?", the mean ORF score of the group that responded "Yes" was compared to the mean ORF score of the group that responded "No" by testing the statistical significance and calculating the associated effect size using Cohen's D between both means. To select significant associations, both statistical significance ($p < 0.05$) and practical significance (Cohen's D greater than 0.35 for teacher and principal data, and greater than 0.25 for student data) were used. Bar charts were created to better illustrate the size and direction of the differences and the corresponding statistical significances and effect sizes are presented next to the chart.

Results of Analyses of Students' Contextual Variables

One thousand eight hundred ninety-eight students responded to the background questionnaires that contain questions about the students themselves and their schools. Male respondents make up 41.2% of the sample while 58.8% are female. In relation to school type, 93.7% reported they attend public schools and 6.3% attend CBEs. In terms of mother tongue, 90.6% reported Dari as their mother tongue while 9.4% speak other languages.

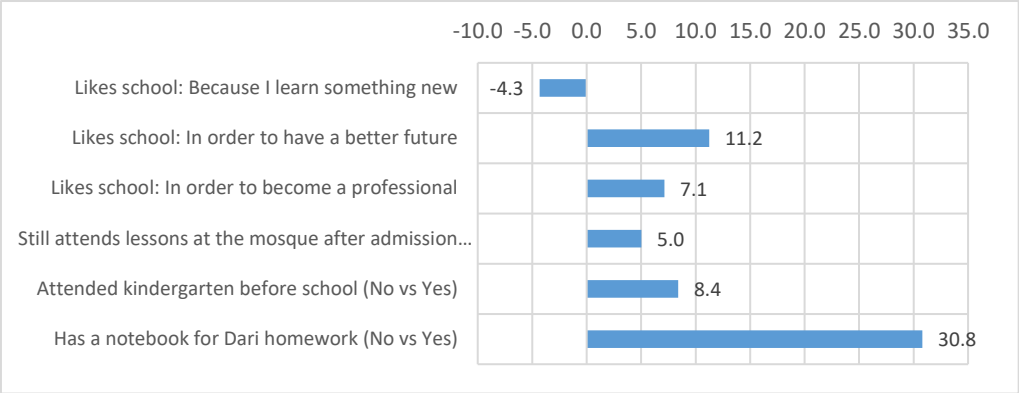
This section presents selected background variables of the student questionnaire based on their content relevance and statistical or practical significance. The variables are grouped into 6 categories: student characteristics; activities by teachers; student absenteeism; home support; home environment; and school environment.

Students' Characteristics

Among the student characteristics that showed positive associations with reading performance are: their expressions about why they like going to school, particularly the responses "In order to have a better future" and "In order to become a professional", while "Because I learn something new" is negatively associated to reading performance.

Current attendance to classes at the mosque, *attendance to kindergarten before school*, and *having a notebook for homework* also show significant association with reading performance. Table 16 presents the results of the analyses.

Table 16: Student Characteristics and their Relationship to ORF



Sig	D
0.01	-0.14
0.00	0.36
0.00	0.22
0.00	0.16
0.00	0.26
0.00	0.97

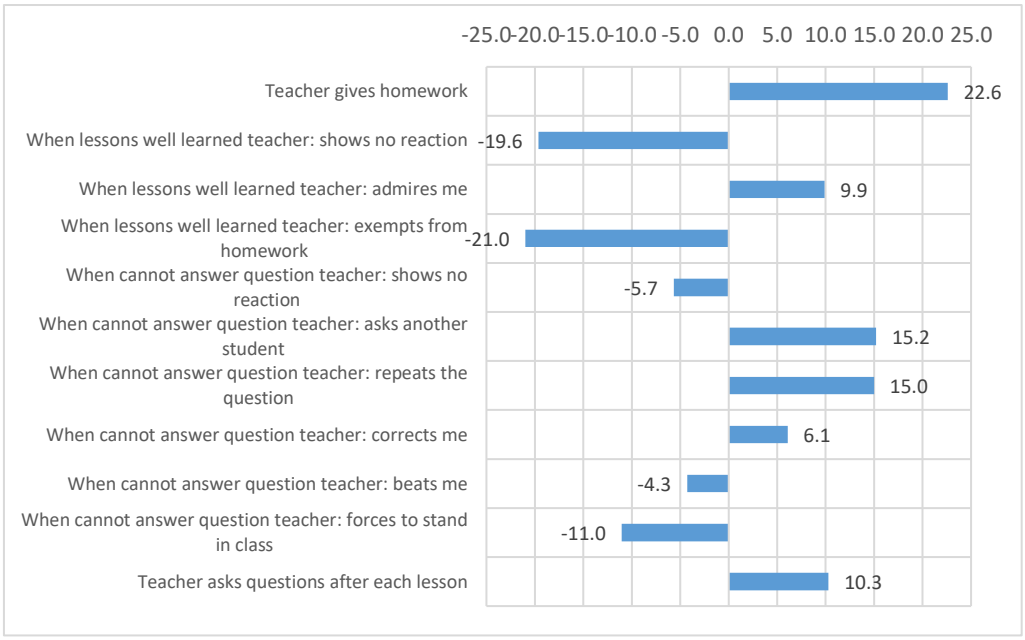
Note: Sig = Significance; D = Cohen's D

Activities of Teachers Reported by Students

Teacher activities positively associated with reading performance are: *teacher assigning homework*; *teachers admiring students when they have good marks*; *teachers asking questions after a lesson*; and *teachers' actions when the student cannot answer a question correctly*, particularly "Asking another student", "repeating the question to clarify it", and "teacher corrects if for me".

Teacher activities that are negatively associated with student performance are related to *negative reactions when the student cannot answer a question correctly* and *not encouraging students when they obtain good marks*. The associations are presented in Table 17 below.

Table 17: Activities by Teachers and their Relationship to ORF



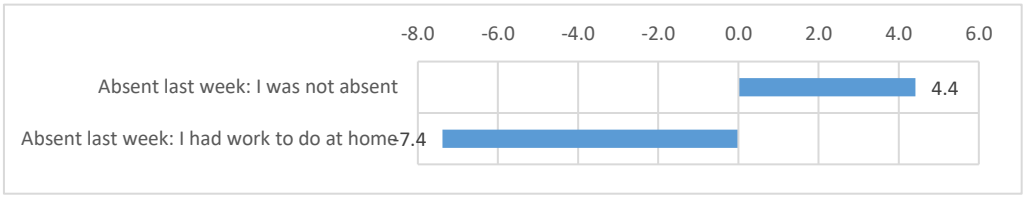
Sig	D
0.01	0.71
0.00	-0.62
0.00	0.31
0.26	-0.66
0.01	-0.18
0.01	0.48
0.00	0.47
0.01	0.19
0.02	-0.13
0.06	-0.35
0.21	0.32

Note: Sig = Significance; D = Cohen's D

Students' Absenteeism

Student attendance at school is a major pre-requisite for successful learning, this statement was confirmed by the analysis of responses related to student absenteeism. As shown in Table 18, two responses are associated with performance in reading. Students who reported they were *not absent from school* shows a positive association while those who responded they were a*bsent because they had work to do at home* is negatively associated with reading performance.

Table 18: Student Absenteeism and its Relationship to ORF



Sig	D
0.01	0.14
0.03	-0.23

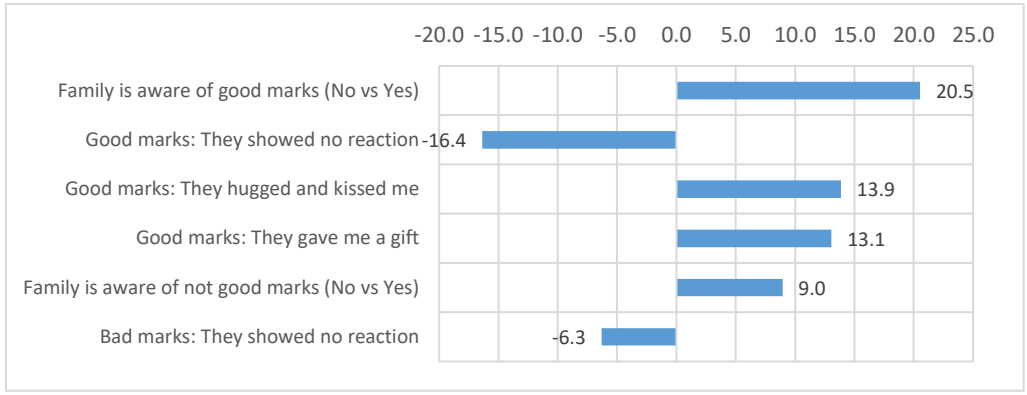
Note: Sig = Significance; D = Cohen's D

Home Support

Home support variables that show positive associations with student performance in reading are: *family being aware of the student obtaining good marks*; *family being aware of the student obtaining not good marks*; and *reactions of the family when students got good marks such as hugging and kissing the student and giving them a gift*.

The variables that showed negative association are related to the *family showing no reaction when the students received either good or not good marks*. These home support variables associated with reading performance are presented in **Error! Reference source not found.** below.

Table 19: Student Home Support and its Relationship to ORF



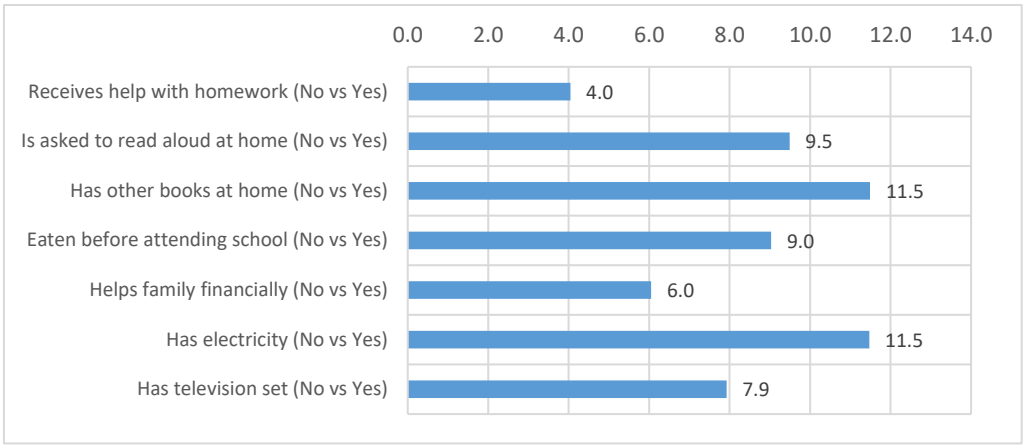
Sig	D
0.00	0.65
0.00	-0.52
0.00	0.44
0.00	0.41
0.00	0.28
0.00	-0.20

Note: Sig = Significance; D = Cohen's D

Home Environment

Seven home environment variables showed positive associations with reading performance as shown in Table 20. Variables related to educational habits are: *receiving help with homework*; *family members asking the student to read aloud at home*; and *having books at home*. In terms of socioeconomic status, the variables are: *eating before attending school*; *having electricity at home*; and *having a television set*. It is interesting to note that students who reported that they *help their family by working* is also associated with higher performance in reading.

Table 20: Student Home Environment and its Relationship to ORF



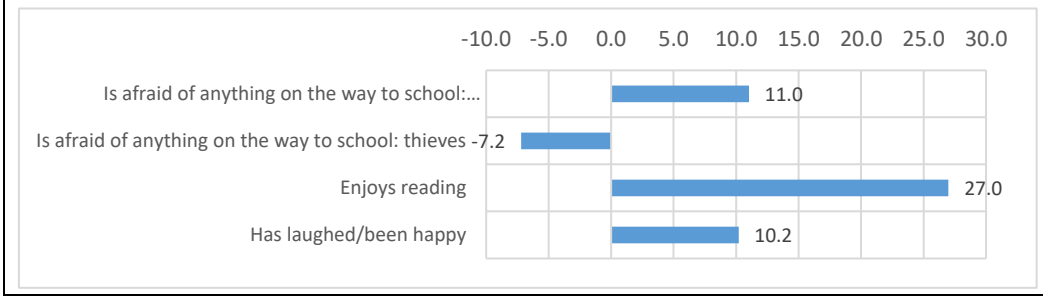
Sig	D
0.02	0.13
0.00	0.30
0.00	0.36
0.00	0.28
0.00	0.19
0.00	0.36
0.00	0.25

Note: Sig = Significance; D = Cohen's D

School Environment Reported by Students

School environment variables are frequently identified as highly relevant factors for student learning progress. The school environment variables included in the student questionnaire focused on social/emotional environment rather than physical resources. Enjoying reading and being happy, both reflecting positive reactions, are positively associated with reading performance. In terms of being afraid on the way to school, *being afraid of thieves* is negatively associated with reading performance while being *afraid of kidnapping* shows a positive association. Table 21 below presents the results of the analyses.

Table 21: School Environment and its Relationship to ORF



Sig	D
0.00	0.35
0.04	-0.23
0.01	0.85
0.06	0.32

Note: Sig = Significance; D = Cohen's D

Results of Analyses of Teachers' Contextual Variables

One hundred seventy-six teachers responded to the background questionnaires that contain questions about themselves and their teaching practices in their schools. Male respondents make up 32.4% of the sample while 67.6% are female. In relation to school type, 81.3% reported they work at public schools and 18.8% work at CBEs. In terms of mother tongue, 96.6% reported Dari as their mother tongue while 3.4% speak other languages. Regarding their level of education, 11.4% reported holding a bachelor's degree; 44.3% graduated from Grade 14, 41.5% graduated from Grade 12, and 2.8% have not finished Grade 12. In terms of attending training provided by Afghan Children Read, 97.7% reported that they have attended the training while 2.3% reported they have not.

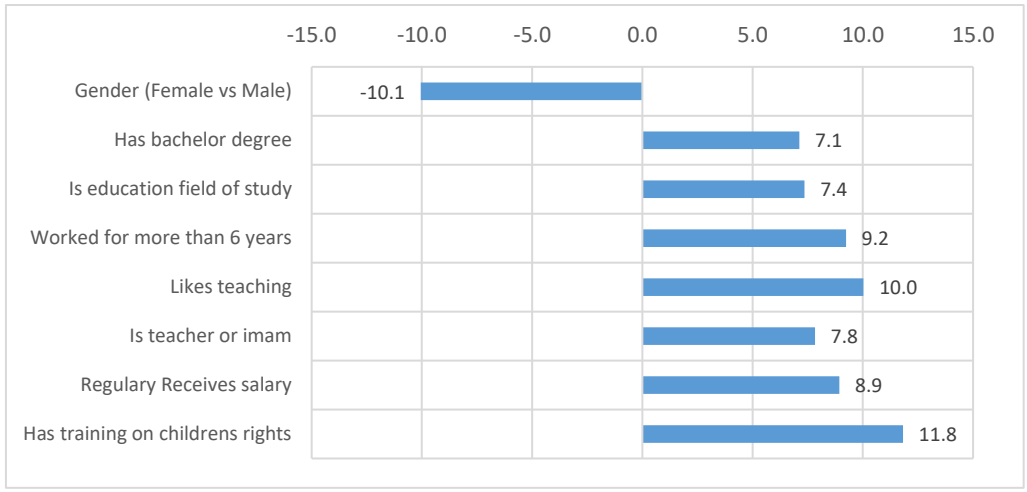
This section presents selected background variables based on their content relevance and statistical or practical significance to review the associations between the contextual variables assessed by the teachers' questionnaire and reading performance of students in their schools. The variables are grouped into 5 categories: *teacher characteristics*; *teacher activities*; *activities by teacher supervisors*; *school environment*; and *Afghan Children Read training*.

Teachers' Characteristics

Teachers characteristics positively associated with reading performance are: *educational level and field of study*; *work experience*; *reasons for becoming a teacher*; *having a second occupation*; *receiving salary on a regular basis*; and *receiving training on children's rights*.

Teachers who reported *having a bachelor's degree* are associated with higher performance in reading than compared to those who *graduated from grade 14 or grade 12*. The same applies for teachers who reported that their *field of study is education* compared to other fields. Teachers with six or more years of experience are also more positively associated than those with less experience; also associated with higher performance in reading is the fact that respondents reported they *decide to become teachers because they like the profession* compared to those that became teachers for other reasons. In terms of additional jobs, those who reported they *have a second job as a private school teacher or imam* show strong association with reading performance. Additionally, teachers that reported *receiving their salaries on a regular basis* is also associated with increased performance in reading. Finally, teachers who reported having *received training on children's rights* shows a positive association with reading performance. As is the case with principals, male teachers are negatively associated with student performance. Table 22 below shows the aforementioned associations.

Table 22: Teacher Characteristics and their Relationship to ORF



Sig	D
0.00	-0.32
0.00	0.22
0.00	0.23
0.00	0.29
0.00	0.32
0.00	0.23
0.00	0.28
0.40	0.37

Note: Sig = Significance; D = Cohen's D

Teachers' Activities

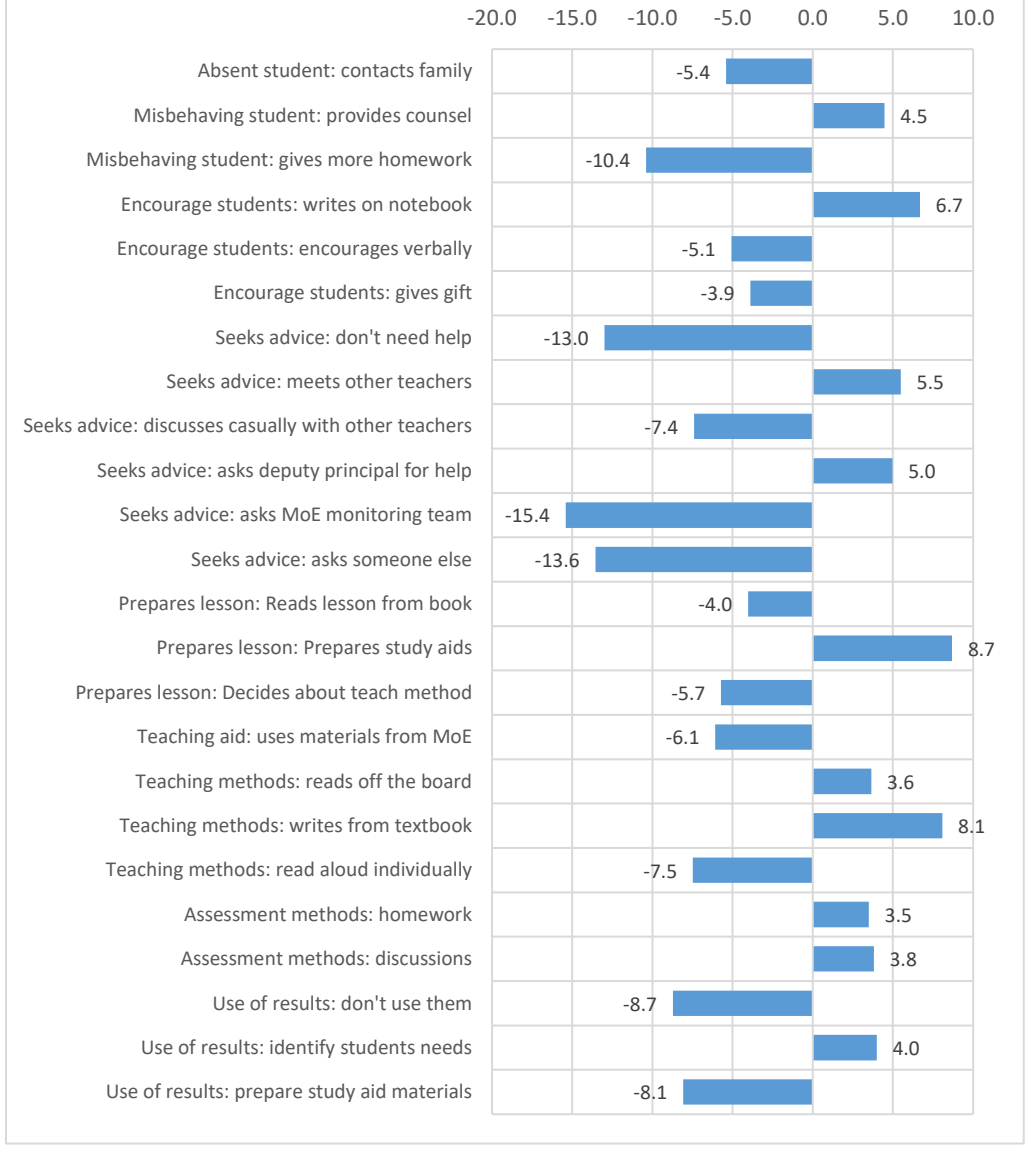
Teacher activities associated with student performance in reading are presented in Table 23. Positively associated teacher activities include: *actions to maintain discipline in class*; *actions to motivate students*; *seeking advice when facing problems*; *lesson plan preparation*; *methods for teaching reading*; *methods to assess students*; and *use of examination results*.

In terms of actions to maintain discipline in class, teachers who reported that they *provide counseling and advice* shows a positive association with reading performance. An action to motivate students that showed positive association is *writing encouraging remarks on the student notebook*. When teachers face problems while teaching, *discussing the problem with other teachers* and *asking the deputy principal for help* are also associated with reading performance. Teachers who *prepare a new lesson by arranging study aid items* and those who *read text written on the board* and *write from the textbook* as methods to teach reading are also positively associated with reading performance. Finally, teachers who *use homework and discussion as methods to assess students' education progress* and those who reported *using the results to identify student needs* are also positively associated with student performance.

Teacher activities that are negatively associated with student performance are the following: *contacting the student family when the student is absent from school*; *giving more homework when students misbehave*; *motivating students by verbally encouraging them or giving them a gift*. In terms of facing problems while teaching, activities negatively associated with reading are: *teachers reporting they do not need help*; *discussing the issue in a casual matter with other teachers*; *asking help from the monitoring team of the Ministry*

of Education and asking someone else. *Reading the lesson from the book* and *deciding about the teaching method* are also negatively associated with reading performance. Regarding the use of teaching aids and teaching methods, teachers who *use materials provided by the Ministry of Education* and who *ask students to read aloud individually* are also negatively associated with reading performance at their schools. Finally, when asked how they use examination results, *not using the results* or *using the results to prepare student aid materials* are both associated with reading performance albeit negatively.

Table 23: Teacher Activities and their Relationship to ORF



Sig	D
0.00	-0.17
0.03	0.14
0.02	-0.33
0.00	0.21
0.00	-0.16
0.01	-0.12
0.02	-0.41
0.00	0.17
0.00	-0.23
0.01	0.16
0.02	-0.48
0.01	-0.42
0.02	-0.13
0.00	0.27
0.01	-0.18
0.01	-0.19
0.02	0.11
0.00	0.25
0.00	-0.24
0.02	0.11
0.03	0.12
0.03	-0.27
0.01	0.12
0.00	-0.25

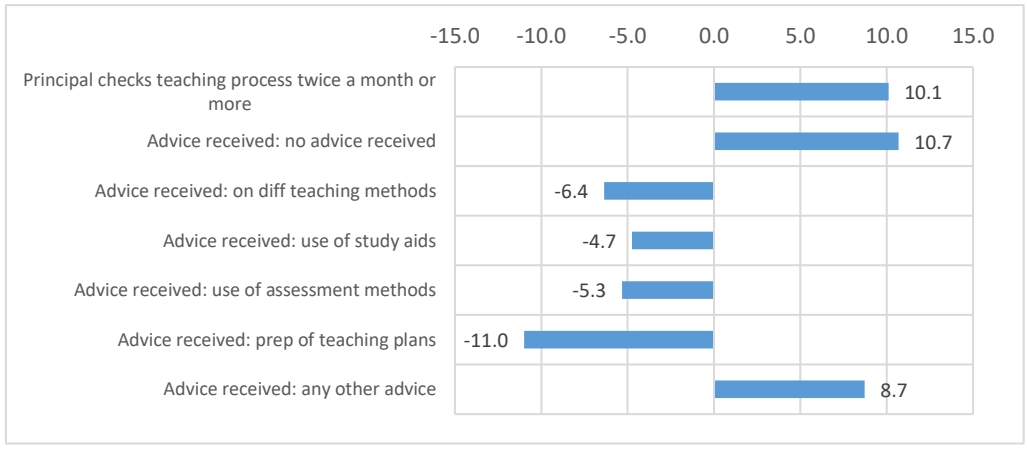
Note: Sig = Significance; D = Cohen's D

Activities by Teachers' Supervisors

Activities by teacher supervisors associated with reading performance are presented in Table 24. Teachers who reported that their *teaching process is supervised once a week or more by the principal or other school official* showed positive association with reading performance. When asked about what type of advice or assistance they have received from the monitoring team, those teachers who responded they have not received advice also showed positive association with reading performance.

Interestingly, activities that are negatively associated with reading performance are all related to receiving advice or assistance from members of the monitoring team. These activities are: *receiving advice on using different teaching methods*; *use of study aids*; *use of different assessment methods*; and *receiving advice on the preparation of teaching plans*.

Table 24: Activities by Teacher Supervisor and their Relationship to ORF



Sig	D
0.00	0.32
0.00	0.34
0.00	-0.20
0.01	-0.15
0.00	-0.17
0.00	-0.35
0.03	0.27

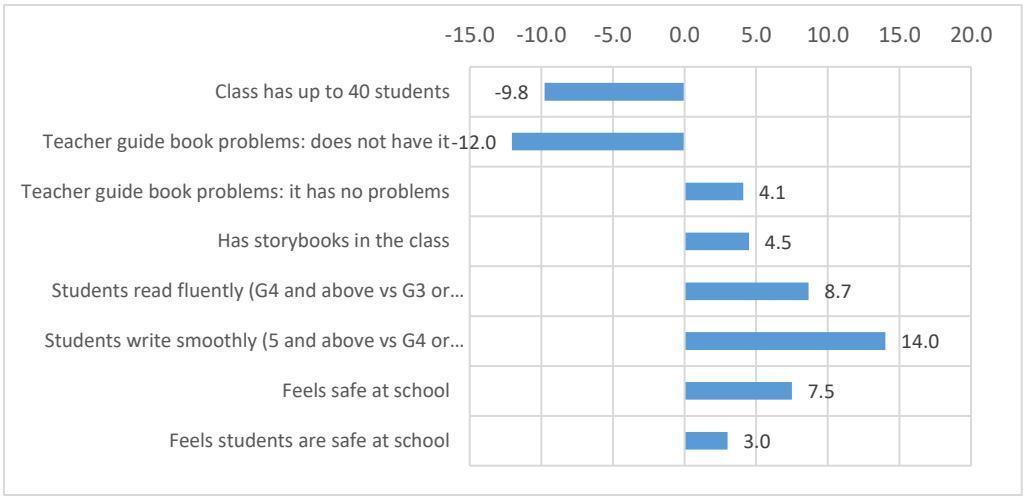
Note: Sig = Significance; D = Cohen's D

School Environment Reported by Teachers

Table 25 presents teacher responses regarding the school environment that are associated with reading performance. Aspects of the school environment that are positively associated with student performance in reading are: *reporting the teacher guidebook has no problems*; *reporting the availability of storybook in the classroom*; expecting that students will read and write fluently in early grades; and school safety for teachers and students.

Aspects of school environment that are negatively associated with reading performance are: *classes with up to forty students*; and *teachers not in possession of the teacher guidebook*.

Table 25: School Environment and its Relationship to ORF



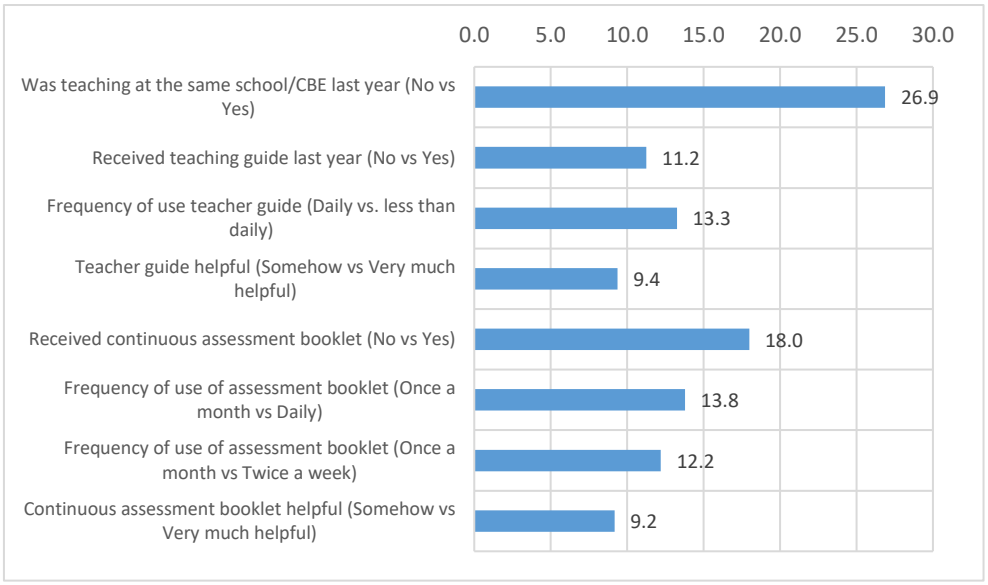
Sig	D
0.00	-0.31
0.00	-0.38
0.01	0.13
0.00	0.14
0.01	0.27
0.00	0.44
0.00	0.24
0.05	0.09

Note: Sig = Significance; D = Cohen's D

Training provided by Afghan Children Read

Teachers' responses regarding training provided by *Afghan Children Read* significantly associated with reading performance in schools are presented in Table 26. Variables that showed positive associations include: *teaching at the same school/CBE as previous year*; *frequency of use of the teacher guide (daily); availability and use of the teaching guide*; *helpfulness of the teaching guide (very much helpful)*; *availability of the continuous assessment booklet*; *frequency of use of the assessment booklet (daily and twice a week)*; and *helpfulness of the assessment booklet (very much helpful)*.

Table 26: Training Provided by *Afghan Children Read* and its Relationship to ORF



Sig	D
0.00	0.85
0.00	0.35
0.00	0.41
0.05	0.29
0.00	0.57
0.00	0.42
0.00	0.38
0.03	0.28

Note: Sig = Significance; D = Cohen's D

Results of Analyses of Principals' Contextual Variables

One hundred forty-two school principals responded to the background questionnaire that contains questions about the principals themselves and teaching practices in their schools. Of this number, 69% are male and 31% are female. Regarding their level of education, 40.8% reported holding a bachelor's degree; 43.7% graduated from Grade 14, 14.8% graduated from Grade 12 and 0.7% have not finished Grade 12. In terms of attending training provided by Afghan Children Read, 95.8% reported that they have attended the training while 4.2% reported that they have not.

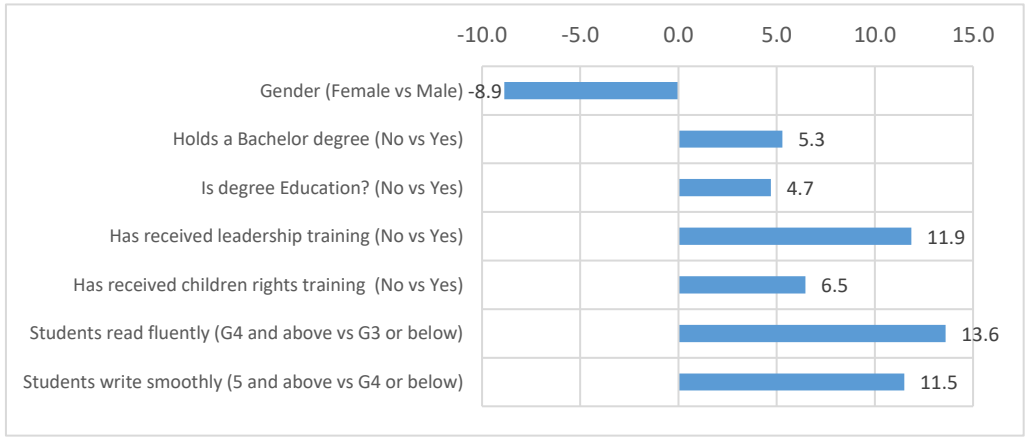
This section presents selected background variables based on their content relevance and statistical or practical significance to review the associations between the contextual variables assessed by the principal questionnaire and reading performance of students in their schools. The variables are grouped into 5 categories: principal characteristics, principal activities, activities by supervisors, school environment, and *Afghan Children Read* training.

Principals' Characteristics

The characteristics of principals that are associated with reading performance in their schools are the following: *educational degree*; *field of educational degree*; *having received training on leadership*; *having received training on children's rights*; and *principal expectations about when students should be able to read and write fluently*.

Holding a bachelor's degree or *having a degree in education* are both positively associated with reading performance at their schools. Principals who reported *having received at least one training on leadership and management* or *on children's rights* are also positively associated with higher performance in reading. In relation to the *principals' expectations about the grade by which students should be able to read and write fluently*, those who responded that students should be able to read and write fluently at or below grade 3 are positively associated with student performance in reading when compared to those that expect students to demonstrate fluency in reading and writing beyond grade 3. In terms of gender, female principals are associated with higher reading performance than their male counterparts. Table 27 below shows the results of the analyses.

Table 27: Principal Characteristics and their Relationship to ORF



Sig	D
0.00	-0.28
0.00	0.17
0.02	0.15
0.00	0.37
0.00	0.20
0.00	0.43
0.00	0.36

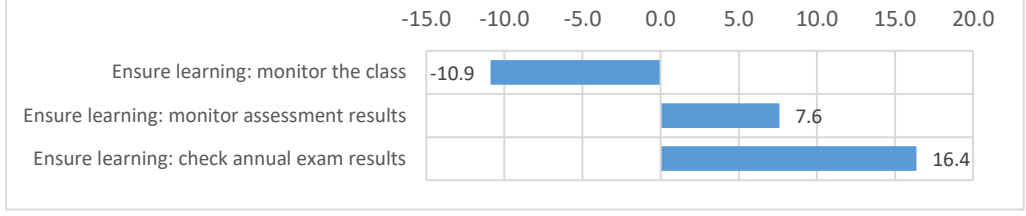
Note: Sig = Significance; D = Cohen's D

Principals' Activities

Principal activities that are associated with student performance in reading are those related to making sure students have learned what they are expected to learn.

Monitoring the examination results of students and *checking annual examinations results* are both positively associated with reading performance. On the other hand, principals who reported *monitoring classes personally* is negatively associated with student performance in reading as seen in Table 28 below.

Table 28: Principal Activities and their Relationship to ORF



Sig	D
0.00	-0.34
0.00	0.24
0.00	0.51

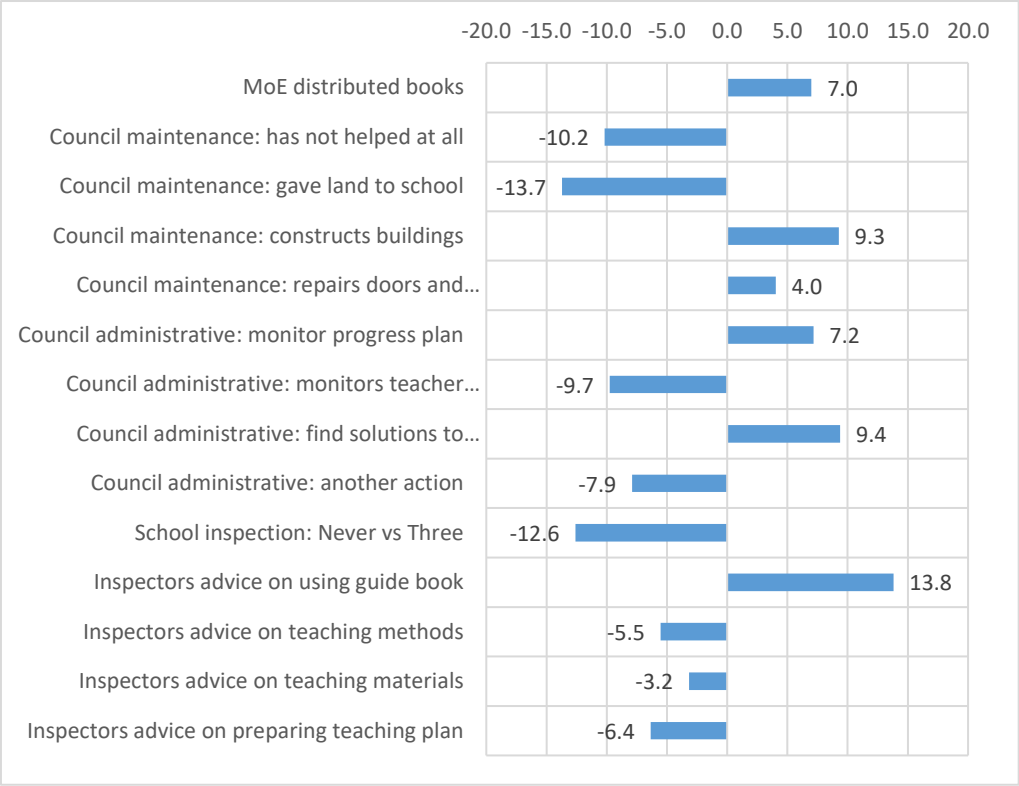
Note: Sig = Significance; D = Cohen's D

Activities by Principals' Supervisors

The activities of school supervising bodies that are associated with reading performance in schools are presented in Table 29. There is a positive association with reading performance when principals reported that *the Ministry of Education distributed books to all students at the beginning of the academic year*. *School council actions related to constructing school buildings and repairing doors and windows* are also positively associated with reading performance. In terms of administrative and academic affairs, actions of the school council, *monitoring the progress plan of the school*, and *finding out about students' problems and discussing their solutions* are also positively related to reading performance. Regarding school inspections by members of the monitoring team, the only action that was positively associated with reading performance is *providing advice on using the guidebook*.

There are also activities by the principals' supervisors that are negatively associated with reading performance. School council actions on school building maintenance with negative associations are: *not helping at all* and *giving land to the school*. *Monitoring teacher absences in a school council action on administrative and academic affairs* also has a negative association. In terms of actions by members of the monitoring team, those negatively associated with reading performance are: *frequency of school inspections*; *providing advice on teaching methods*; *providing advice on teaching materials*; and *providing advice on preparing teaching plans*.

Table 29: Activities by Principals' Supervisors and their Relationship to ORF



Sig	D
0.00	0.22
0.00	-0.32
0.03	-0.43
0.00	0.29
0.05	0.13
0.00	0.22
0.00	-0.30
0.00	0.29
0.00	-0.25
0.06	-0.42
0.00	0.43
0.00	-0.17
0.04	-0.10
0.00	-0.20

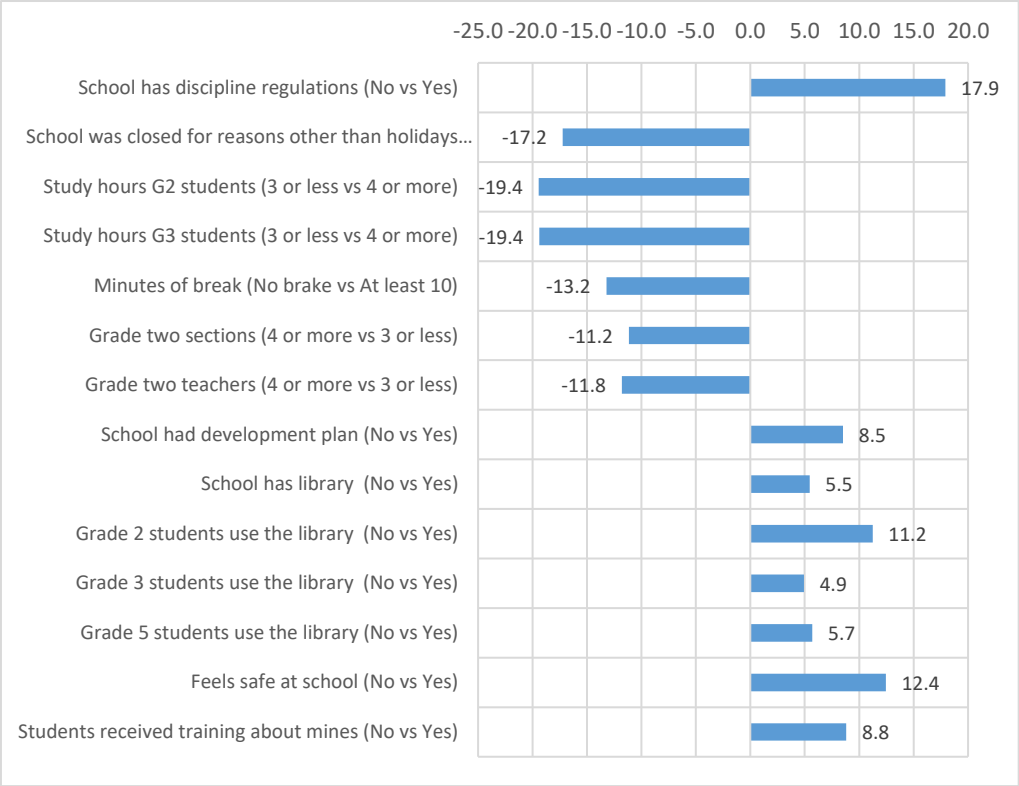
Note: Sig = Significance; D = Cohen's D

School Environment Reported by Principals

School environment variables that are positively associated with reading performance are: *having discipline/order regulations at school*; *implementation of a development plan*; *existence and use of the school library*; *principals feeling safe at school*; and *students receiving training about mines and unexploded devices*. These aspects are presented in Table 30 below.

Principals who reported their *schools have and implement discipline regulations* is strongly associated with reading performance, as is *having designed and implemented a school development plan during the previous academic year*. The *existence of a school library* is positively associated with reading performance, as is *use of the library by grade 2, 3, and 5 students*. Regarding safety at school, principals who reported *feeling safe at school* and that *their students received training about mines and unexploded devices* also show significantly positive associations with reading performance.

Table 30: School Environment and its Relationship to ORF



Sig	D
0.00	0.56
0.03	-0.54
0.00	-0.62
0.00	-0.62
0.00	-0.42
0.00	-0.35
0.00	-0.37
0.00	0.27
0.00	0.17
0.00	0.35
0.01	0.15
0.02	0.17
0.00	0.39
0.00	0.28

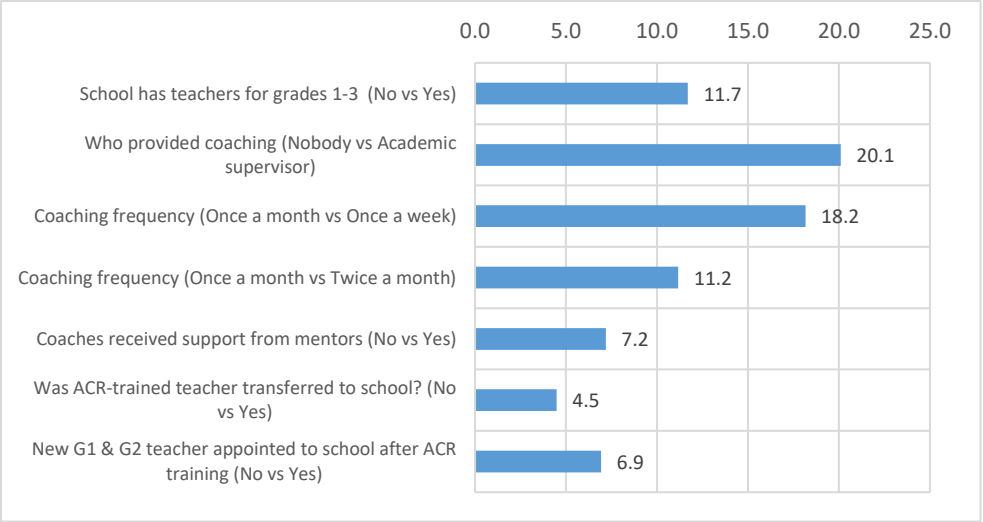
Note: Sig = Significance; D = Cohen's D

Factors that are negatively associated with reading performance are: *school closures for reasons other than holidays*; *grade 2 or grade 3 students studying more than four hours a day at school*; *students not taking breaks during the school day*; and *having three or less grade 2 classes and grade 2 teachers*.

Training provided by Afghan Children Read

Principals' responses regarding training provided by *Afghan Children Read* significantly associated with reading performance in schools are presented in Table 31. The variables that are positively associated with reading performance are: *the school having teachers for each early grade class (grades 1-3)*; *academic supervisors providing coaching support at the school*; *frequency of coaching (particularly once a week and twice a month)*; *coaches receiving support from mentors*; *transfer to the school of any grade 1 of grade 2 teacher trained by ACR*; and the *appointment to the school of any new grade 1 or grade 2 teachers after the ACR trainings*.

Table 31: Training Provided by *Afghan Children Read* and its Relationship to ORF



Sig	D
0.00	0.37
0.00	0.61
0.03	0.67
0.10	0.41
0.00	0.22
0.00	0.14
0.02	0.21

Note: Sig = Significance; D = Cohen's D

Appendix 4: Comparison of Baseline and Midline Factors Associated with Reading Achievement

This section highlights the factors associated with reading achievement that showed significant associations in both the baseline and midline studies. Table 42 through Table 37 below show the statistical significance ($p<0.05$ indicated in yellow cells) and the practical significance (Cohen's D greater than 0.35 for teacher and principal data, and greater than 0.25 for student data indicated in pink cells) for Dari language for the baseline and the midline. Empty cells indicate that the factor showed neither statistical nor practical significance.

Comparison of Analyses of Student Contextual Variables

Student characteristics that are significant in both the baseline and midline include: *students like to go to school to have a better future* and *student reporting that they have feel happy.* One factor that was significant in the midline but not in the baseline is *students reporting they enjoy reading.*

Table 32: Student Characteristics and their Relationship to ORF

Student characteristics	Baseline		Midline	
	Sig	D	Sig	D
Likes school: have a better future	0	0.3	0	0.36
Enjoys reading			0.01	0.85
Has laughed/been happy	0	0.39	0.06	0.32

Note: Sig = Significance; D = Cohen's D

Regarding actions by teachers, factors significantly associated with reading performance in both the baseline and midline include: *teachers actions when students learn their lessons well (showing no reaction and admiring students)* and *teachers actions when student cannot answer a question in class (asking the question again* and *asking another student if the answer is wrong).* One factor that did not show significant association with reading performance in the baseline but did so in the midline is *teachers forcing students to stand in class when they cannot answer a question correctly*. Two factors that were significant in the baseline but not in the midline include *students having a homework notebook* and *teacher checking students' homework.*

Table 33: Actions by Teachers and their Relationship to ORF

Actions by teachers	Baseline		Midline	
	Sig.	D	Sig	D
Has homework notebook	0	0.46		
Checks homework	0	0.64		
Lessons well learned teacher: shows no reaction	0	-0.42	0	-0.62
Lessons well learned teacher: admires me	0	0.3	0	0.31
When can't answer question teacher: asks the question in another way	0	0.38	0	0.47
When can't answer question teacher: asks another student	0.03	0.23	0.01	0.48
When can't answer question teacher: forces to stand in class			0.06	-0.35

Note: Sig = Significance; D = Cohen's D

Regarding student absenteeism, *being absent from school the week before the interview was conducted (was absent last week)* are significantly associated with reading performance in both the baseline and midline, albeit in opposite direction. The factor *being absent because the student had to work at home* has a negative association with reading performance in both the baseline and midline.

Table 34: Student Absenteeism and its Relationship to ORF

Student absenteeism	Baseline		Midline	
	Sig.	D	Sig	D
Was absent last week	0	-0.14	0.01	0.14
Was absent because had to work at home	0	-0.34	0.03	-0.23

Note: Sig = Significance; D = Cohen's D

Regarding student home support, four factors related to the student obtaining good marks were significant in both the baseline and the midline: *family knows when student gets good marks*, *family showing no reaction*, *family hugging and kissing the student*, and *family giving a gift to the student*.

Table 35: Student Home Support and its Relationship to ORF

Student home support	Baseline		Midline	
	Sig.	D	Sig	D
Family aware of good marks	0	0.37	0	0.65
Good marks: They showed no reaction	0	-0.55	0	-0.52

Student home support	Baseline		Midline	
	Sig.	D	Sig	D
Good marks: They hugged and kissed me	0	0.32	0	0.44
Good marks: They gave me a gift	0	0.39	0	0.41

Note: Sig = Significance; D = Cohen's D

In terms of student home environment, factors like *having books other than the textbook at home*, *eating before attending school*, and *having electricity at home*, are all significantly associated with reading performance both in the baseline and in the midline.

Table 36: Student Home Environment and its Relationship to ORF

Student home environment	Baseline		Midline	
	Sig.	D	Sig	D
Has other books at home	0	0.25	0	0.36
Eaten before attending school	0	0.18	0	0.28
Has electricity	0	0.33	0	0.36

Note: Sig = Significance; D = Cohen's D

For school environment, the midline showed two factors significantly associated with reading performance. These factors (*being afraid of kidnapping* and *being afraid of thieves*) are both related to students being afraid on the way to school.

Table 37: School Environment and its Relationship to ORF

School environment	Baseline		Midline	
	Dari		Dari	
	Sig.	D	Sig	D
On way to school: Is afraid of kidnapping			0	0.35
On way to school: Is afraid of thieves/dacoits			0.04	-0.23

Note: Sig = Significance; D = Cohen's D

Comparison of Analyses of Teachers' Contextual Variables

For teacher characteristics, *being a male teacher, having education as field of study,* and *having more than 6 years of experience* were significant factors in both the baseline and midline. On the other hand, two factors that were significantly associated with reading performance on the midline but not on the baseline are: *teachers having a bachelor's degree,* and *teachers reporting they like teaching.*

Table 38: Teacher Characteristics and their Relationship to ORF

Teacher characteristics	Baseline		Midline	
	Sig	D	Sig	D
Gender (Is male)	0.01	-0.29	0	-0.32
Has bachelor's degree			0	0.22
Education is field of study	0.06	0.57	0	0.23
Worked for more than 6 years	0.02	0.27	0	0.29
Likes teaching			0	0.32

Note: Sig = Significance; D = Cohen's D

For teacher activities, factors that showed significant associations with reading performance in both the baseline and midline include: *teachers reporting that they do not need help when facing problems, teachers reporting that they seek advice from deputy principal when facing problems,* and *teachers reporting that they do not use assessment results to improve teaching.*

Table 39: Teacher Activities and their Relationship to ORF

Teacher activities	Baseline		Midline	
	Sig	D	Sig	D
Seeks advice: Doesn't need help	0.1	-0.56	0.02	-0.41
Seeks advice: Asks deputy principal for help	0.05	0.27	0.01	0.16
Use of results: Doesn't use them	0.06	-0.48	0.03	-0.27

Note: Sig = Significance; D = Cohen's D

For activities by teacher supervisors, three factors were significant in the midline but not in the baseline: *principals checking the teaching process at least twice a month* and *receiving advice from supervisors on preparation of teaching plans.* The one activity that was significant in the baseline but not in the midline is *department head checking the teaching plan at least twice a month.*

Table 40: Activities by Teachers' Supervisors and their Relationship to ORF

Activities by teachers' supervisors	Baseline		Midline	
	Sig	D	Sig	D
Principal checks teaching process twice a month or more			0	0.32
Department head checks teaching plan twice a month or more	0.01	0.37		
Advice received: no advice received			0	0.34
Advice received: prep of teaching plans			0	-0.35

Note: Sig = Significance; D = Cohen's D

Regarding school environment, three factors that did not show significant association with reading performance in the baseline did show significant associations in the midline, these are: *teachers reporting they do not possess the teacher guidebook*, *teachers expecting students to read a text fluently in Grade 4 and above,* and finally *and teachers expecting students to write a text smoothly in Grade 4 and above.*

One factors that was significant in the baseline but not in the midline is: *teachers reporting they lack teaching aids*. Additionally, reporting that the *class size is up to 40 students* was significant in both the baseline and midline.

Table 41: School Environment and its Relationship to ORF

School environment	Baseline		Midline	
	Sig	D	Sig	D
G3 class has up to 40 students (vs 50 or more)	0	-0.67	0	-0.31
Teaching aids availability: no aids items in the class	0.01	-0.47		
Teacher guidebook problems: does not have it			0	-0.38
Students read fluently (G4 and above vs G3 or below)			0.01	0.27
Students write smoothly (5 and above vs G4 or below)			0	0.44

Note: Sig = Significance; D = Cohen's D

Comparison of Analyses of Principals' Contextual Variables

For principal characteristics, factors that showed significant associations with reading performance in both the baseline and midline include *having a bachelor's degree*; *having a degree in education*; and *gender*

(being a male principal), which in both the baseline and midline showed a negative association. *Years of experience (2 or more)* was a significant factor in the baseline but not in the midline.

Table 42: Principal Characteristics and their Relationship to ORF

Principal Characteristics	Baseline		Midline	
	Sig	D	Sig	D
Has 2 or more years of service	0.01	0.46		
Has a degree in Education	0.22	0.43	0.02	0.15
Has a bachelor's degree or higher	0	0.62	0	0.17
Gender (is male)	0	-0.64	0	-0.28

Note: Sig = Significance; D = Cohen's D

For principal activities, factors that are significant in the midline but not in the baseline include *checking annual exam results*, and *monitoring assessment results* and *monitoring the class*, both related to ensure that students are learning their lessons.

Table 43: Principal Activities and their Relationship to ORF

Principal Activities	Baseline		Midline	
	Sig	D	Sig	D
Ensure learning: check annual exam results			0	0.51
Ensure learning: monitor assessment results			0	0.24

Note: Sig = Significance; D = Cohen's D

For activities by principals' supervisors, the midline shows two significant factors that did not occur in the baseline: *inspectors providing advice on using the guidebook*; and *distribution of books by the MoE*.

Table 44: Activities by Principals' Supervisors and their Relationship to ORF

Supervisor Activities	Baseline		Midline	
	Sig	D	Sig	D
Inspectors provide advice on using guidebook			0	0.43
MoE distributed books			0	0.22

Note: Sig = Significance; D = Cohen's D

For school environment, factors that showed significant associations with reading performance in both the baseline and midline include: *students receiving training on mines and unexploded devices*, *schools having more than 3 grade 2 teachers/sections, and principals' opinion that students in grade 4 and above can read and write simple texts smoothly*. The midline shows five significant factors that were not significant in the baseline: *school having a library*, *students of grades 2-5 using the school library*; *having a development plan at school*; *principals feeling safe at school, and school closing for reasons other than a holiday*.

Table 45: School Environment and its Relationship to ORF

School environment	Baseline		Midline	
	Sig	D	Sig	D
Students received training on mines and unexploded devices	0.03	0.58	0	0.28
Feels safe at school			0	0.39
Grade 2 students use the library			0	0.35
Grade 3 students use the library			0.01	0.15
Grade 5 students use the library			0.02	0.17
School has library			0	0.17
Warm area: has more than 3 Grade 2 teachers	0	0.82	0	-0.37
Warm area: has more than 3 Grade 2 sections	0	0.78	0	-0.35
School was closed for reasons other than holidays			0.03	-0.54
School has discipline regulations	0.11	0.39	0	0.56
School has a development plan			0	0.27
Grade 4 and above can write simple texts smoothly	0.05	-0.51	0	0.36
Grade 4 and above can read simple texts fluently	0	-0.63	0	0.43

Note: Sig = Significance; D = Cohen's D