

EARLY GRADE READING ASSESSMENT, ENDLINE REPORT

USAID READ WITH ME PROJECT



Photo credit: USAID Read With Me

July 2021

This publication was produced for review by the United States Agency for International Development. It was prepared by School-to-School International and Chemonics International Inc.

Early Grade Reading Assessment, Endline Report

USAID READ WITH ME PROJECT

Contract No. AID-176-C-16-00003

DISCLAIMER

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States government.

ACKNOWLEDGMENTS

Many organizations and individuals contributed to the successful completion of the Read with Me Project (RWM) Endline Early Grade Reading Assessment (EGRA). The following organizations in Tajikistan provided insight, expertise, and guidance throughout the assessment process, from the review of assessment tools to data collection monitoring—the Ministry of Education and Science (MoES); National Testing Center; Academy of Education; Education Development Institute under the Academy of Education; Republican Teaching and Methodological Center; Republican Teachers’ Training Institute and the five regional Teacher Training Institutes; and the district and city education departments. We would like to extend special gratitude to Mr. Imomzoda MuhammadyusufSaidali, Minister of Education and Science; Mr. Salomiyon Muhammaddovud Qayum, First Deputy Minister of Education and Science; Ms. Ghanizoda Gulchehra Sami, Deputy Minister of Education and Science; Ms. Latofat Naziri Qahor, Advisor to Head of Inspection and Control Department under the President’s Office; Ms. Karimova Irina Kholovna, Deputy President of Academy of Education; Mr. Qurbonzoda Khonali, Head of Republican In-Service Teacher Training Institute; Mr. Yormuhamzoda Shermahmad Rajabali, Head of Republican Teaching and Methodological Center; Ms. Imatova Lyutfiya, Head of Education Development Institute; Mr. Khurshed Ikromi, Director of National Testing Center; and Mr. Sabzali Jafarov; Deputy Director of National Testing Center and Ms. Gulchehra Ghanizoda, Deputy Minister of Education and Science, for their guidance and support, as well as to Mr. Abdujabbor Aliev, RWM Coordinator from MoES, for his assistance throughout the assessment.

This report would not have been possible without the support of the American people through the United States Agency for International Development (USAID) RWM project. USAID provided critical input on the EGRA process through Andrew Colburn, USAID Regional Education Advisor; Paul Richardson, USAID Tajikistan Country Office Director; Mavjuda Nabieva, USAID Tajikistan Contracting Officer’s Representative; Angelique R. Mahal, Acting Deputy Mission Director; Inna Kirilyuk, USAID Almaty Youth Development Specialist; Mitch Kirby, USAID Asia Senior Education Advisor; Rebeca Martinez, Reading and Instructional Program Advisor; and Melissa Chiappetta, Senior Research and Learning Advisor.

The EGRA was implemented with the technical assistance of School-to-School International (STS) and Chemonics International. The assessment was planned and administered with the support of staff members in the United States and in the USAID RWM project office in Tajikistan. We appreciate RWM project staff’s collaboration and support. We also appreciate the hard work of many other individuals at STS, for contributions to tool development, programming, data collection training, data analysis, and this EGRA endline report.

Finally, we thank the 64 data collectors, whose efforts to reach schools in every region of Tajikistan made this report possible, as well as the school directors, teachers, and students who welcomed data collectors to their schools and participated in this study.

CONTENTS

Acknowledgments	3
Contents	4
Acronyms	5
List of Tables	6
List of Figures	7
Executive Summary	9
Study Purpose	9
Project Background	9
Design, Methods, and Limitations	9
Summary of findings	10
Conclusions and Recommendations	15
Introduction	17
Read With Me Project Background	17
Early Grade Reading Outcomes in Tajikistan	18
Purpose of the Study	19
Research Questions	20
Methods and Limitations	21
Tool Development	21
Endline Sampling	22
Endline Data Collection	24
Analytic Methods	25
Limitations	27
Results	29
Recommendations	88
Annexes	90
Annex A: Operational Data Collection	90
Annex B: Analytic Methods	92
Annex C: Reliability Measures	96
Annex D: Intraclass Correlations	106
Annex E: Detailed EGRA Tables	107
Annex F: Detailed Teacher SSME Tables	123
Annex G: Detailed Director SSME Tables	135
Annex H: Detailed School Inventory SSME Tables	138
Annex I: Detailed Classroom Observation SSME Tables	141

ACRONYMS

CWPM	Correct Words per Minute
DED	District Education Department
DRS	Districts of Republican Subordination
EGRA	Early Grade Reading Assessment
GBO	Gorno-Badakhshan Autonomous Oblast
ICC	Intraclass Correlation
IRB	Institutional Review Board
IRR	Inter-rater reliability
LOI	Language of Instruction
LTA	Learning Together Activity
MoES	Ministry of Education and Science
ORF	Oral Reading Fluency
PTA	Parent-teacher Association
QCO	Quality Control Officer
QLP	Quality Learning Project
QRP	Quality Reading Project
RWM	Read with Me
SE	Standard Error of the Mean
SES	Socioeconomic Status
SSME	Snapshot of School Management Effectiveness
STS	School-to-School International
USAID	United States Agency for International Development

LIST OF TABLES

Table 1. Change in Proportion of Students Meeting Benchmarks.....	11
Table 2. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Sex, between Baseline and Endline.....	11
Table 3. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Urbanicity, between Baseline and Endline	12
Table 4. Geographic Subgroups.....	13
Table 5. Associations between Teacher Behavior and Oral Reading Fluency by Grade and Language.....	13
Table 6. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language.....	14
Table 7. Endline EGRA Subtasks.....	22
Table 8. Midline SSME Tools.....	22
Table 9. Number of Students Assessed, Tajik Sample, 142 Schools	23
Table 10. Number of Students Assessed, Russian Sample, 60 Schools	23
Table 11. Timeline of RWM Interventions by Cohort	25
Table 12. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask, between Baseline and Endline	29
Table 13. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Sex, between Baseline and Endline.....	34
Table 14. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Urbanicity, between Baseline and Endline	37
Table 15. Trends in Tajik Grade 2 Student Reading Performance by Subtask and Region, between Baseline and Endline.....	41
Table 16. Trends in Tajik Grade 4 Student Reading Performance by Subtask and Region, between Baseline and Endline.....	41
Table 17. Trends in Russian Grade 2 Student Reading Performance by Subtask and Region, between Baseline and Endline	42
Table 18. Trends in Russian Grade 4 Student Reading Performance by Subtask and Region, between Baseline and Endline	42
Table 19. Subgroup performance for Oral Reading Fluency – Tajik Grade 2.....	46
Table 20. Subgroup performance for Oral Reading Fluency – Tajik Grade 4.....	47
Table 21. Subgroup performance for Oral Reading Fluency – Russian Grade 2.....	47
Table 22. Subgroup performance for Oral Reading Fluency – Russian Grade 4.....	48
Table 23. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language.....	49
Table 24. Associations between Total Evaluation Methods and Uses of Assessment Results and ORF by Grade and Language.....	53
Table 25. Associations between RWM Materials or Observed Teaching Practices and Oral Reading Fluency by Grade and Language.....	54
Table 26. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language.....	65
Table 27. Student and Teacher Resources Indexes, Means by Language and Grade Level	66
Table 28. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language.....	68
Table 29. Associations between Student Variables and Oral Reading Fluency by Grade and Language	71
Table 30. Associations between Teacher Variables and Oral Reading Fluency by Grade and Language	75
Table 31. Associations between Director Variables and Oral Reading Fluency by Grade and Language.....	77
Table 32. Associations between School Inventory Variables and Oral Reading Fluency by Grade and Language.....	78

LIST OF FIGURES

Figure 1. Trends in Tajik Grade 2 Reading Performance by Time Point: Fluency Tasks	30
Figure 2. Trends in Tajik Grade 2 Reading Performance by Time Point: Non-Fluency Tasks.....	30
Figure 3. Trends in Tajik Grade 4 Reading Performance by Time Point: Fluency Tasks	31
Figure 4. Trends in Tajik Grade 4 Reading Performance by Time Point: Non-Fluency Tasks.....	31
Figure 5. Trends in Russian Grade 2 Reading Performance by Time Point: Fluency Tasks.....	32
Figure 6. Trends in Russian Grade 2 Reading Performance by Time Point: Non-Fluency Tasks.....	32
Figure 7. Trends in Russian Grade 4 Reading Performance by Time Point: Fluency Tasks.....	33
Figure 8. Trends in Russian Grade 4 Reading Performance by Time Point: Non-Fluency Tasks.....	33
Figure 9. Trends in Tajik Grade 2 Reading Performance by Time Point and Sex: Oral Reading Fluency	35
Figure 10. Trends in Tajik Grade 4 Reading Performance by Time Point and Sex: Oral Reading Fluency	36
Figure 11. Trends in Russian Grade 2 Reading Performance by Time Point and Sex: Oral Reading Fluency	36
Figure 12. Trends in Russian Grade 4 Reading Performance by Time Point and Sex: Oral Reading Fluency	37
Figure 13. Trends in Tajik Grade 2 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency .	38
Figure 14. Trends in Tajik Grade 4 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency .	39
Figure 15. Trends in Russian Grade 2 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency	39
Figure 16. Trends in Russian Grade 4 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency	40
Figure 17. Trends in Tajik Grade 2 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF).....	43
Figure 18. Trends in Tajik Grade 4 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF CWPM).....	44
Figure 19. Trends in Russian Grade 2 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF).....	44
Figure 20. Trends in Russian Grade 4 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF).....	45
Figure 21. Factors associated with group with lowest scores: Tajik grade 2.....	46
Figure 22. Factors associated with group with lowest scores: Tajik grade 4.....	47
Figure 23. Factors associated with group with lowest scores: Russian grade 2	48
Figure 24. Factors associated with group with lowest scores: Russian grade 4	49
Figure 25. Teachers Preferred Evaluation Methods, by Language and Grade Level	50
Figure 26. Evaluation Methods used by Teachers (%), by Language and Grade Level.....	51
Figure 27. Use of Assessment Results, by Language and Grade Level	51
Figure 28. Total Uses of Assessment Results, by Language and Grade Level.....	52
Figure 29. Change in Use of Assessment Results Across Time, Tajik.....	52
Figure 30. Evolution in Use of Assessment Results Across Time, Russian	53
Figure 31. Overall Good Teaching Practices, by Language and Grade Level	55
Figure 32. Teachers’ Prereading Activities, by Language and Grade Level.....	55
Figure 33. Teachers’ While-Reading Activities – Part I, by Language and Grade Level	56
Figure 34. Teachers’ While-Reading Activities – Part II, by Language and Grade Level	57
Figure 35. Reading Skills Developed, by Language and Grade Level	57
Figure 36. Teachers’ Post-Reading Activities, by Language and Grade Level.....	59
Figure 37. Teacher Strategies Used, by Language and Grade Level	60
Figure 38. Percentage of Tajik Grade 2 Teachers Using Specific Teaching Techniques, Midline and Endline.....	61
Figure 39. Percentage of Tajik Grade 2 Teachers Focusing on Specific Reading Skills, Midline and Endline	61
Figure 40. Percentage of Tajik Grade 4 Teachers Using Specific Teaching Techniques, Midline and Endline.....	62
Figure 41. Percentage of Tajik Grade 4 Teachers Focusing on Specific Reading Skills, Midline and Endline	62
Figure 42. Percentage of Russian Grade 2 Teachers Using Specific Teaching Techniques, Midline and Endline .	63
Figure 43. Percentage of Russian Grade 2 Teachers Focusing on Specific Reading Skills, Midline and Endline...	63
Figure 44. Percentage of Russian Grade 4 Teachers Using Specific Teaching Techniques, Midline and Endline .	64
Figure 45. Percentage of Russian Grade 4 Teachers Focusing on Specific Reading Skills, Midline and Endline...	64
Figure 46. Books Other than Textbooks, by Language and Grade Level	65
Figure 47. Books Provided by the Project are Available, by Language and Grade Level	66

Figure 48. Percentage of Students with Textbooks, Language Exercise and Vocabulary Exercise Books, by Language and Grade Level.....	67
Figure 49. Materials Available to Teachers, by Language and Grade Level.....	67
Figure 50. Teacher with Developed Lesson Plan, by Language and Grade Level	68
Figure 51. Other RWM Efforts, by Language and Grade Level	69
Figure 52. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Tajik Grade 279	
Figure 53. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Tajik Grade 2	80
Figure 54. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Tajik Grade 480	
Figure 55. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Tajik Grade 4	81
Figure 56. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Russian Grade 2	82
Figure 57. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Russian Grade 2.....	82
Figure 58. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Russian Grade 4	83
Figure 59. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Russian Grade 4.....	83

EXECUTIVE SUMMARY

STUDY PURPOSE

This is the endline final report of a five-year study that seeks to understand trends in early grade reading performance in Tajikistan at the national level and among schools providing Tajik and Russian language of instruction and benefiting from the United Agency for International Development (USAID) Read with Me (RWM) project. Findings from this study aim to learn from the implementation of RWM and support the Ministry of Education and Science (MoES) improve early grade reading in Tajikistan.

PROJECT BACKGROUND

RWM is a five-year (2016–2021) project implemented by Chemonics International that aims to improve reading outcomes for 75 percent of all primary grade students nationwide over the course of implementation. It is designed to achieve this goal through improved reading instruction in grades 1 through 4, improved access to and availability of quality materials that support the development of reading in grades 1 through 4, increased innovations and partnerships supporting literacy outcomes, and increased government support to improve reading. RWM accomplishes these goals both by working with the MoES at all levels in the system and schools to support the improvement of reading outcomes.

RWM is the latest USAID-funded project intervening in primary education in Tajikistan. It builds on the gains made most recently by two previous education projects—the 2007–12 Quality Learning Project (QLP) and the 2013–17 Quality Reading Project (QRP)—and is followed by the 2020-2025 Learn Together Activity.

DESIGN, METHODS, AND LIMITATIONS

STUDY DESIGN AND PURPOSE

This is a five-year study with data collection taking place in spring 2018 as a baseline, spring 2019 as a midline, and spring 2021 as an endline.¹

The purpose of this report is twofold, building on earlier results:

1. To provide a nationally representative “snapshot” of reading performance in Tajikistan over time (from baseline to endline); and
2. To provide RWM, the Learn Together Activity, and the MoES and district education departments with further insights to further align interventions to meet the needs of the schools and students.

METHODS

RWM drew a representative sample of schools from the project’s second of four cohorts.² The sample represents the full population of schools in Tajikistan based on region, language of instruction (Russian and Tajik), and location (urban and rural). Students were sampled at the class level by sex.

¹ This five-year study originally set out to hold baseline data collection in spring 2017. Due to timelines external to the project, data collection for the baseline occurred in spring 2018. In addition, full implementation of the intervention was staggered throughout the 2018-2019 school year. Midline results from spring 2019 thus show results from less than a full year of project implementation. These endline results from spring 2021 show results from three years of project implementation in the second cohort of schools, which received training inputs starting in July 2018.

² More information can be found in the Methods section of this report.

Students were assessed using Tajik- and Russian-language adaptations of the Early Grade Reading Assessment (EGRA), which measures student performance on the basic foundational skills required for fluency in reading. Alongside the EGRA, students took a brief survey that collected information about their background and home environment. Additionally, school-level data was collected using Snapshot of School Management Effectiveness (SSME) tools, including a school director questionnaire, a teacher questionnaire, a school inventory checklist, a classroom inventory checklist, and a language (Tajik classes) or reading (Russian classes) lesson observation. The purpose of the student survey and SSME data was to understand student and school contextual factors that may influence students' reading performance.

Endline data collection took place in April and early May 2021. Data were collected by trained enumerators, many of whom were proposed by the MoES, who passed three assessor accuracy quizzes. In total, enumerators visited 202 schools—142 in the Tajik sample and 60 in the Russian sample, given the proportionate population sizes. They collected data from 1,342 Tajik grade 2 students, 1,348 Tajik grade 4 students, 569 Russian grade 2 students, and 599 Russian grade 4 students. Details of sampling and the census approach are found in Methods and Limitations.

The oral reading fluency (ORF) subtasks from the endline EGRA tools were equated with—or brought to a common scale as—the ORF subtasks used in this study's 2018 baseline data collection. Equating was done to allow for appropriate comparisons with data reported at baseline and midline. Sampling weights were calculated and applied to the assessment data to minimize bias in the results. RWM scored weighted and equated data using the following measures: mean scores, percentage correct scores, fluency scores, zero scores, and benchmark scores. Trends across data collection points and for each of these measures were analyzed using t-tests.³ Endline results were also disaggregated by groups of interest, and differences within each group were analyzed using t-tests.

Data from the student survey and SSME questionnaires, inventories, and observations were correlated with the ORF scores to identify student and school characteristics linked to the subtask scores. Those items with statistically significant correlations with the ORF subtask scores were included in linear regression models to understand the extent to which they predicted the students' reading performance.

SUMMARY OF FINDINGS

In this summary of findings from the report, note that when measuring the relationship between student performance and contextual factors such as teacher practices or home conditions, these contextual factors were examined in relation to ORF scores only. ORF scores were used for these correlations because they provide a robust picture of students' ability, with substantial range and variability in results, instead of comprehension scores, which are more limited and less reliable given that the comprehension subtask consisted of only five questions.

ORF and Comprehension

Oral reading fluency (ORF) consists of reading speed, accuracy, and prosody (appropriate intonation). Comprehension consists of understanding the meaning of what has been read. ORF has been shown to be predictive of reading comprehension (USAID: EGRA Toolkit 2.0, p. 18) and is sometimes used as a proxy for comprehension. However, in some instances, students can read quickly (high ORF score) but not understand what they are reading (low comprehension score). For this reason, separate measures of fluency (ORF) and comprehension are preferred because they provide a more accurate picture of each student's reading ability than ORF alone.

³ A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, whether independent or dependent. STS used this methodology to compare subgroups because they were suitable in a context where different subgroups had to be compared within or across timepoints, in absence of longitudinal data.

All four groups of students—Tajik grade 2 and 4 and Russian grade 2 and 4—showed statistically significant improvements from baseline to endline. All groups except for Russian grade 4 students made significant gains on at least five subtasks. Notably, significant gains were made from baseline to endline in reading comprehension by grade 2 students in Tajik and Russian, and on the silent reading comprehension subtask by grade 2 and grade 4 students in both languages, which was a rarely used instructional method prior to RWM interventions. Importantly, no group saw statistically significant declines on any subtask from baseline to endline. The proportion of students meeting the reading with comprehension benchmarks in each group increased from baseline to endline as follows:

- Tajik grade 2: 14.70 percent at baseline, 28.11 percent at endline
- Tajik grade 4: 22.50 percent at baseline, 28.52 percent at endline
- Russian grade 2: 16.20 percent benchmark at baseline, 41.25 percent at endline
- Russian grade 4: 41.40 percent benchmark at baseline, 41.07 percent at endline

The proportion of students meeting benchmarks increased significantly from baseline to endline in three of the four groups of students assessed. The following table indicates that the proportion of students performing at or above benchmarks significantly increased from baseline to endline in three groups—grade 2 Tajik students, grade 4 Tajik students, and grade 2 Russian students. The proportion of grade 4 Russian students meeting benchmarks remained statistically unchanged from baseline to endline, as shown in Table 1. The proportion of students meeting the ORF benchmark in each group increased from baseline to endline as follows:

- Tajik grade 2: 50.60 percent at baseline, 54.59 percent at endline
- Tajik grade 4: 18.80 percent at baseline, 41.45 percent at endline
- Russian grade 2: 48.70 percent benchmark at baseline, 54.59 percent at endline
- Russian grade 4: 38.60 percent benchmark at baseline, 41.74 percent at endline

Table 1. Change in Proportion of Students Meeting Benchmarks

Students meeting benchmarks: Endline relative to baseline				
EGRA Subtask	Tajik Grade 2	Tajik Grade 4	Russian Grade 2	Russian Grade 4
ORF (Grade 2: 40 CWPM, Grade 4: 80 CWPM)	B=50.60% E=54.59%	↑ B=18.80%, E=41.45%	B=48.70% E=54.59%	B=38.60% E=41.74%
Reading Comprehension: 80% score (4 of 5 questions)	↑ B=14.70%, E=28.11%	B=22.50% E=28.52%	↑ B=16.20%, E=41.25%	B=41.40% E=41.07%

Note: B and E indicates baseline and endline proportions of students meeting benchmarks. An up arrow (↑) indicates that the proportion of students at endline was statistically significantly higher than the proportion of students at baseline. For all comparisons, statistically significant differences are reported at the $p < 0.05$ level.

By endline, both girls and boys were making significant improvements across grades and language groups. Rates of improvement were comparable for Tajik boys and girls; both showed statistically significant gains on four to five subtasks at both grade levels from baseline to endline. Russian girls in grade 2 also showed statistically significant improvement on five subtasks from baseline to endline. Notably, significant gains were observed on the silent reading comprehension task for both girls and boys in Tajik and Russian at both grade levels. Fewer gains were achieved by grade 2 Russian boys and grade 2 and grade 4 Russian boys and girls.

Table 2. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Sex, between Baseline and Endline

EGRA Subtask	Endline relative to baseline							
	Tajik Grade 2		Tajik Grade 4		Russian Grade 2		Russian Grade 4	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls

Letter name identification (CLNPM)	↑	↔	N/A	N/A	↓	↔	N/A	N/A
Initial sound identification	↑	↑	N/A	N/A	↑	↔	N/A	N/A
Familiar word reading (CFWPM)	↔	↔	↑	↑	↔	↑	↔	↓
Nonword reading (CNWPM)	↑	↑	↑	↑	↔	↑	↔	↔
ORF (equated CWPM)	↔	↔	↑	↑	↔	↑	↔	↔
Reading comprehension	↑	↑	↔	↔	↑	↑	↔	↔
Silent reading comprehension	↑	↑	↑	↑	↑	↑	↑	↑
Listening comprehension	↔	↔	↑	↑	↔	↔	↑	↔

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a down arrow (↓) indicates that the mean score for students at endline was statistically significantly lower than the mean score for students at baseline; and a left-right arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the $p < 0.05$ level. Differences were computed by sex, language and grade level.

Tajik and Russian grade 2 rural students saw the greatest proportion of gains on EGRA subtasks over time, while urban students, especially Tajik grade 2 and Russian grade 4, saw the fewest gains. As the following table shows, gains were posted amongst all groups, and no statistical declines were found by urbanicity. Importantly, rural students saw more gains on EGRA subtasks than their urban counterparts in grade 2 and 4 Tajik classrooms and grade 2 Russian classrooms:

Table 3. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Urbanicity, between Baseline and Endline

EGRA Subtask	Endline relative to baseline							
	Tajik Grade 2		Tajik Grade 4		Russian Grade 2		Russian Grade 4	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Letter name identification (CLNPM)	↔	↑	N/A	N/A	↔	↑	N/A	N/A
Initial sound identification	↔	↑	N/A	N/A	↑	↑	N/A	N/A
Familiar word reading (CFWPM)	↔	↑	↔	↑	↔	↔	↔	↔
Nonword reading (CNWPM)	↔	↑	↔	↑	↑	↑	↔	↑
ORF (equated CWPM)	↔	↑	↑	↑	↑	↑	↔	↑
Reading comprehension	↔	↑	↔	↔	↑	↑	↔	↔
Silent reading comprehension	↑	↑	↑	↑	↑	↑	↑	↔
Listening comprehension	↔	↔	↑	↑	↔	↔	↑	↔

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a left-right arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the $p < 0.05$ level. Differences were computed by urbanicity, language and grade level.

Geographic subgroups varied in performance. Across regions, the proportion of student groups (disaggregated by sex, urbanicity, and region) meeting the ORF benchmark was lowest in DRS, followed by Sughd, as shown in the following table. All groups highlighted in pink or red require supplementary attention. Of these, Russian grade 2 girls in rural schools in DRS struggled most at 25.54 correct words per minute (CWPM). When examining the student groups struggling most in each language by grade level by language, three common themes were language spoken at home, homework, and having family members who cannot read:

- Grade 2 Tajik boys who came from Uzbek-speaking homes and received homework less frequently than their peers.
- Grade 4 Tajik boys who came from Uzbek-speaking homes, received homework less frequently than their peers, and had brothers who cannot read.

- Grade 2 Russian girls who did not have a father who can read, did not attend preschool, and had no reading books at school to take home.
- Grade 4 Russian boys who came from homes where Russian is not spoken.

Table 4. Geographic Subgroups

Region	Tajik Grade 2				Tajik Grade 4				Russian Grade 2				Russian Grade 4			
	Urban		Rural		Urban		Rural		Urban		Rural		Urban		Rural	
	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl
Khatlon-Bokhtar	51.55	49.51	41.38	42.66	83.68	86.54	71.7	72.55	35.17	37.53	n/a	n/a	69.14	73.1	n/a	n/a
Dushanbe	47.47	49.83	n/a	n/a	102.56	89.11	n/a	n/a	42.56	52.17	n/a	n/a	78.83	73.55	n/a	n/a
Khatlon-Kulob	46.88	39.67	47.42	44.15	68.17	76.27	66.34	70.52	41.9	31.02	n/a	n/a	196.26	84.73	n/a	n/a
DRS	35.25	45.16	36.69	47.18	67.22	68.06	72.5	72.22	34.77	31.45	30.21	25.54	59.7	65.8	75.2	62.81
Sughd	37.69	58.87	33.05	41.38	71.53	85.3	56.27	74.3	40.62	47.22	37.55	38.24	72.05	77.63	55.82	72.7
GBAO	43.75	35.38	47.26	57.76	69.44	108.25	76.09	82.71	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note: Highlighted cells show means that are below the ORF benchmark for the corresponding grade level and language. In addition, the stronger the color, the lower the score.

Among regions, Dushanbe saw the least improvement. Though nearly all student groups met the ORF benchmark in Dushanbe—the highest proportion of any region in this evaluation (see previous point)—students’ scores in Dushanbe nevertheless stagnated or declined more than any other region in both grades in Tajik and Russian grade 4. This may have been due to already-higher performance levels, making it more difficult to make gains the same size as those of groups starting at lower levels.

When teachers used instructional practices promoted by RWM, students made greater gains. For example, in both Russian grade 2 and Tajik grade 2 classes, teachers who were observed explicitly articulating the objectives of the lesson and relating classroom activities to those objectives were associated with ORF increases of nearly 6 and nearly 5 CWPM, respectively. Teachers who included more prereading, while-reading, and post-reading teaching strategies were associated with modest but statistically significant increases in CWPM in Tajik grade 2. Similar results were seen in Tajik grade 4. In addition, in classrooms where teachers a) rephrased and explained a question if a student was unable to answer it correctly or b) encouraged such students to try harder, ORF scores generally improved from baseline to endline. Conversely, in classrooms where teachers used negative discipline measures such as hitting students, ORF scores generally declined, demonstrating the need for SEL training for teachers.

Table 5. Associations between Teacher Behavior and Oral Reading Fluency by Grade and Language

Student Survey Variable	Associated increase/decrease in ORF score (CWPM)			
	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
If unable to answer a question - teacher puts mark			-2.64	-5.33
If unable to answer a question - teacher rephrases, explains	5.47	7.63	4.18	
If unable to answer a question - teacher encourages student to try harder	8.45	7.99	6.26	5.50
If unable to answer a question - teacher scolds student		-11.40	-5.84	
If unable to answer a question - teacher hits student	-4.70	-10.91	*	*
If unable to answer a question - teacher sends student to the corner of the classroom	-6.55		*	*

* Results are omitted due to extremely small number of affirmative responses

When teachers used selected assessment practices promoted by RWM, students made greater gains, yet many types of evaluation methods remain underutilized. For all but Russian grade 4, using different types of evaluation methods, as well as using assessment results for multiple purposes, tended to correlate positively with gains in reading ability over time. For example, for teachers of grade 4 Tajik students, measuring student progress using oral evaluations was associated with an ORF increase of 12.97 CWPM (see table). Furthermore, when teachers reported using more evaluation methods, students consistently performed better on ORF. Few teachers, however, reported using the full range of types of evaluation encouraged by RWM, opting for simpler and more traditional forms of evaluation, including oral evaluations. Teachers also reported infrequent use of end-of-term evaluations, written tests, portfolios, and few teachers reported using assessments to plan teaching activities or adapt teaching to better suit students' needs.

Table 6. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language

Associated increase/decrease in ORF score (CWPM)				
Materials or Instructional Practices	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
Assessment and evaluation practices	Teacher measures students' progress using written tests	8.83		8.32
	Teacher measures students' progress using oral evaluations		12.97	-9.11
	Teacher measures students' progress using portfolios and other projects	7.48		9.98
	Teacher measures students' progress using homework			7.23
	Teacher measures students' progress using end of term evaluations	6.01		
	Use results of students' oral and written assessments to grade students		8.47	5.74
	Use results of students' oral and written assessments to evaluate students' understanding of subject matter	4.72	12.64	7.93
	Use results of students' oral and written assessments to adapt teaching to better suit students' needs	4.65		10.22

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. In particular, green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

Experienced and trained teachers tended to be associated with improved ORF scores.

More experienced teachers were associated with improvements in performance as measured by the ORF subtask (CWPM) in every grade and language. Attaining higher levels of education, receiving support at school in the past year, and attending in-service training or professional development in the last year, especially in teaching reading, were also associated with improvements, especially in Tajik grade 2. Similarly, directors who reported that their teachers had received training on how to teach reading outside of in-service teacher training were associated with sometimes large improvements in every grade and language.

Parental involvement predicted better reading outcomes. In all but Tajik grade 4, students with parents who reviewed their homework were associated with improvements in ORF. In both Russian grades, teachers who reported they were satisfied with parental involvement in the classroom were also associated with improvements. In both Tajik grades, schools holding regular parent-teacher association (PTA) meetings were associated with improved student performance. In Russian grade 2, directors who were satisfied with the level of support the PTA provided to the school were also associated with improved student performance. In addition, having books at home was positively associated with ORF in all grades and languages.

The level of resources available was associated with improvements in ORF. The positive relationship between ORF scores and resources such as the appropriate textbook (Mother Tongue in Tajik classes and reading in Russian classes), language exercise books, vocabulary books, a teachers' board, chalk or markers, an interactive board, other visual aids, or a developed lesson plan or lesson summaries suggests that reading proficiency is resource-sensitive, indicating a path to further improvement in reading outcomes. Some resources, such as reading corners and the use of logbooks, were provided directly by RWM and were associated with improvements in ORF scores ranging from 5.99 to 11.59 CWPM.

CONCLUSIONS AND RECOMMENDATIONS

This EGRA endline evaluation found substantial evidence of progress made by RWM over the life of the project. From baseline to endline, all four groups of students—Tajik grade 2 and 4 and Russian grade 2 and 4—showed statistically significant improvements from baseline to endline, and for three of the four groups, the proportion of students meeting ORF or comprehension benchmarks increased during that time. By endline, girls and boys were performing comparably in most cases—effectively equalizing performance among Tajik students after girls had been performing better at midline. Importantly, boys and girls in grades 2 and 4 in Tajik and grade 2 in Russian improved on several tasks from baseline to endline, and all students, including those in Russian grade 4, improved significantly in silent reading comprehension. Factors associated with reading gains included:

- the use of a variety of assessment types,
- having reading books at home,
- having teachers who use positive discipline strategies such as encouraging students to try harder (punitive practices such as hitting the student were negatively correlated with achievement),
- having experienced and trained teachers,
- having more materials in schools and at home, and
- having parents who review students' homework and motivate reading and learning.

RWM's role in providing some of these types of training and materials suggests that the project played a significant role in improving student outcomes.

The following is a list of key considerations and recommendations.

Assessment practice correlated with performance. While gains were greater when students had teachers who used selected assessment methods, students' reading achievement was also stronger for all but Russian grade 4 when teachers reported using different methods, and were stronger for Tajik students in grade 2 and Russian students in grade 4 when teachers used results of students' oral and written assessments to adapt teaching to better suit their students' needs.

Recommendation 1: Continue to improve teachers' use of a range of formative assessment practices. Future efforts should encourage teachers to use a variety of evaluation methods, as well as strategies for using assessment results to adapt teaching to better suit their students' needs.

Professional development correlated with performance. This EGRA found a positive relationship between training activities like in-service training and mentoring, especially focused on reading, and improved student outcomes.

Recommendation 2: Identify strategies that could ensure ongoing professional development over the near to medium term—e.g., providing three to five days per year to teachers in reading strategies—in order to sustain and build on the gains realized in RWM.

Classroom management practices correlated with performance. Though the focus of this EGRA was not on classroom management, this evaluation found evidence that positive discipline strategies like encouraging students were associated with better learning outcomes, while negative strategies such as hitting had the reverse effect.

Recommendation 3: Expand teachers' use and appreciation of positive discipline strategies and identify ways to monitor and correct cases of more punitive approaches.

Support at-risk student groups with targeted interventions. While high proportions of student groups (disaggregated by region, sex, urbanicity, grade, and language) were struggling in DRS and Sughd and, therefore, require supplementary attention, particular attention should be paid to the lowest-scoring student groups.

Recommendation 4: Support the most at-risk student groups as follows:

- **Grade 2 and 4 Tajik boys:** Encourage assigning homework more frequently and providing Tajik linguistic support for students who speak Uzbek at home.
- **Grade 2 Russian girls:** Encourage providing reading books at school to take home, promoting preschool attendance, and providing Tajik linguistic support for girls who speak Uzbek at home.
- **Grade 4 Russian, both girls and boys:** Encourage providing Russian language support tailored to the needs of linguistically diverse groups or linguistic minorities.

Since each of the four groups is characterized by a gap between languages spoken at home and in the classroom, provide linguistic support for these students by encouraging teachers to find out which languages students use most often and their level of fluency in a second language, and identify ways to bring students' language and culture into the classroom, especially in the early grades.⁴

Conditions at home and in the classroom matter. This evaluation found that some students showed greater improvement when there were books at home, when parents reviewed their homework, and when parents were involved in the school. It also found positive correlations between performance and the existence of materials in schools such as the appropriate textbooks, language exercise books, and vocabulary books.

Recommendation 5: Improve the provision of resources at school and in the home.

The selection of these resources should be made in reference to the classroom index, school environment index, and lists of reading materials at home identified in this evaluation.

⁴ For additional explanation and strategies, see Save the Children/UK (2009) Steps Toward Learning: A guide to overcoming language barriers in children's education. <https://resourcecentre.savethechildren.net/library/steps-towards-learning-guide-overcoming-language-barriers-childrens-education>

INTRODUCTION

READ WITH ME PROJECT BACKGROUND

The Government of the Republic of Tajikistan has implemented reforms to improve the quality of its education system since independence and the country's five-year civil war in the 1990s.⁵ The Government of Tajikistan and the Ministry of Education and Science (MoES) view inclusive access to quality education as a requirement of the country's transition to a market economy and are committed to reforming the education sector as outlined in its National Strategy for Education Development (2012–2020) and Midterm Development Strategy (2016–2020).⁶ Investment in education—particularly early childhood education and the inclusion of girls, rural students, and students with disabilities—supports building the capacity of Tajikistan's workforce and contributing to the economic development of the country.⁷ Although Tajikistan's education budget has steadily increased as a percentage of the country's gross domestic product, the level of financing is inadequate to meet system needs, according to a funding gap analysis conducted by the MoES.⁸

Aid agencies and international organizations in Tajikistan help to fill this gap through initiatives like early childhood education, inclusive education, quality improvements in education, and child-friendly classrooms, as well as other educational programming. The United States Agency for International Development (USAID) has provided support to the government's reforms to primary education for years through efforts such as the 2007–2012 Quality Learning Project (QLP) and the 2013–2017 Quality Reading Project (QRP).⁹ USAID has been supporting the Read with Me project (RWM) to build on the gains made by these two projects in reading skills, including comprehension, and RWM is followed by the 2020–2025 Learn Together Activity.

RWM is a five-year (2016–2021) project that builds on both QLP and QRP, as well as other education programs in Tajikistan; it supports the National Strategy for Education Development 2012–2020 and the Midterm Development Strategy for 2016–2020. USAID's RWM project involves both the MoES and individual schools to support the improvement of reading outcomes, as well as reaches out to communities and the private sector. It looks to introduce the role of science, technology, innovations, and partnership in supporting reading outcomes. The RWM project provides support directly to schools in a series of four cohorts from 2018 to 2021.

RWM aims to improve reading outcomes for students in grades 1 to 4 in targeted schools in Tajikistan by:

- a. Increasing the availability of age-appropriate reading materials in Tajik and Russian languages, which will develop students' core reading skills;
- b. Providing educators with reading-specific in-service training, supplementary materials, and the integration of assessments to monitor reading progress;
- c. Increasing innovations and partnerships supporting literacy outcomes; and,

⁵ Government of the Republic of Tajikistan, *National Strategy for Education Development for 2012-2020*. (2012), 3–4. http://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/tajikistan_ed_sector_plan_2012-2020.pdf and United Nations in Tajikistan. *Annual UNDAF Results Report*. (2016).

https://untj.org/files/Publications/UNDAF/UNDAF_Annual_Report_English_Version_2016.pdf

⁶ Ibid.

⁷ Government of the Republic of Tajikistan, *National Strategy for Education Development for 2012-2020*. (2012), 3, 9–12.

⁸ The World Bank Group, *Tajikistan Partnership Program Snapshot* (October 2015), 5.

<http://pubdocs.worldbank.org/en/645741444794465533/Tajikistan-Snapshot.pdf>

⁹ RWM expands on QRP's work, which reached over 60 percent of schools in Tajikistan. QRP supported reading instruction in grades 1 to 4 through in-service training and classroom-based mentoring for educators, as well as Tajik and Russian-language reading materials and other school-based reading activities and governmental policy support to reading. See American Institutes for Research, *USAID Quality Reading Project Republic of Tajikistan: Final EGRA and Impact Report, 2013–2017*, (2017).

- d. Increasing government support for reading through capacity building in reading instruction and mentoring, as well as increasing dialogue on the reading environment.¹⁰

RWM has supported 3,001 schools, 28,415 teachers, and 665,830 students across the country in Tajik and Russian classrooms over the course of implementation.. RWM has been active in every region of Tajikistan and expands USAID’s support of early grade reading education to the Gorno-Badakhshan Autonomous Oblast (GBAO).

EARLY GRADE READING OUTCOMES IN TAJIKISTAN

The RWM Early Grade Reading Assessment (EGRA) builds on a foundation of research in early grade reading outcomes in Tajikistan. Previous USAID-funded projects have administered multiple EGRAs, which provide the most extensive research into early grade reading outcomes in Tajikistan. In 2011, the first USAID-funded EGRA in Tajikistan analyzed results from students in Tajik and Russian in grades 2 to 4 and included a qualitative study of early grade reading pedagogy in Tajikistan.¹¹ This study, implemented in the last year of QLP, found that students generally had strong foundations in early grade reading skills, including the alphabetic principle and recognition of sight words; however, they had a weakness in decoding unfamiliar words and comprehending texts. The report suggested links between classroom practices and these outcomes, as pedagogy at that time focused more on rote memorization of words and speed reading rather than on phoneme segmentation, decoding skills, and comprehension strategies.¹²

QRP’s interventions were designed to address these areas for improvement by drawing on predictors of reading success identified by the EGRA, such as the promotion of a “culture of literacy” with parents and the publication of grade-level texts. In addition to program interventions, QRP collected EGRA data with a baseline assessment in 2014, a midline assessment in 2016, and an endline assessment in 2017 as part of a randomized control trial in primary grades in all regions of the country except for GBAO. Early QRP results were consistent with the 2011 EGRA report.¹³

QRP reported overall trends of growth in Tajik- and Russian-reading outcomes from 2014 to 2016, with declines in 2017 that reverted to 2014 levels. Overall, the QRP EGRA data suggest that by 2017, students in Tajikistan generally demonstrated foundational reading skills. Phoneme segmentation and dictation appear to have improved from 2011. The rate of zero scores—or students who were unable to answer at least one item correctly on a subtask—was low across all grades, languages, and subtasks. Overall, the oral reading fluency (ORF) subtask scores fluctuated over the course of QRP, with less than half of students in Tajik- or Russian-language schools attaining proficiency scores by 2017. The reading comprehension subtask scores, while showing gains among students in grade 4 who benefitted from QRP interventions, remained fairly low; a large number of students performed poorly.¹⁴

¹⁰ Chemonics International, Inc, *USAID Read With Me Project Annual Report Year 1*. (2017).

https://pdf.usaid.gov/pdf_docs/PA00N7RF.pdf

¹¹ Mirka Tvaruzkova and Duishon Shamatov, *Review of Early Grade Reading Teaching and Skills: The Kyrgyz Republic and Tajikistan Final Report*. (Washington, DC: United States Agency for International Development, 2012).

¹² The interviews focused on current pedagogical practices and materials used in early grade reading instruction, as well as factors that affect early grade reading acquisition. Interview guides were prepared and used in a semi-structured format with stakeholder groups. A total of 25 interviews were conducted in Tajikistan. The sample included education officials from the MoES, pre-service teacher training institution staff, Academy of Education members, authors of children’s literature, representatives of international organizations, primary school teachers, and parents.

¹³ Mirka Tvaruzkova and Duishon Shamatov, *Review of Early Grade Reading Teaching and Skills: The Kyrgyz Republic and Tajikistan Final Report*. (Washington, DC: United States Agency for International Development, 2012).

¹⁴ American Institutes for Research, *USAID Quality Reading Project Republic of Tajikistan: Final EGRA and Impact Report, 2013–2017*. (2017).

RWM completed a baseline EGRA in 2018 that built on previous QRP results, followed by a midline EGRA in 2019. As with previous assessments, RWM’s EGRA results for students studying in Tajik and Russian showed strong reading outcomes overall. The ORF subtask mean scores for grade 2 students in Tajik-language schools and grade 4 students in Russian-language schools approached the established benchmarks for their respective grades—40 correct words per minute (CWPM) and 80 percent comprehension for grade 2, and 80 CWPM and 80 percent comprehension for grade 4. The minimum acceptable standard on the ORF subtask was based on standards set by the MoES.¹⁵ The minimum acceptable standard for the reading comprehension subtask is 80 percent—four out of five questions—a standard that is used most often globally in association with the EGRA reading comprehension subtask.¹⁶ Regardless of language or grade, the percentage of students with zero scores across most EGRA subtasks was extremely low—less than five percent for most subtasks and approximately zero for many of the basic reading skills—indicating that nearly all students have some ability to read.

The RWM baseline EGRA established patterns of performance for subgroups that continued with the RWM midline. Girls generally outperformed boys, especially on timed subtasks such as letter name identification, familiar word reading, nonword reading, and ORF. Students from urban areas generally outperformed students from rural areas, especially in Tajik-language schools.

As the first EGRAs in Tajikistan designed to compare performance by region, the RWM baseline and midline yielded valuable region-level results. At both time points, students from schools located in Dushanbe and GBAO generally outperformed students from the other regions in the country—the Districts of Republican Subordination (DRS), Khatlon-Bokhtar, Khatlon-Kulob, and Sughd. In Tajik-language schools, students in Dushanbe and GBAO typically had the highest mean scores relative to students in other regions. In Russian-language schools, students in Dushanbe typically had the highest mean scores relative to other regions (there are no Russian Medium Schools in GBAO). These results were consistent across both grade 2 and grade 4.

The RWM baseline and midline EGRAs also examined predictors of student outcomes. At the student level, several factors had positive associations with reading outcomes, including parents’ ability to read in the language of instruction (LOI), the availability of books to read at home, the ability to take non-textbooks home from school for reading for pleasure, and preschool attendance. At the school level, several factors had positive associations with reading outcomes, including class size, teacher professional development on reading instruction, and the availability of student resources such as reading or mother language textbooks, reading workbooks, and writing utensils. That is, while controlling for other factors, students with these characteristics tended to have higher learning outcomes.

PURPOSE OF THE STUDY

This report examines reading trends at the national level among RWM schools. The study’s purposes were defined in collaboration with USAID and in consultation with the MoES. First, the report provides nationally representative results over time. This examination of change from the RWM baseline to endline provides detailed results for subgroups, including sex, urbanicity, and region. Second, the report seeks to provide insights to RWM and MoES to ensure the project meets the needs of the schools and students.

¹⁵ Ministry of Education and Science of the Republic of Tajikistan, *Primary Education Subject Standards*, (Dushanbe: Maorif, 2017) and American Institutes for Research. *USAID Quality Reading Project Kyrgyz Republic: Final EGRA and Impact Report 2013–2017*. (Washington, DC: United States Agency for International Development, 2017), p. 30.

¹⁶ RTI International. *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*. (Washington, DC: United States Agency for International Development, 2015). p. 110.

To meet these goals, the study draws on data collected at baseline in April and May 2018, at midline in April and May 2019, and at endline in April and May 2021. At each phase, data collection captured student, teacher, and school data using the following tools—the EGRA, which measured student performance on the basic foundational skills required for fluency in reading; a student survey, which captured background information about the student being assessed; and Snapshot of School Management Effectiveness (SSME) tools, which captured teachers’ and administrators’ perceptions of school management and pedagogic practices in the school.¹⁷ At midline and endline, the study also used two different lesson observation tools.

RESEARCH QUESTIONS

This endline EGRA report seeks to answer five research questions:

1. How do Tajik and Russian reading outcomes of students in grade 2 and grade 4 nationally vary by subgroups and across time points (baseline, midline, endline)?¹⁸
2. Which RWM geographic subgroups require supplementary attention, and what kind of supplementary attention is required?¹⁹
3. Do materials provided by RWM, or teacher instructional practices supported by RWM, serve as predictors of Tajik and Russian reading outcomes of students in grade 2 and grade 4 in schools supported by RWM? If yes, which materials or practices are predictors?
4. Which contextual factors or other classroom measures are predictors of Tajik and Russian reading outcomes of students in grade 2 and grade 4 in schools supported by RWM?
5. What proportion of students can read and understand the meaning of grade-level text (Standard Foreign Assistance (F) Indicators ES.1-1 and ES.1-2) at each time point in schools served by RWM in grade 2 and grade 4 in Tajik and in Russian?²⁰

Results in this report are organized by the language of the assessment—Tajik or Russian—and by grade level—grade 2 and grade 4. Results are further disaggregated by sex, region, and rural or urban status.²¹ This study examines only public schools with primary grades that offer reading instruction in Russian or Tajik. For data collection, RWM drew a representative sample of the Tajik schools in RWM cohort 2 and assessed every Russian school in RWM cohort 2.²²

¹⁷ The SSME tools were designed to capture indicators that are believed to affect student learning, and the results of these tools can inform education stakeholders about current practices in schools and classrooms in Tajikistan.

¹⁸ Subgroups include gender, urban/rural status, and region.

¹⁹ Subgroups requiring supplemental attention were detected by examining EGRA and SSME results of regions by urban/rural status. The type of supplementary attention was based on student EGRA subtask performance or evidence of the intervention in the school measured by the SSME.

²⁰ ES.1-1 is calculated by finding the proportion of students in Grade 2 who read more than 40 CWPM on the ORF subtask. ES.1-2 is calculated by finding the proportion of students in grade 4 who read more than 80 CWPM on the ORF subtask.

²¹ Dushanbe is considered one region for disaggregating. Khatlon is separated into two separate regions, Khatlon-Bokhtar and Khatlon-Kulob.

²² Cohort 2 includes those schools in which the RWM intervention began during the 2018-19 school year.

METHODS AND LIMITATIONS

This section describes the methods used to address the study’s research questions, including tool development, study design, sampling, and analysis, as well as the study’s limitations. Of note, RWM team members and USAID were consulted on decisions related to study design and tool development, but individuals independent from RWM carried out certain activities related to the study—including tool piloting and finalization, data collection, and data analysis—to mitigate potential bias or influence on the study.

TOOL DEVELOPMENT

EGRA BASELINE ADAPTATION, PILOTING, AND EQUATING

The RWM endline EGRA builds on previous USAID-funded early grade reading research in Tajikistan. In 2017, following guidance from USAID/Central Asia, RWM reviewed existing EGRA tools used in the QRP endline assessment and conducted an EGRA adaptation workshop with the project’s assessment working group comprised of local education experts, officials, and stakeholders. The workshop updated EGRA tools for both languages and grades as needed to bring existing tools into closer alignment with the latest EGRA toolkit. The student survey and SSME tools were also developed in consultation with the working group. RWM piloted these revised tools in December 2017 and used them for baseline data collection in April and May 2018.

EGRA MIDLINE AND ENDLINE UPDATES

Prior to both midline and endline data collections, RWM revised existing EGRA tools to prevent any leakage effects in case schools had obtained copies of the tools. Revisions included re-randomization of the order of items in subtasks that did not require any equating, including letter name identification, familiar word reading, and nonword reading. Passage-based subtasks—including ORF with reading comprehension, silent reading competition, and listening comprehension—had certain words changed while keeping the overall story structure and difficulty as close to that of the baseline’s as possible. ORF scores were equated using a common-persons design. Comprehension questions were not equated due to the low number of items; this decision is supported by EGRA Toolkit 2.0 guidance. To maintain consistency, the initial sound identification subtask was not modified. RWM specialists and the National Testing Center reviewed revised drafts to verify that difficulty was unchanged.

Prior to the midline data collection, RWM also made targeted revisions to the student survey and SSME tools to remove items that were not found to be useful during the baseline analysis, reporting, and validation process. Several new questions were also added at the suggestion of RWM technical staff. In addition, RWM developed a new lesson observation tool, which was incorporated into the SSME package.²³

For the endline study, RWM replaced the midline lesson observation tool with a new version intended to align more closely with RWM activities. RWM led a four-day remote training in February 2021 for four pilot testers, who completed 20 lesson observations in Russian and Tajik in grades 2 and 4. After STS analyzed the pilot data and made initial changes to the tool, RWM provided final feedback, and the classroom observation tool was finalized in March.

INSTITUTIONAL REVIEW BOARD

²³ This tool was adapted from the open-source World Bank Teach tool. Revisions were made to more closely align with RWM activities. World Bank Group, “Teach Training Manual,” <http://saber.worldbank.org/index.cfm?indx=5&sub=7> (accessed March 25, 2019).

Institutional review boards (IRBs) are responsible for ascertaining the acceptability of proposed research regarding institutional commitments and regulations, applicable laws, standards of professional conduct and practice, and ethical and societal norms. IRBs examine subject recruitment procedures, proposed remuneration, and the informed consent process. IRBs also evaluate the potential risks and benefits to participants outlined in each protocol. Solutions IRB, an accredited IRB in the United States, approved this EGRA study.

EGRA AND SSME TOOLS

The RWM endline EGRA tools included the subtasks described in Table 7. The silent reading comprehension subtask is an additional measure of comprehension that RWM developed. It resembles the reading comprehension subtask in that students are provided with a short, written grade-level text and asked five comprehension questions. It differs from the subtask in that students have three minutes to read it; they may read silently; they are able to refer to the text while responding to questions; and it is a slightly more challenging text.

Table 7. Endline EGRA Subtasks

Subtask	Grade
Letter name identification	Grade 2 only
Initial sound identification	Grade 2 only
Familiar word reading	Both grade 2 and grade 4
Nonword reading	Both grade 2 and grade 4
Oral reading fluency	Both grade 2 and grade 4
Reading comprehension	Both grade 2 and grade 4
Listening comprehension	Both grade 2 and grade 4
Silent reading comprehension	Both grade 2 and grade 4

In addition to the EGRA subtasks, each student was administered a student survey.

The SSME tools were used to provide a multifaceted picture of school and classroom management practices are described in Table 8.

Table 8. Midline SSME Tools

Tools	Respondents	Number administered
School inventory	Completed by enumerator, accompanied by school director or deputy school director, if appropriate	One per school
Classroom inventory	Completed by enumerator in grade 2 and grade 4 classes sampled	One per grade, two per school
School director questionnaire	School director or deputy school director	One per school
Teacher questionnaire	Grade 2 and grade 4 teachers of the classes sampled	One per grade, two per school
Lesson observation	Completed by enumerator in grade 2 and grade 4 classes sampled	One per grade, two per school

ENDLINE SAMPLING

At a minimum, *power calculations* indicated that a sample of 270 students per language per grade—10

students per school per grade—was required to generalize results to the project population.²⁴ In addition, the study was designed to generalize results for various subgroups, which increased the sample size. The EGRA baseline study sample was designed to draw statistical comparisons between different intervention arms.²⁵ This design was not used for the midline or endline studies; one arm of schools was removed after the baseline analysis—those not served by RWM—while sampled schools served by RWM were retained in the midline and endline samples. The sample of schools was stratified and drawn in accordance with the representation of the full population of schools in Tajikistan based on region, language of instruction (LOI), and urbanicity. Students were stratified at the class level by sex. Given the low numbers of Russian-language schools, a census approach was taken, and enumerators assessed students from all Russian-language schools served by RWM. Between midline and endline data collection, seven Russian-language schools could no longer be included because they had closed, ceased teaching in Russian, or did not have grade 2 or 4 students. In addition, six schools that had previously not taught in Russian began using it as the LOI and were added to the endline sample. Only one Tajik-language school, in GBAO from Cohort 2, was removed from the endline sample, as it had not participated in RWM activities

FINAL SAMPLE

At each school, 10 students in each grade—five girls and five boys—were randomly selected and assessed. No schools were replaced during data collection. Table 9 and Table 10 show the achieved sample size by language. Enumerators assessed a total of 3,858 students at 202 schools.

Table 9. Number of Students Assessed, Tajik Sample, 142 Schools

	Grade 2	Grade 4	Total by Sex
Girls	670	677	1,347
Boys	672	671	1,343
Total by Grade	1,342	1,348	2,690

Table 10. Number of Students Assessed, Russian Sample, 60 Schools

	Grade 2	Grade 4	Total by Sex
Girls	283	300	583
Boys	286	299	585
Total by Grade	569	599	1,168

While the number of Tajik students assessed was slightly below the desired target due to insufficient numbers of students at some schools, this sample size was sufficient for the analyses conducted and

²⁴ To compare by subgroups in Tajik, this sample assumes a stated power of 0.80, a margin of error of 3.7 percent, an ICC of 0.2, and an α (alpha) of 0.05 to be able to detect an effect size, as identified by *Cohen's d* of 0.25. With these assumptions, this sample size also allows for statistical comparisons by student sex and location with the ability to detect an effect size of 0.32. With the assumptions noted above and representative coverage of the five regions, the study can detect effect sizes equal to 0.22. This sample assumes an ICC of 0.2, the average for these types of studies; a standard deviation of the ORF subtask scores, based on scores from the QRP Midline Report; and a desired confidence band width of 10, with 95 percent confidence that the ORF subtask scores are ± 5 points of the mean.

²⁵ The baseline study aimed to provide estimates of RWM intervention effects on reading progress at the primary level by comparing performance between populations receiving and not receiving RWM, including those that had participated in QRP. The three groups were 1) "QRP-only," defined as public schools with primary grades that received QRP interventions but did not receive RWM interventions; 2) "QRP+RWM," defined as public schools with primary grades that received both QRP and RWM interventions; and 3) "RWM-only," defined as public schools with primary grades that did not receive QRP interventions but did receive RWM interventions. The first group was removed at midline.

for generalizing to the subgroups of interest. The number of students assessed in Russian was also sufficient for the analyses conducted and generalizing within the census population.

ENDLINE DATA COLLECTION

This section describes endline operational data collection, including enumerator training, operational data collection, and data entry and cleaning. For more details about these activities and the processes followed during them, see Annex A.

ENUMERATOR TRAINING

RWM took precautions to minimize risk of COVID-19 during both training and data collection. All participants were tested for COVID before training. No international STTA traveled for the training. Instead, STTA helped conduct the training of supervisors and observers remotely.

Data collection teams included a supervisor, a classroom observer, and two enumerators. These three groups were trained in separate groups. First, a training of EGRA trainers was held March 24–26 in Dushanbe. This training prepared a set of experienced, regionally based trainers who included the RWM Regional M&E specialists and two external trainers to lead enumerator trainings in each region of the country. These trained individuals also served as Quality Control Officers (QCO). Second, the supervisor and QCO training took place March 28–29 in Dushanbe. Then, the QCOs traveled to their home regions to deliver regionally based EGRA training sessions over four consecutive days, including one day to practice the tools and procedures at a school to provide enumerators with the opportunity to practice in real-world conditions. Regional trainings covered the tools, sampling, data collection protocols, data management, and reporting requirements. These regional EGRA enumerator trainings took place between April 2 and 10, depending on the region.

DATA COLLECTION

Endline data collection took place in every region of Tajikistan in April and May 2021. Sixteen teams—11 Tajik-speaking and five Russian-speaking—collected data from April 7 to May 3. Each team visited one school per day. Each of the 16 teams consisted of one supervisor, who led sampling and administered the teacher interview, director interview, and school inventory; one classroom observer; and two EGRA enumerators. In total, 202 schools were assessed.

RWM implemented a variety of strategies to track the progress of data collection, as well as provide oversight and quality assurance checks on the EGRA and SSME data collection. Each region was assigned a QCO, who visited every team in the assigned region at least once. School-to-School International (STS) local staff in Tajikistan also visited 14 schools in person to observe data collection. Thus, each data collection team had at least one on-site spot check, and many teams were visited several times.

DATA ENTRY AND CLEANING

Throughout operational data collection, RWM followed the guidance laid out in the *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*, also known as the EGRA Toolkit 2.0, by regularly uploading and reviewing data to better manage and track data collection issues and progress.²⁶ QCOs ensured data collection procedures were followed and submitted daily reports that logged any discrepancies in the number and type of data collected that differed from the intended sample.²⁷

²⁶ RTI International. *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*. (Washington, DC: United States Agency for International Development, 2015). p. 103.

²⁷ These reports documented the school demographics, type and number of each assessment or questionnaire collected, status of data upload, and any other issues or challenges encountered that day in the school.

These reports were later cross-referenced against the uploaded data in Tangerine and Ona.²⁸ Disposition codes were applied to categorize the various issues or problems that emerged during the data collection process. These codes were used in determining cleaning rules that were incorporated into the database using syntax to clean the data accordingly. These coding and flagging procedures helped to ensure the various and nuanced contexts of data collection at schools were sufficiently cataloged and considered during the data cleaning, analysis, and reporting process.

ANALYTIC METHODS

This section describes how endline results were equated with results from a previous project; how weights were calculated and applied to ensure representativeness in results; and how findings were generated. For more technical details about these methods, see Annex B.

STUDY DESIGN

RWM implementation has been staggered across four cohorts of schools. Cohort 1 schools received RWM interventions beginning in January 2018. Cohort 2 schools began receiving the intervention at the beginning of the 2018–19 school year in September 2018. Cohort 3 and 4 schools began receiving the intervention at the beginning of the school year in 2019 and 2020, respectively.

This study drew a *stratified random sample* of schools from cohort 2 schools served by RWM. The study first gathered pre-implementation baseline data on student reading outcome data from these schools in spring 2018. This cohort of schools received RWM interventions in a staggered rollout. The earlier intervention began with teacher training in August 2019, though full implementation was not reached until March 2019. This endline report thus examines cohort 2 schools after nearly three academic years of exposure. Table 11 describes the rollout of RWM cohorts by the month of first program implementation.

Table 11. Timeline of RWM Interventions by Cohort

Cohort	January–March 2018	April 2018	August 2018 – March 2019	April 2019	August–December 2019	August–December 2020	April 2021
Cohort 1	Teacher training conducted						
Cohort 2 (EGRA sample)		Baseline EGRA	Teacher training conducted	Midline EGRA			Endline EGRA
Cohort 3					Teacher training conducted		
Cohort 4						Teacher training conducted	

EQUATING

One of the objectives of the current study is to compare the average performance of students participating in RWM over time. To that end, and similarly to midline, it was necessary to conduct a procedure known as *statistical equating*, which brings the scores of two forms of a test into a

²⁸ Tangerine is a commonly used application to collect EGRA data. Ona is a mobile data collection application, built on an Open Data Kit Collect platform, used by RWM to collect SSME data.

common scale. Statistical equating is a required procedure to establish comparisons across EGRA administrations because some versions of the ORF passage over time may be more difficult than others; this makes it impossible to gauge the magnitude of the differences in the performance of students who take different forms. Equating techniques vary according to their data collection design and the statistical methods chosen.

At endline, equating was carried out for the ORF subtask and was meant to render scores from the RWM endline forms equivalent to those of the baseline (and midline) forms. The analysts followed equating methodologies recommended by the EGRA Toolkit 2.0 and used a single-group design in which the same students read both ORF passages—endline and baseline—enabling analysts to directly attribute differences in difficulties to the items included in each form. Please see Annex B for further details.

Only scores from the ORF subtask were equated to baseline. Statistical equating was not carried out on the other EGRA subtasks. For letter naming, familiar word reading, and nonword reading, items from the baseline forms were re-randomized to ensure a common scale between data collection points. For initial sound identification, the same form used at baseline was used at midline and endline to ensure direct comparability. Passage-based comprehension subtasks—such as reading comprehension, silent reading comprehension, and listening comprehension—provided too few scores to be reliably equated. Instead, these subtasks underwent targeted changes in word choice while keeping the overall story structure and difficulty as close as possible to that of the baseline and midline.

STATISTICAL WEIGHTING

As with other time points, the RWM endline analysis used *sampling weights* to minimize bias on the estimates conducted in the sample of students. Random sampling does not account for the fact that some students have a lower probability of being selected when they are in schools of varying size or represent smaller subgroups within the population; sampling weights allow the analysts to account for these differences in probabilities.

Analysts computed the weights using variables, including the type of school, region, and the number of grade 2 and 4 classrooms and the students in each classroom at each school. Weights were computed separately for each language and grade level. STS collected information from the project via Education Management Information System datasets; sampling weights were updated based on the latest information provided.

CHARACTERISTICS OF ASSESSMENT TOOL

Analyzing the quality of the assessment ensures that the conclusions drawn about student performance on the assessment are valid. The quality of the assessment tool was analyzed in four ways. First, correlations between each subtask of the EGRA were analyzed. Second, the internal consistency of the full EGRA assessment was analyzed using coefficient alpha. Third, the difficulty of each task was analyzed using percentage correct scores. Fourth, item analyses of difficulty and discrimination were completed. All results are reported in Annex C.

GENERATION OF FINDINGS

After applying the equating and weighting functions to the clean datasets, analysts generated mean-scores, zero-scores, percent-correct scores, and reading-benchmark estimates for endline EGRA data for each dataset and for each of the disaggregated groups of interest. In addition, and in accordance with midline analysis, analysts generated composites for the different SSME questionnaires administered. Descriptive statistics from the EGRA, student survey, and the SSME questionnaires and inventories were generated, and findings were triangulated where possible.

To respond to the research questions, country-level estimates were based on aggregated scores,

while differences between within-group categories for one or multiple time points were analyzed using independent groups t-tests and Stata outputs on weighted standard errors and confidence intervals. The identification of subgroups at risk was based on a hierarchical linear regression²⁹ analysis by region, urbanicity, and sex. The analysis of classroom-level or other predictors was conducted via hierarchical linear regression analyses, which controlled for the region and the urbanicity of schools. Lastly, the analysis of benchmarks was conducted using intervals of oral reading fluency and t-tests to compare results across time.

All analysis was completed using Stata version 16 software.

Findings of this report will be validated through a remote or an in-person meeting with USAID and an assessment working group comprised of officials from the MoES and related education agencies in Tajikistan. RWM will work with government counterparts through the validation workshop to ensure they are grounded in the local context and meaningful to stakeholders. Interpretation of results will encourage and support MoES further use of EGRA results to determine policy objectives. For example, following the midline 2019 validation, the MoES was supportive of further presentations to regional and district education departments as well as other local and international stakeholders. A similar process is expected following validation of these endline results.

LIMITATIONS

The following limitations should be kept in mind while reviewing the results reported in this document:

1. The non-experimental nature of this study design limits the level of attribution of the RWM intervention on reading outcomes as measured by EGRA. The EGRA uses a non-experimental design because RWM is primarily an implementation project that does not randomly assign schools to treatment conditions or cohorts. As there is no counterfactual, this study cannot attribute results to RWM.
2. As with most EGRA studies, these results do not provide appropriate data for cross-linguistic comparisons; that is, student reading skills in Tajik should not be directly compared with student reading skills in Russian. Acquisition of language and reading development depend on several factors, including the different levels of orthographic transparency, visual complexity, and phonology. In addition, the composition of the Tajik and Russian samples differs considerably, which further highlights the importance of avoiding comparisons across languages.
3. This design assumes that Cohort 2 schools are representative of the RWM school population to generalize results to schools participating in the other cohorts. Additionally, implementation is assumed to be uniform across all schools. While Cohort 2 schools may benefit from refinements in the intervention, later cohorts were assumed to similarly benefit from lessons learned during implementation. Of course, implementation was not uniform across all cohorts. For example, teacher training in Cohort 4 employed a blended learning approach. Teachers accessed all RWM materials, such as training modules, through an online platform. RWM developed a community-based mentoring system that combined several schools into a single unit. RWM trained a highly experienced teacher within each unit, who then trained the other teachers within that unit. The online platform recorded training progress. Project mentors reinforced and expanded training during regular monitoring visits to the basic schools.
4. Baseline results in this report differ from baseline results in the RWM EGRA baseline report published in 2018 to provide a more accurate estimation of student performance in the

²⁹ Except for Tajik grade 4 which presented a very low level of ICC (< 6%).

sample population. Results presented in this report are aligned to those from the midline report.

5. The analysis of factors associated to ORF scores (research questions 3 and 4) should not be interpreted as causal relations. First, the analysis is based on hierarchical linear regression over a non-experimental design. Second, for simplicity and richness, each factor was analyzed separately; it is possible that an aggregated regression analysis would modify some of the reported relations.

RESULTS

Evaluation Question 1: How do Tajik and Russian reading outcomes of students in grade 2 and grade 4 nationally vary by sub-groups and across time points (baseline, midline, endline)?

Overall and across grades and languages, students’ performance was better at endline than baseline, with the difference being significant on most subtasks. Students across grades and languages consistently improved their performance on the silent reading comprehension subtask, as displayed in Table 12, suggesting that the RWM project successfully improved their ability to understand written texts.

Table 12. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask, between Baseline and Endline

EGRA Subtask	Endline relative to baseline			
	Tajik Grade 2	Tajik Grade 4	Russian Grade 2	Russian Grade 4
Letter name identification (CLNPM)	↑	N/A	↔	N/A
Initial sound identification	↑	N/A	↑	N/A
Familiar word reading (CFWPM)	↔	↑	↔	↔
Non-word reading (CNWPM)	↑	↑	↑	↔
Oral reading fluency (equated CWPM)	↔	↑	↑	↔
Reading comprehension	↑	↔	↑	↔
Silent reading comprehension	↑	↑	↑	↑
Listening comprehension	↔	↑	↔	↑

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a horizontal arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the $p < 0.05$ level.

Part 1: EGRA Results by Year

TAJIK GRADE 2 RESULTS BY YEAR

Tajik grade 2 students performed better at endline than baseline on most subtasks, with the difference being significant for five subtasks—letter name identification, initial sound identification, nonword reading, reading comprehension, and silent reading comprehension. As shown in Figures 1 and 2, there was an increase in performance for fluency and non-fluency subtasks, particularly between midline and endline. Listening comprehension was the only subtask in which Tajik grade 2 students did not show growth between baseline and endline, but this result is likely due to an initial ceiling effect at baseline. In other words, students’ listening comprehension scores were high at baseline, and the number of questions asked as part of the subtask—only five—was not sufficient to capture more nuanced differences at the higher end of the scale.

Figure 1. Trends in Tajik Grade 2 Reading Performance by Time Point: Fluency Tasks

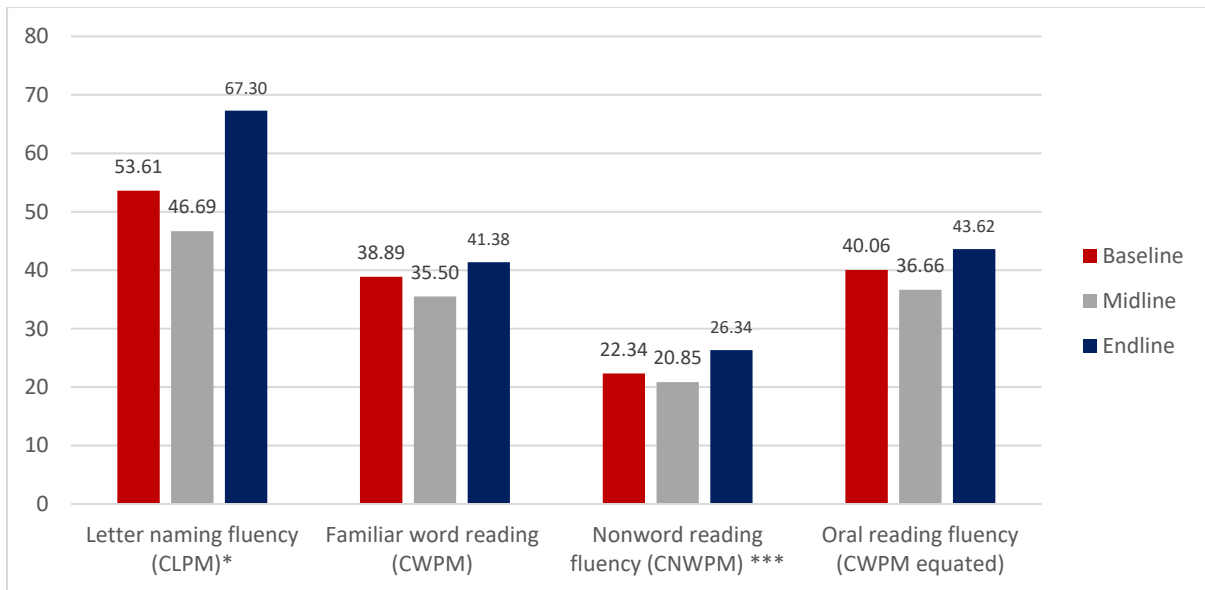
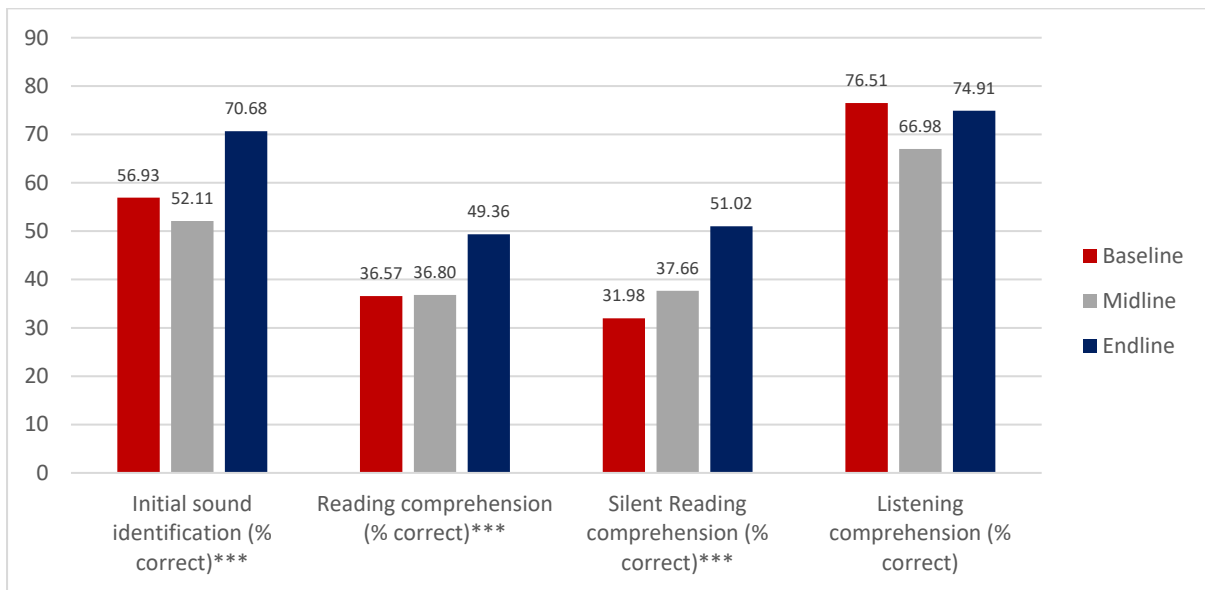


Figure 2. Trends in Tajik Grade 2 Reading Performance by Time Point: Non-Fluency Tasks



TAJIK GRADE 4 RESULTS BY YEAR

Tajik grade 4 students performed better at endline than baseline on all subtasks, with the difference being significant for every subtask except for reading comprehension. As illustrated in Figures 3 and 4, the most consistent gains were observed with ORF, silent reading comprehension, and listening comprehension; these skills were gained evenly throughout the life of the project. In other words, there was an increase in performance between baseline and midline and an increase in performance between midline and endline. However, for all other tasks, the performance at baseline and midline remained relatively unchanged, meaning that the gains in performance occurred between midline and endline.

Figure 3. Trends in Tajik Grade 4 Reading Performance by Time Point: Fluency Tasks

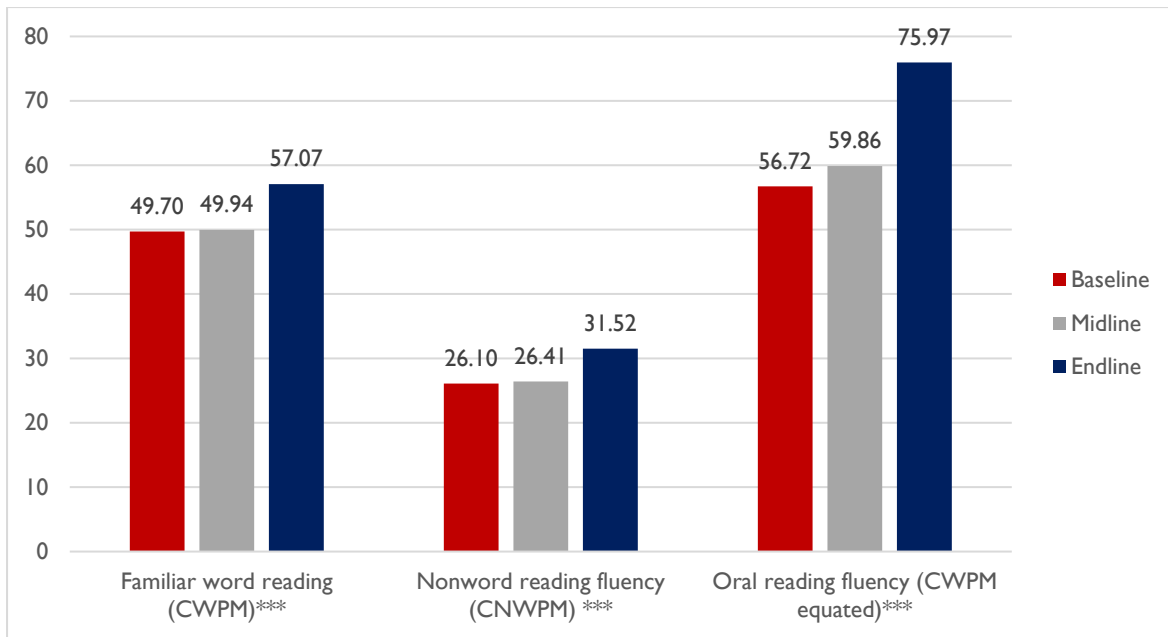
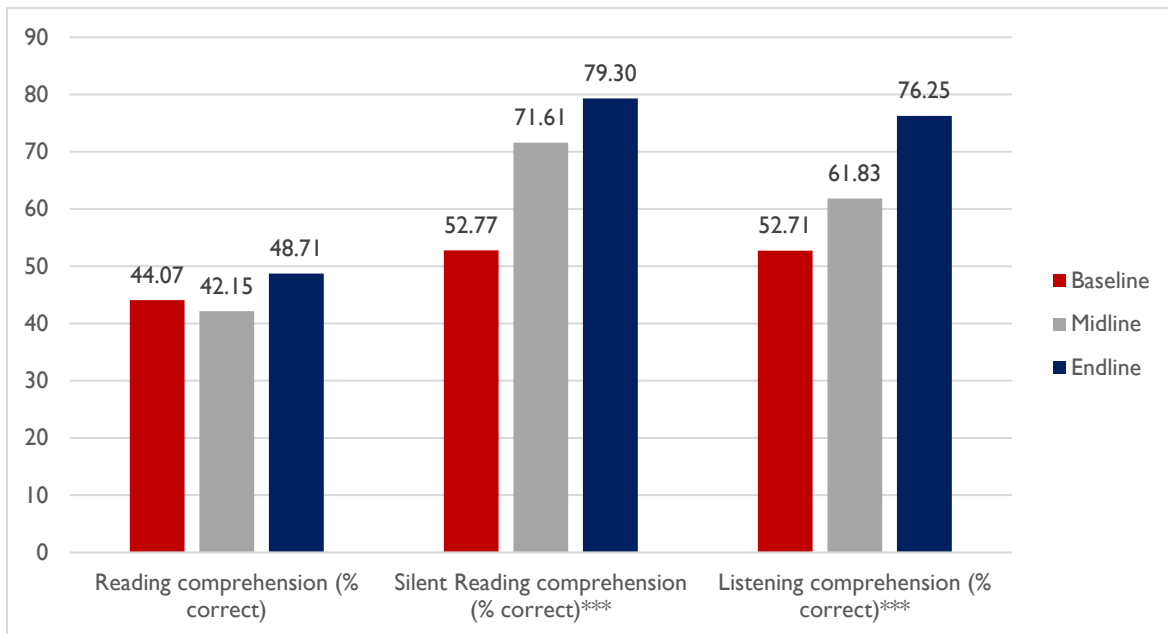


Figure 4. Trends in Tajik Grade 4 Reading Performance by Time Point: Non-Fluency Tasks



RUSSIAN GRADE 2 RESULTS BY YEAR

Russian grade 2 students performed better at endline than baseline on most subtasks, with the difference being significant for five subtasks—initial sound identification, nonword reading, ORF, reading comprehension, and silent reading comprehension. As displayed in Figures 5 and 6, gains were more pronounced for initial sound identification, reading comprehension, and silent reading comprehension. As for letter name identification, students’ performance was slightly lower at endline than baseline, but this difference was not statistically significant.

Figure 5. Trends in Russian Grade 2 Reading Performance by Time Point: Fluency Tasks

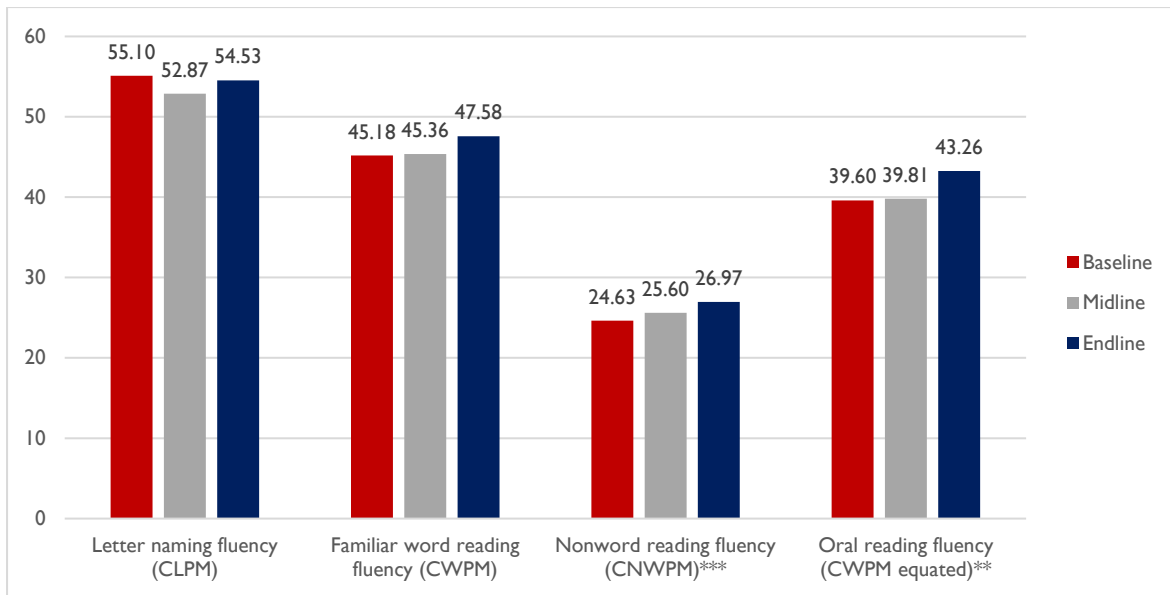
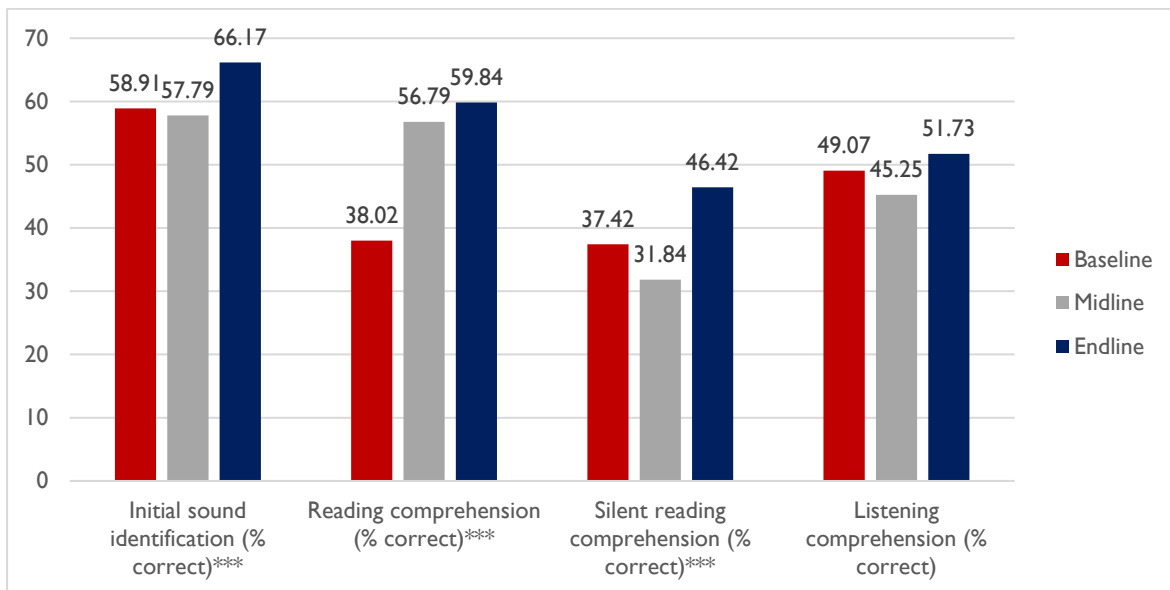


Figure 6. Trends in Russian Grade 2 Reading Performance by Time Point: Non-Fluency Tasks



RUSSIAN GRADE 4 RESULTS BY YEAR

Russian grade 4 students performed better at endline than baseline on most subtasks, with the difference being significant for silent reading comprehension and listening comprehension. As illustrated in Figure 7, students' performance on familiar word reading and nonword reading was lower at endline than baseline, but these differences were not statistically significant. As shown in Figures 7 and 8, the trends across the three time points were mixed. While performance on fluency subtasks was relatively flat, it was more varied on non-fluency tasks. In particular, while students' performance gradually improved on reading comprehension and silent reading comprehension, it was worse at endline than midline on listening comprehension. It is difficult to infer strong conclusions, however, because only four questions were asked as part of the listening comprehension subtask. Still, students' performance was significantly lower at endline than midline on this subtask.

Figure 7. Trends in Russian Grade 4 Reading Performance by Time Point: Fluency Tasks

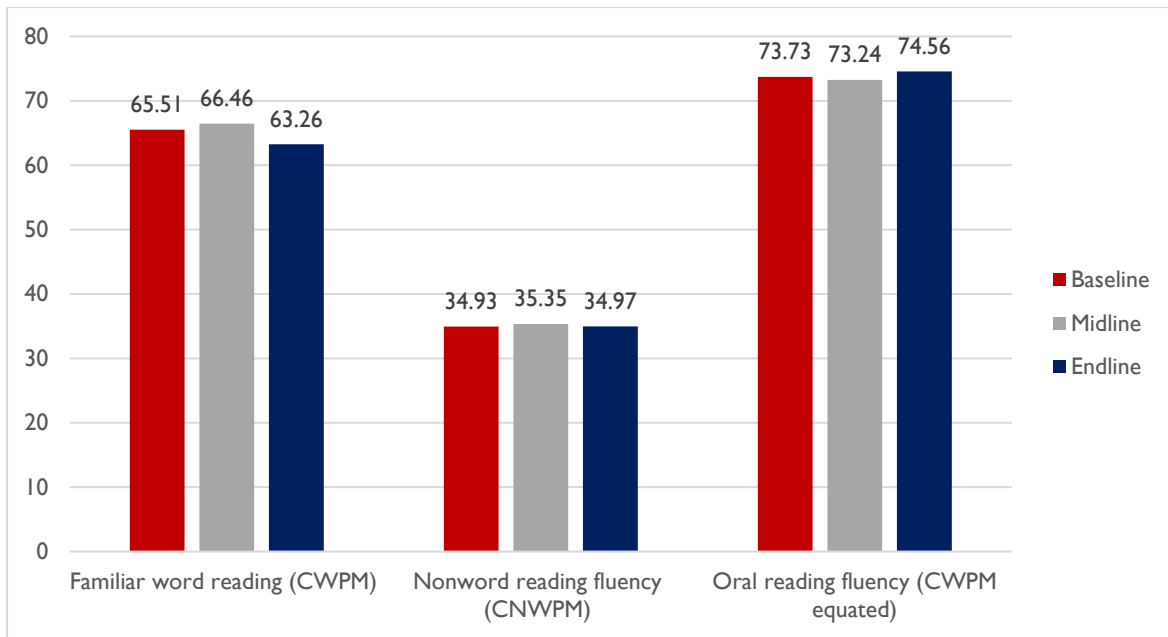
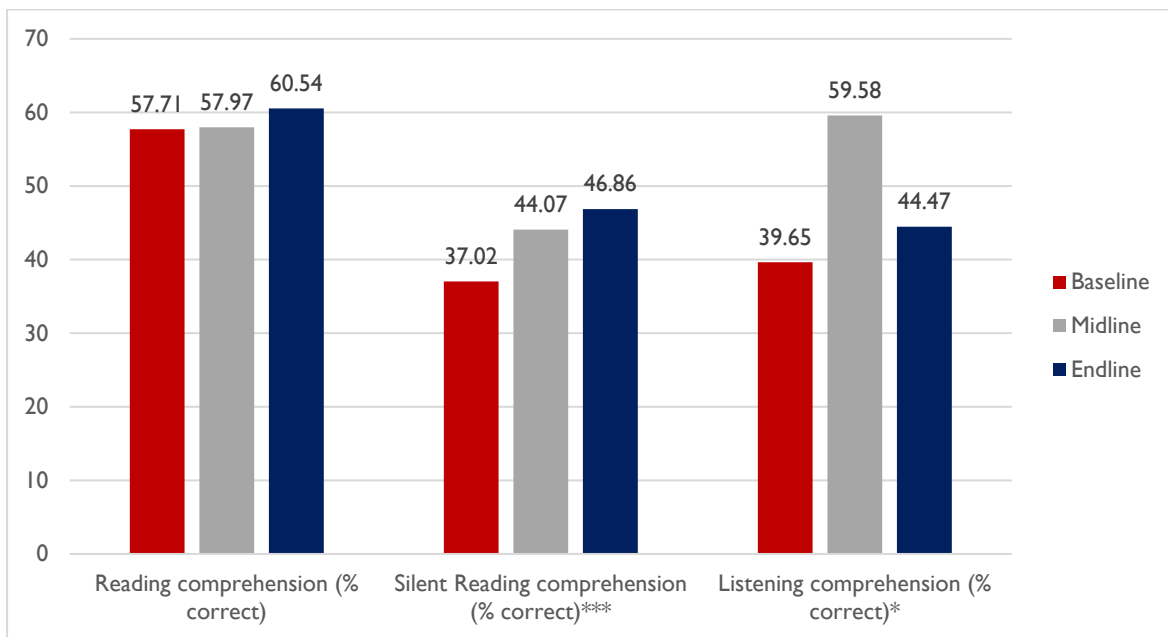


Figure 8. Trends in Russian Grade 4 Reading Performance by Time Point: Non-Fluency Tasks



Part 2: EGRA Results by Sex

To a large degree, trends in the differences between boys' and girls' results at baseline and endline mirrored those found in the overall results, as displayed in Table 7, including students showing improvement in silent reading comprehension; grade 2 students showing gains in reading comprehension; most of grade 2 students showing gains in initial sound identification; most of grade 4 students showing gains in listening comprehension; and Tajik grade 4 students showing gains in

familiar words, non-word reading and ORF. However, several differences were notable.³⁰ In particular, some statistical differences in the overall results were driven either by boys' or girls' improved performance, but not both. For example, Tajik grade 2 students' letter name identification scores were significantly higher at endline than baseline. However, this difference resulted due to an increase only in boys' letter name fluency skills, as shown in Table 13. In addition, although overall performance did not decline significantly from baseline to endline on any subtask across languages and grades, it did so in two cases when analyzing results disaggregated by sex. Russian grade 2 boys' scores on letter name identification were significantly lower at endline than baseline, as were Russian grade 4 girls' scores on familiar word reading.

Table 13. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Sex, between Baseline and Endline

EGRA Subtask	Endline relative to baseline							
	Tajik Grade 2		Tajik Grade 4		Russian Grade 2		Russian Grade 4	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Letter name identification (CLNPM)	↑	↔	N/A	N/A	↓	↔	N/A	N/A
Initial sound identification	↑	↑	N/A	N/A	↑	↔	N/A	N/A
Familiar word reading (CFWPM)	↔	↔	↑	↑	↔	↑	↔	↓
Nonword reading (CNWPM)	↑	↑	↑	↑	↔	↑	↔	↔
ORF (equated CWPM)	↔	↔	↑	↑	↔	↑	↔	↔
Reading comprehension	↑	↑	↔	↔	↑	↑	↔	↔
Silent reading comprehension	↑	↑	↑	↑	↑	↑	↑	↑
Listening comprehension	↔	↔	↑	↑	↔	↔	↑	↔

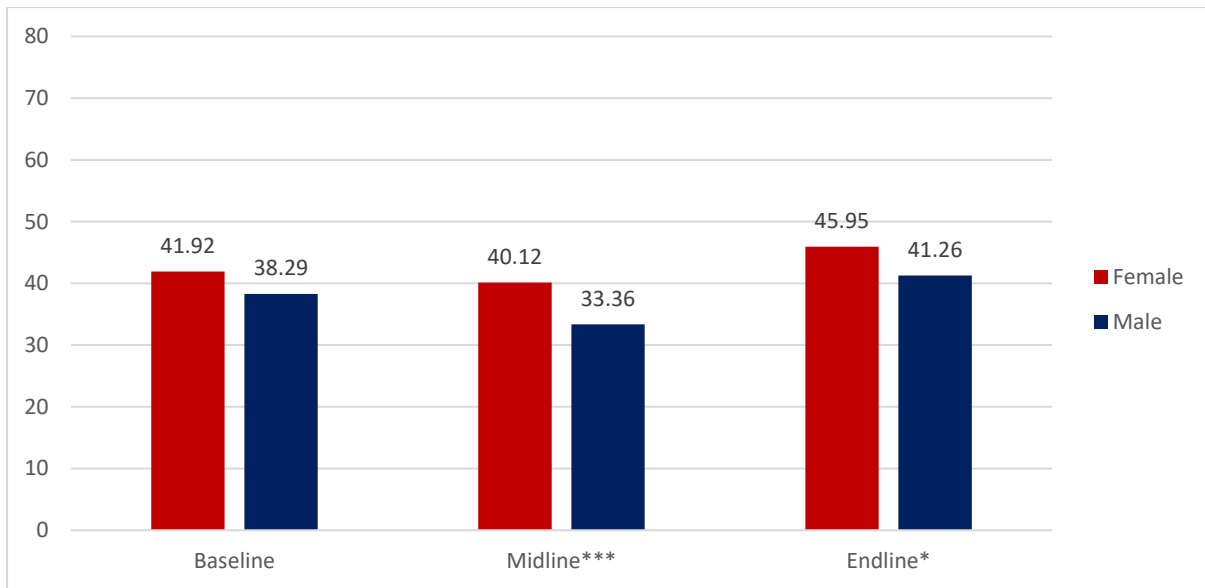
Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a down arrow (↓) indicates that the mean score for students at endline was statistically significantly lower than the mean score for students at baseline; and a left-right arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the p<0.05 level. Differences were computed by sex, language and grade level.

TAJIK GRADE 2 RESULTS BY SEX

Tajik grade 2 girls outperformed their male peers in ORF at all three time points, as shown in Figure 9, with the difference being statistically significant at both midline and endline. These results suggest that gender gaps have not been reversed.

³⁰ Enumerators recorded students' perceived sex, not self-identified gender.

Figure 9. Trends in Tajik Grade 2 Reading Performance by Time Point and Sex: Oral Reading Fluency



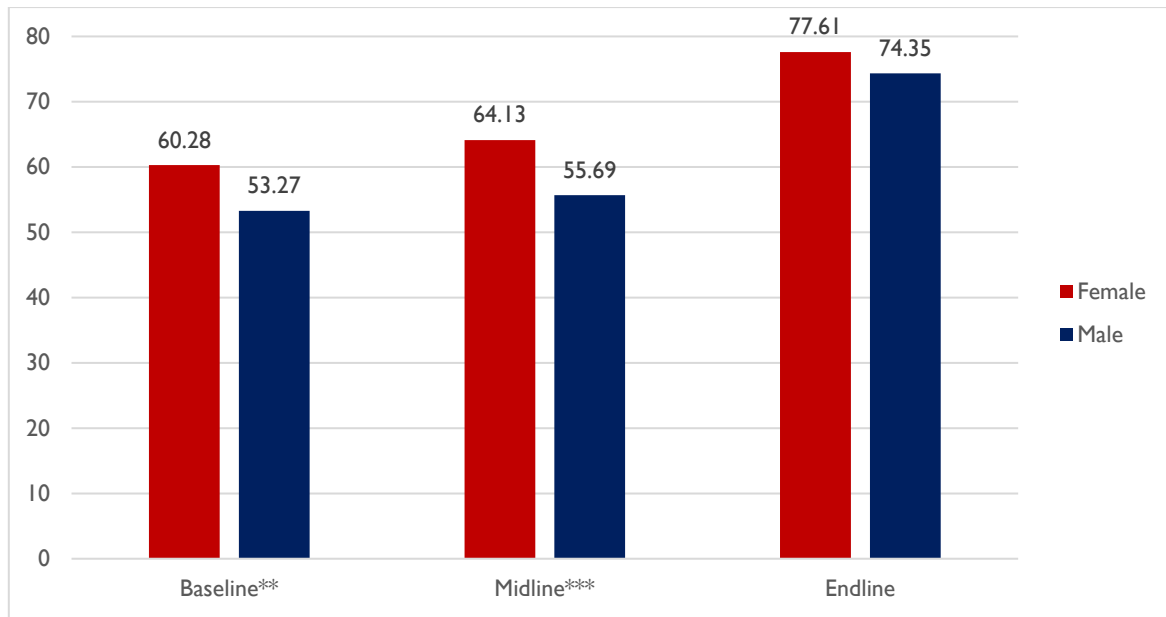
Note³¹: Three asterisks (***) denotes differences between girls and boys that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between girls and boys that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between girls and boys that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between girls and boys was not statistically significant.

TAJIK GRADE 4 RESULTS BY SEX

Tajik Grade 4 girls outperformed their male peers in ORF at all three time points, as shown in Figure 10, with the difference being statistically significant at only baseline and midline. These results suggest that gender gaps have decreased for this population of students.

³¹ As a note, statistical significance is not the same as the size of the differences: two means may be different by the same amount with one of them being statistically significant and the other one not. Statistical significance is also a function of standard errors meaning that variables that are less reliable (i.e. which have higher standard errors) are less likely to be found statistically significant (as the difference may fall within the expected limits of the corresponding distribution).

Figure 10. Trends in Tajik Grade 4 Reading Performance by Time Point and Sex: Oral Reading Fluency

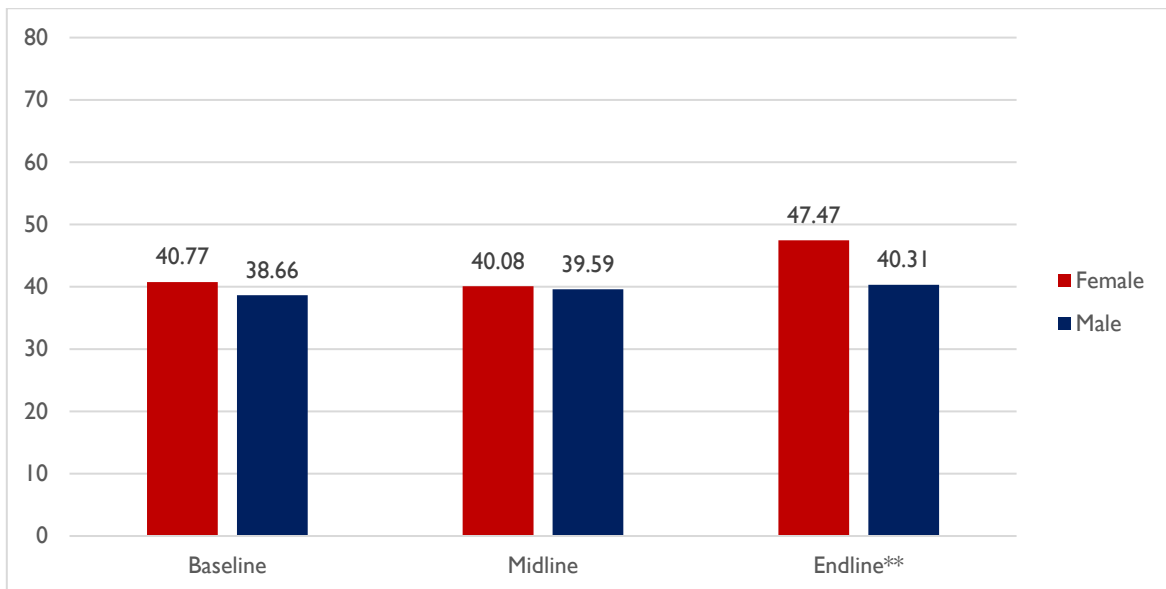


Note: Three asterisks (***) denotes differences between girls and boys that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between girls and boys that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between girls and boys that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between girls and boys was not statistically significant.

RUSSIAN GRADE 2 RESULTS BY SEX

Russian grade 2 girls outperformed their male peers in ORF at all three time points, as illustrated in Figure 11, with the difference being statistically significant only at endline. These results suggest that gender gaps exist and have increased across time.

Figure 11. Trends in Russian Grade 2 Reading Performance by Time Point and Sex: Oral Reading Fluency

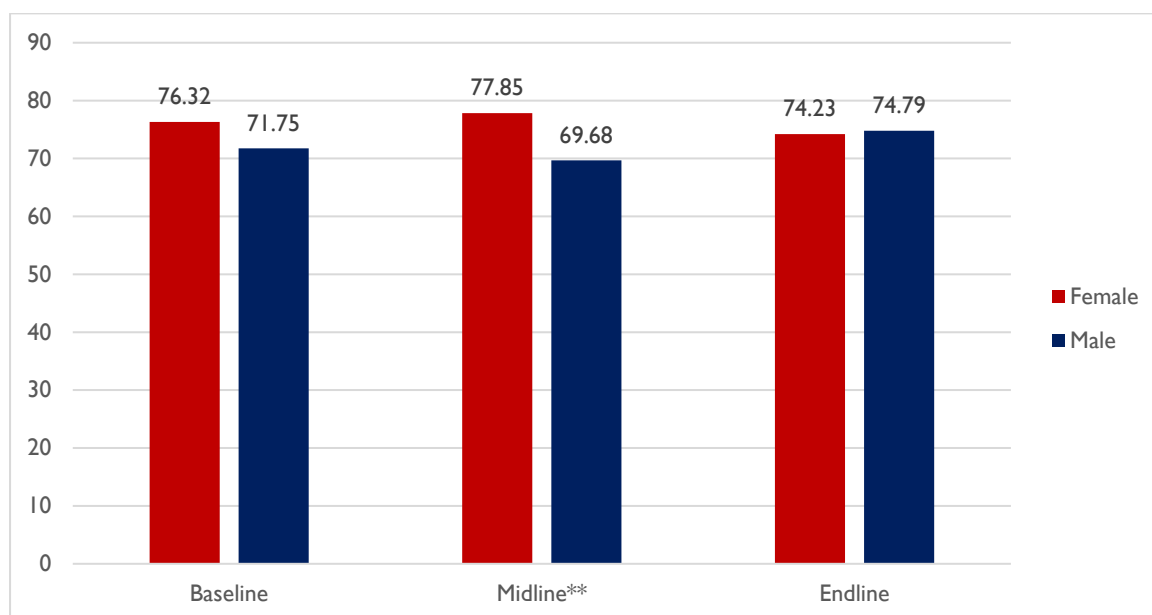


Note: Three asterisks (***) denotes differences between girls and boys that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between girls and boys that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between girls and boys that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between girls and boys was not statistically significant.

RUSSIAN GRADE 4 RESULTS BY SEX

While Russian Grade 4 girls outperformed their male peers in ORF at baseline and midline, as shown in Figure 12, the reverse trend was observed at endline. The difference between the performance of girls and boys was statistically significant only at midline. These results suggest that gender gaps have meaningfully changed over time for this population of students.

Figure 12. Trends in Russian Grade 4 Reading Performance by Time Point and Sex: Oral Reading Fluency



Note: Three asterisks (***) denotes differences between girls and boys that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between girls and boys that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between girls and boys that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between girls and boys was not statistically significant.

Part 3: EGRA Results by Urbanicity

To a large degree, trends in the differences between the performance of schools located in urban and rural areas mirrored those found in the overall results, as shown in Table 8, but several differences are noted. In particular, some statistical differences were driven either by students in urban or rural locations, but not both. Notably, in Tajik schools, rural students' improved performance drove most differences, while, in Russian schools, differences were more balanced between urban and rural schools. For instance, for Russian grade 4, while urban schools accounted for the improvement in silent reading comprehension and listening comprehension, rural schools were responsible for the gains in nonword reading and ORF.

Table 14. Trends in Grade 2 and Grade 4 Student Reading Performance in Tajik by Subtask and Urbanicity, between Baseline and Endline

EGRA Subtask	Endline relative to baseline							
	Tajik Grade 2		Tajik Grade 4		Russian Grade 2		Russian Grade 4	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Letter name identification (CLNPM)	↔	↑	N/A	N/A	↔	↑	N/A	N/A
Initial sound identification	↔	↑	N/A	N/A	↑	↑	N/A	N/A
Familiar word reading (CFWPM)	↔	↑	↔	↑	↔	↔	↔	↔
Nonword reading (CNWPM)	↔	↑	↔	↑	↑	↑	↔	↑

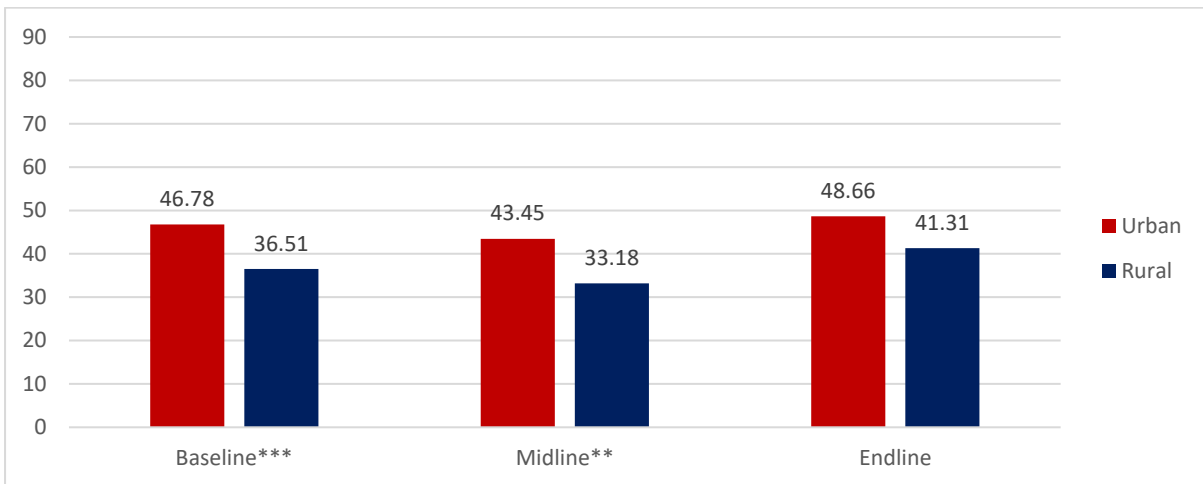
EGRA Subtask	Endline relative to baseline							
	Tajik Grade 2		Tajik Grade 4		Russian Grade 2		Russian Grade 4	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
ORF (equated CWPM)	↔	↑	↑	↑	↑	↑	↔	↑
Reading comprehension	↔	↑	↔	↔	↑	↑	↔	↔
Silent reading comprehension	↑	↑	↑	↑	↑	↑	↑	↔
Listening comprehension	↔	↔	↑	↑	↔	↔	↑	↔

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a left-right arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the $p < 0.05$ level. Differences were computed by urbanicity, language and grade level.

TAJIK GRADE 2 RESULTS BY URBANICITY

Across all time points, Tajik Grade 2 urban students outperformed their rural peers in ORF, as shown in Figure 13. However, the difference was statistically significant only at baseline and midline, suggesting a reduction in the performance gap over time.

Figure 13. Trends in Tajik Grade 2 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency

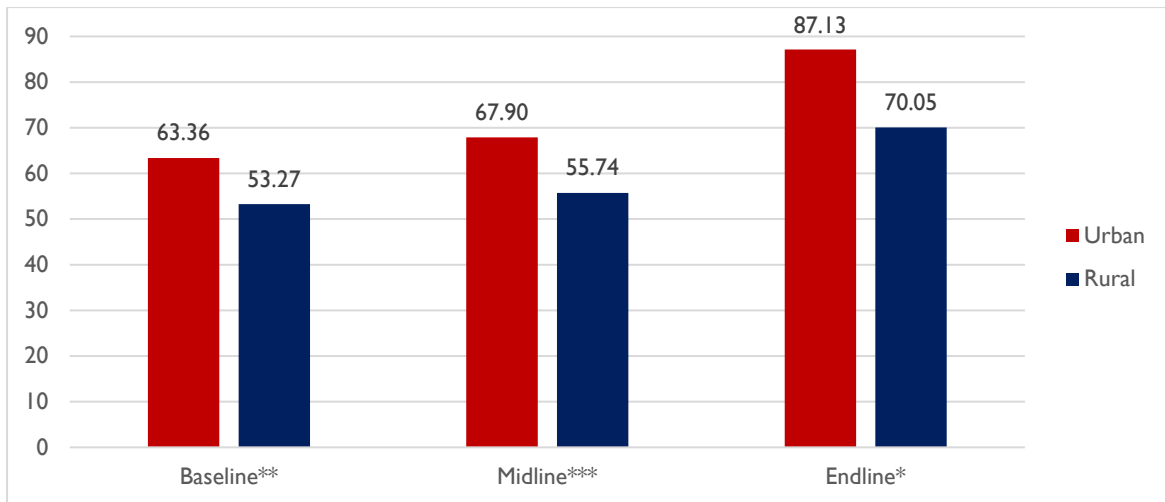


Note: Three asterisks (***) denotes differences between urban and rural students that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between urban and rural students that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between urban and rural students that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between urban and rural students was not statistically significant.

TAJIK GRADE 4 RESULTS BY URBANICITY

Across all time points, Tajik grade 4 urban students outperformed their rural peers in ORF, as shown in Figure 14, with the difference being statistically significant at baseline, midline, and endline. These results suggest a persistent gap between students at different locations.

Figure 14. Trends in Tajik Grade 4 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency

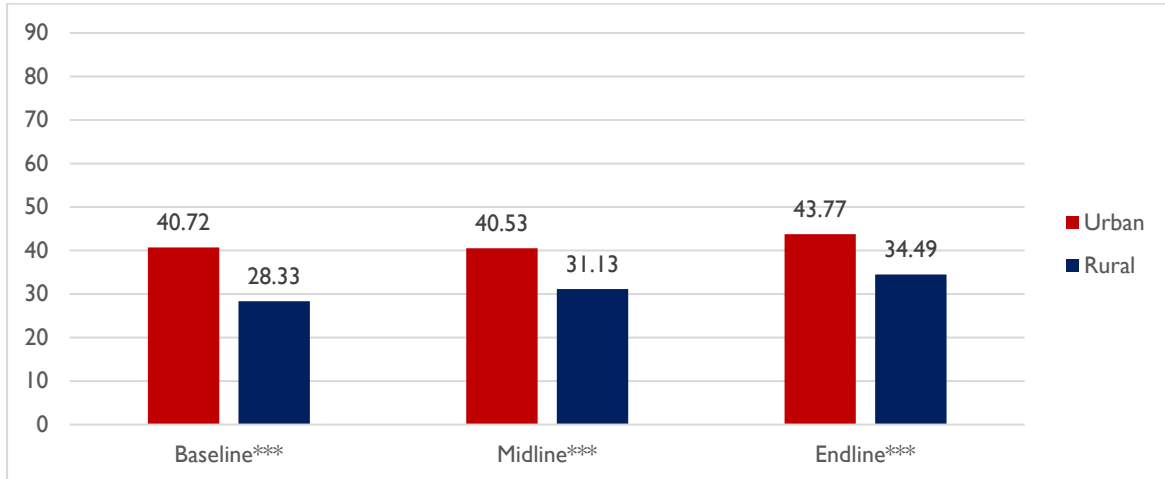


Note: Three asterisks (***) denotes differences between urban and rural students that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between urban and rural students that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between urban and rural students that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between urban and rural students was not statistically significant.

RUSSIAN GRADE 2 RESULTS BY URBANICITY

Across all time points, Russian grade 2 urban students outperformed their rural peers in ORF, as shown in Figure 15, with the difference being statistically significant at baseline, midline, and endline. These results suggest a persistent gap between students at different locations.

Figure 15. Trends in Russian Grade 2 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency

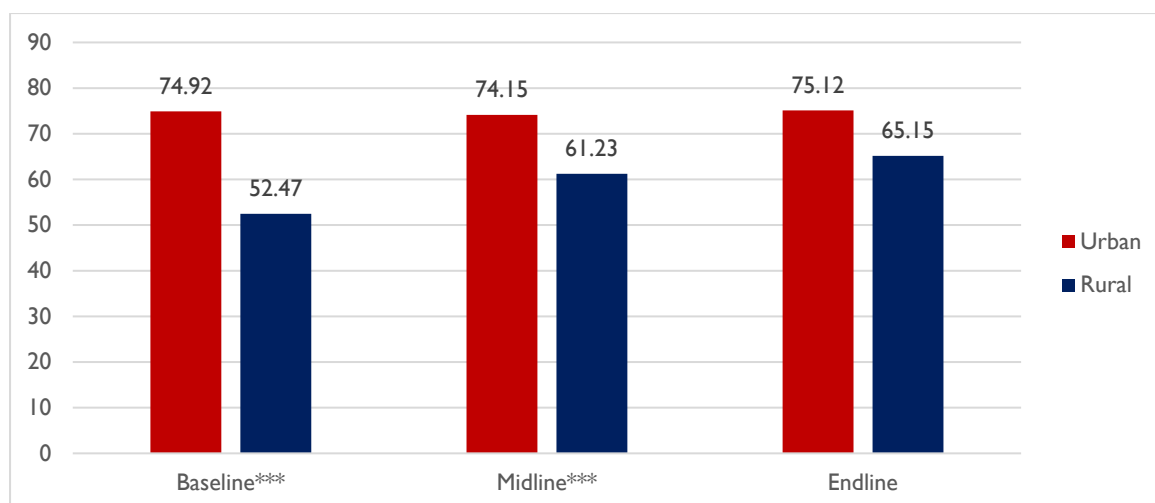


Note: Three asterisks (***) denotes differences between urban and rural students that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between urban and rural students that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between urban and rural students that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between urban and rural students was not statistically significant.

RUSSIAN GRADE 4 RESULTS BY URBANICITY

Across all time points, Russian grade 4 urban students outperformed their rural peers in ORF, as shown in Figure 16. However, the difference was statistically significant only at baseline and midline, suggesting a reduction in the performance gap over time.

Figure 16. Trends in Russian Grade 4 Reading Performance by Time Point and Urbanicity: Oral Reading Fluency



Note: Three asterisks (***) denotes differences between urban and rural students that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between urban and rural students that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between urban and rural students that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between urban and rural students was not statistically significant.

Part 4: EGRA Results by Region

The changes in performance from baseline to endline disaggregated by region shed light on which areas of Tajikistan drove overall gains in student performance, as shown in Tables 9–12. For Tajik schools, students in the regions of GBAO, Khatlon-Bokhtar, Khatlon-Kulob, and Sughd achieved most of the gains, as illustrated in Table 15 and Table 16. While the specific differences varied by subtask, the results showed very consistent gains for these regions. For Russian schools, however, the gains were more balanced across all regions, as shown in Table 17 and Table 18, with the highest gains corresponding to the region of Sughd.³² Although performance improved or remained unchanged on subtasks in most regions in both Tajik and Russian schools, results were more mixed for schools in Dushanbe. While students in Dushanbe had significant improvements on some subtasks, they also had significant decreases in other subtasks, most notably for Tajik grade 2 students.

³² The Russian results disaggregated by region exclude the regions of GBAO and Khatlon-Kulob. The Russian sample did not include schools from GBAO because Russian-language schools were not present in this region at the time of sampling or data collection, and results from Khatlon-Kulob are not shown due to an extremely small sample size (1 school, 10 students).

Table 15. Trends in Tajik Grade 2 Student Reading Performance by Subtask and Region, between Baseline and Endline

EGRA Subtask	Endline relative to baseline - Tajik Schools					
	Grade 2					
	DRS	Dushanbe	GBAO	Khatlon-Bokhtar	Khatlon-Kulob	Sughd
Letter name identification (CLNPM)	↔	↓	↑	↔	↑	↔
Initial sound identification	↔	↓	↑	↑	↔	↑
Familiar word reading (CFWPM)	↔	↔	↔	↔	↔	↔
Nonword reading (CNWPM)	↑	↔	↔	↔	↑	↔
ORF (equated CWPM)	↔	↔	↔	↔	↑	↔
Reading comprehension	↔	↔	↑	↑	↑	↑
Silent reading comprehension	↑	↔	↑	↑	↑	↑
Listening comprehension	↓	↓	↓	↔	↔	↔

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a down arrow (↓) indicates that the mean score for students at endline was statistically significantly lower than the mean score for students at baseline; and a left-right arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the p<0.05 level. Differences were computed by urbanicity, language and grade level.

Table 16. Trends in Tajik Grade 4 Student Reading Performance by Subtask and Region, between Baseline and Endline

EGRA Subtask	Endline relative to baseline - Tajik Schools					
	Grade 4					
	DRS	Dushanbe	GBAO	Khatlon-Bokhtar	Khatlon-Kulob	Sughd
Letter name identification (CLNPM)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Initial sound identification	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Familiar word reading (CFWPM)	↔	↔	↔	↑	↑	↔
Nonword reading (CNWPM)	↔	↑	↔	↑	↑	↑
ORF (equated CWPM)	↑	↔	↑	↑	↑	↑
Reading comprehension	↔	↓	↑	↑	↔	↑
Silent reading comprehension	↑	↑	↑	↑	↑	↑
Listening comprehension	↑	↑	↑	↑	↑	↑

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a down arrow (↓) indicates that the mean score for students at endline was statistically significantly lower than the mean score for students at baseline; and a left-right arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the p<0.05 level. Differences were computed by urbanicity, language and grade level.

Table 17. Trends in Russian Grade 2 Student Reading Performance by Subtask and Region, between Baseline and Endline³³

EGRA Subtask	Endline relative to baseline - Russian Schools			
	Grade 2			
	DRS	Dushanbe	Khatlon-Bokhtar	Sughd
Letter name identification (CLNPM)	↔	↔	↔	↔
Initial sound identification	↑	↔	↑	↑
Familiar word reading (CFWPM)	↔	↔	↔	↔
Nonword reading (CNWPM)	↔	↔	↔	↑
ORF (equated CWPM)	↔	↔	↔	↑
Reading comprehension	↔	↑	↑	↑
Silent reading comprehension	↔	↑	↔	↑
Listening comprehension	↔	↔	↔	↔

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a down arrow (↓) indicates that the mean score for students at endline was statistically significantly lower than the mean score for students at baseline; and a left-right arrow (↔) indicates that the mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the p<0.05 level. Differences were computed by urbanicity, language and grade level.

Table 18. Trends in Russian Grade 4 Student Reading Performance by Subtask and Region, between Baseline and Endline³⁴

EGRA Subtask	Endline relative to baseline - Russian Schools			
	Grade 4			
	DRS	Dushanbe	Khatlon-Bokhtar	Sughd
Letter name identification (CLNPM)	↔	↔	↔	↔
Initial sound identification	↔	↔	↔	↔
Familiar word reading (CFWPM)	↔	↔	↔	↔
Nonword reading (CNWPM)	↔	↓	↑	↔
ORF (equated CWPM)	↔	↔	↔	↔
Reading comprehension	↔	↔	↔	↔
Silent reading comprehension	↔	↔	↑	↑
Listening comprehension	↔	↔	↑	↑

Note: An up arrow (↑) indicates that the mean score for students at endline was statistically significantly higher than the mean score for students at baseline; a down arrow (↓) indicates that the mean score for students at endline was statistically significantly lower than the mean score for students at baseline; and a left-right arrow (↔) indicates that the

³³ The Russian results disaggregated by region exclude the regions of GBAO and Khatlon-Kulob. The Russian sample did not include schools from GBAO because Russian-language schools were not present in this region at the time of sampling or data collection, and results from Khatlon-Kulob are not shown due to an extremely small sample size (1 school, 10 students).

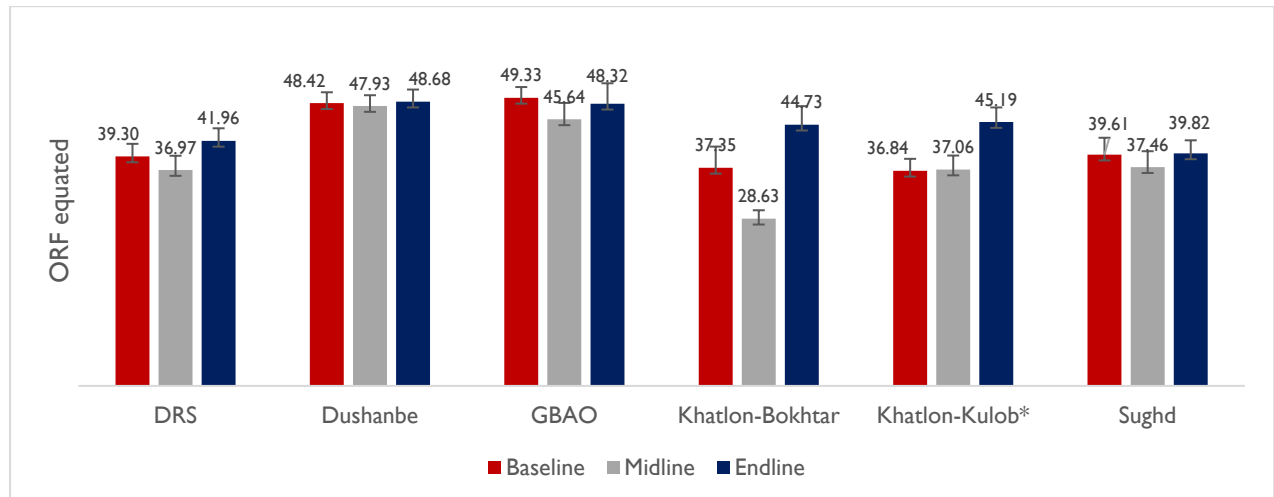
³⁴ The Russian results disaggregated by region exclude the regions of GBAO and Khatlon-Kulob. The Russian sample did not include schools from GBAO because Russian-language schools were not present in this region at the time of sampling or data collection, and results from Khatlon-Kulob are not shown due to an extremely small sample size (1 school, 10 students).

mean score for students in both time points was comparable. For all comparisons, statistically significant differences are reported at the $p < 0.05$ level. Differences were computed by urbanicity, language and grade level.

TAJIK GRADE 2 RESULTS BY REGION

Trends in Tajik grade 2 ORF performance varied by region across time points, as shown in Figure 17, with most regions showing gains of different magnitude between baseline and endline. The difference was only significant for Khatlon-Kulob. Overall, scores showed a more even distribution across regions at endline, corroborating the finding that gains were mostly driven by the increase among students in certain previously lower-performing regions.

Figure 17. Trends in Tajik Grade 2 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF)



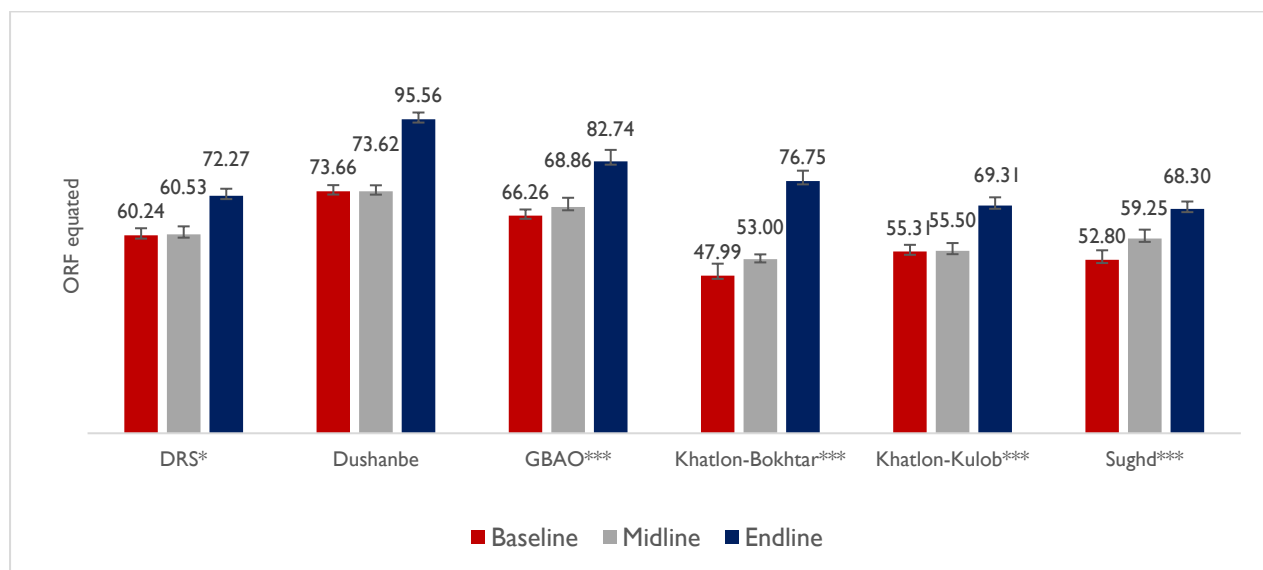
Note: Three asterisks (***) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between urban and rural students was not statistically significant.

TAJIK GRADE 4 RESULTS BY REGION

Regional trends in Tajik grade 4 ORF performance were more consistent than those for Tajik grade 2 across time points, as shown in Figure 18, with all regions showing important gains between baseline and endline. The differences were statistically significant in all regions but Dushanbe.³⁵ Performance appeared more balanced across regions at endline.

³⁵ While the differences appear large in Dushanbe, the standard error was correspondingly large.

Figure 18. Trends in Tajik Grade 4 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF CWPM)

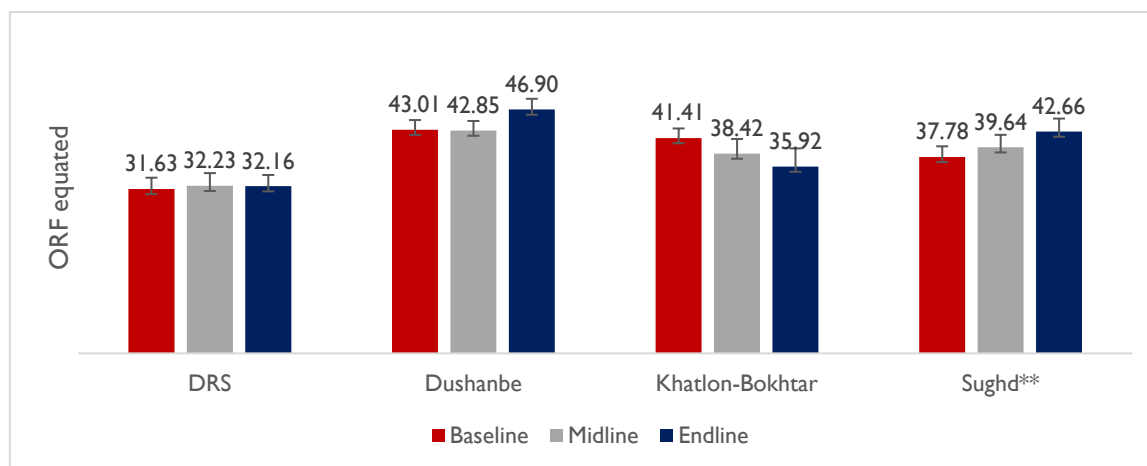


Note: Three asterisks (***) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between urban and rural students was not statistically significant.

RUSSIAN GRADE 2 RESULTS BY REGION³⁶

Regional trends in Russian grade 2 ORF performance varied across time points, as shown in Figure 19. Statistically significant gains were observed in Sughd, while performance changes in the other regions were not statistically significant.

Figure 19. Trends in Russian Grade 2 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF)³⁷



Note: Three asterisks (***) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.05$. No asterisks indicates that the

³⁶ The Russian results disaggregated by region exclude the regions of GBAO and Khatlon-Kulob. The Russian sample did not include schools from GBAO because Russian-language schools were not present in this region at the time of sampling or data collection, and results from Khatlon-Kulob are not shown due to an extremely small sample size (1 school, 10 students).

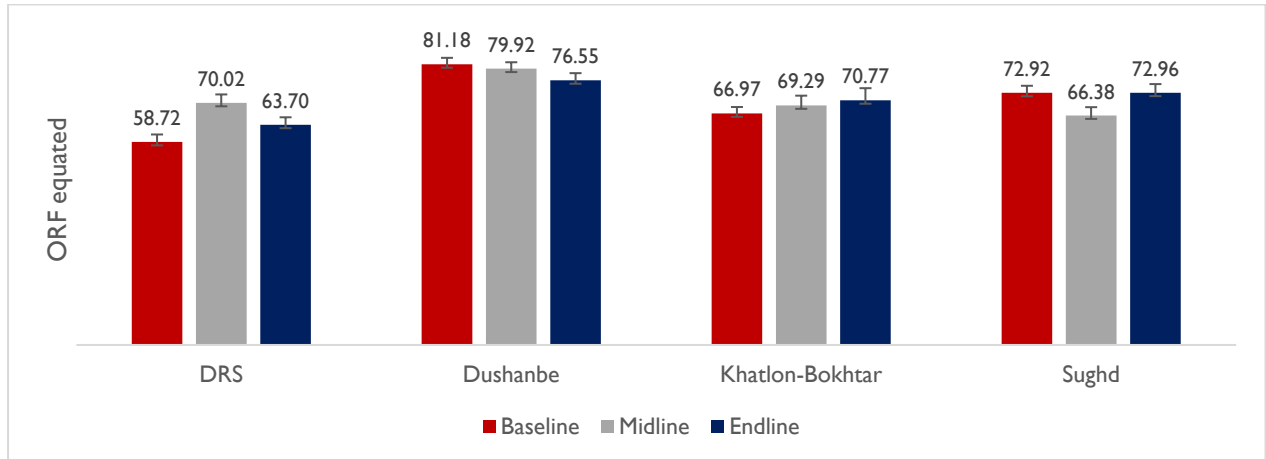
³⁷ Ibid.

difference between urban and rural students was not statistically significant.

RUSSIAN GRADE 4 RESULTS BY REGION³⁸

Figure 20 shows that regional trends in Russian grade 4 ORF were stable across time, with no region showing statistical gains between baseline and endline.

Figure 20. Trends in Russian Grade 4 Reading Performance by Time Point and Region: Oral reading fluency (equated ORF)³⁹



Note: Three asterisks (***) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between baseline and endline students for the corresponding region that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between urban and rural students was not statistically significant.

Evaluation Question 2: Which RWM geographic sub-groups require supplementary attention, and what kind of supplementary attention is required?

To respond to the second evaluation question, mean scores in ORF were analyzed by region, urbanicity, and sex, as displayed in Table 19 through Table 22. While at-risk groups may be identified in different ways, a proficiency approach (i.e., an approach based on learning outcomes) was utilized in this case, with low performing subgroups assumed to be at higher risk. Highlighted cells show subgroups whose average performance was below the corresponding ORF standard.

TAJIK GRADE 2

As shown in Table 19, five Tajik 2 subgroups scored below the ORF standard and thus require supplementary attention—DRS urban boys, Sughd urban boys, GBAO urban girls, DRS rural boys, and Sughd rural boys. Among these subgroups, Sughd rural boys require the most immediate attention.

³⁸ Ibid.

³⁹ The Russian results disaggregated by region exclude the regions of GBAO and Khatlon-Kulob. The Russian sample did not include schools from GBAO because Russian-language schools were not present in this region at the time of sampling or data collection, and results from Khatlon-Kulob are not shown due to an extremely small sample size (1 school, 10 students).

Table 19. Subgroup performance for Oral Reading Fluency – Tajik Grade 2

Region	Urban		Rural	
	Boy	Girl	Boy	Girl
Khatlon-Bokhtar	51.55	49.51	41.38	42.66
Dushanbe	47.47	49.83		
Khatlon-Kulob	46.88	39.67	47.42	44.15
DRS	35.25	45.16	36.69	47.18
Sughd	37.69	58.87	33.05	41.38
GBAO	43.75	35.38	47.26	57.76

Note: Highlighted cells show means that are below the ORF benchmark for the corresponding grade level and language. In addition, the deeper the red, the lower the score.

Analyses of factors associated with lower performance of Tajik grade 2 rural boys in Sughd revealed two types of findings. First, as displayed in Figure 21, these boys attended preschool and have mothers who can read, which are factors normally associated with stronger performance. These results do not mean that attending preschool or having mothers who can read cause lower performance, but rather the contrary—despite having attended preschool and having mothers who can read, these students are at risk. In particular, these boys live in Uzbek-speaking homes and receive homework with less frequency than their peers, suggesting that appropriate interventions to improve their proficiency should focus on homework practices and Tajik linguistic support for students who speak Uzbek at home.

Figure 21. Factors associated with group with lowest scores: Tajik grade 2

Language at Home	Family Literacy
Speak Uzbek	Mothers can read
Preschool	Homework/ Books
Did attend preschool	Get homework with less frequency

TAJIK GRADE 4

As displayed in Table 20, fifteen Tajik grade 4 subgroups scored below the ORF standard and thus require supplementary attention—rural students in Khatlon-Bokhtar (girls and boys), all subgroups in DRS and Sughd, boys (urban and rural) and rural girls in Sughd, and boys (urban and rural) in GBAO. All rural students except for girls in GBAO require supplementary attention. Among these subgroups, Sughd rural boys require the most immediate attention, as was the case with Tajik grade 2 students.

Table 20. Subgroup performance for Oral Reading Fluency – Tajik Grade 4

Region	Urban		Rural	
	Boy	Girl	Boy	Girl
Khatlon-Bokhtar	83.68	86.54	71.70	72.55
Dushanbe	102.56	89.11		
Khatlon-Kulob	68.17	76.27	66.34	70.52
DRS	67.22	68.06	72.50	72.22
Sughd	71.53	85.30	56.27	74.30
GBAO	69.44	108.25	76.09	82.71

Note: Highlighted cells show means that are below the ORF benchmark for the corresponding grade level and language. In addition, the deeper the red, the lower the score.

As with grade 2 above, analyses of factors associated with lower performance of Tajik grade 4 rural boys in Sughd revealed two types of findings. First, as displayed in Figure 22, these boys are of higher SES and have mothers who can read, which are factors normally associated with stronger performance. These results do not mean that high SES or having mothers who can read cause lower performance, but rather the contrary—despite higher levels of SES and having mothers who can read, these students are at risk. As with the grade 2 cohort, these boys also live in Uzbek-speaking homes and receive homework with less frequency than their peers. Moreover, struggling Tajik grade 4 boys also have brothers who cannot read. As with their counterparts in grade 2, Tajik grade 4 rural boys would likely benefit from interventions that support more frequent homework and Tajik linguistic support for students who speak Uzbek at home.

Figure 22. Factors associated with group with lowest scores: Tajik grade 4

Language at Home	Family Literacy
Speak Uzbek	Mothers can read Brothers cannot read
Socioeconomic Status	Homework/ Books
Higher socio-economic status	Get homework with less frequency

RUSSIAN GRADE 2

As displayed in Table 21, seven Russian grade 2 subgroups scored below the ORF standard and thus require supplementary attention—urban students in Khatlon-Bokhtar (girls and boys), urban girls in Khatlon-Kulob, all subgroups in DRS, and rural students in Sughd (girls and boys). Among these subgroups, DRS rural girls require the most immediate attention.

Table 21. Subgroup performance for Oral Reading Fluency – Russian Grade 2

Region	Urban		Rural	
	Boy	Girl	Boy	Girl
Khatlon-Bokhtar	35.17	37.53		
Dushanbe	42.56	52.17		
Khatlon-Kulob	41.90	31.02		
DRS	34.77	31.45	30.21	25.54
Sughd	40.62	47.22	37.55	38.24

Note: Highlighted cells show means that are below the ORF benchmark for the corresponding grade level and language. In addition, the deeper the red, the lower the score.

Analyses of factors associated with lower performance of Russian grade 2 rural girls in DRS revealed two types of findings. First, as displayed in Figure 23, these girls have reading books at home, which is a factor normally associated with stronger performance. However, these girls also live in Uzbek-speaking homes, have fathers who cannot read, did not attend preschool, and have no reading books at school to take home.⁴⁰ As such, Russian grade 2 rural girls in DRS would likely benefit from interventions that provide reading books at school to take home, promote preschool attendance, and provide Tajik linguistic support for girls who speak Uzbek at home.

Figure 23. Factors associated with group with lowest scores: Russian grade 2

Language at Home	Family Literacy
Speak Uzbek	Fathers cannot read
Preschool	Homework / Books
Did not attend preschool	Have reading books at home No reading books at school to take home

RUSSIAN GRADE 4

As displayed in Table 22, eleven Russian grade 4 subgroups scored below the ORF standard and thus require supplementary attention—urban students in Khatlon-Kulob (girls and boys), urban girls in Dushanbe, and all subgroups in DRS and Sughd. Among these subgroups, Sughd rural boys require the most immediate attention.

Table 22. Subgroup performance for Oral Reading Fluency – Russian Grade 4⁴¹

Region	Urban		Rural	
	Boy	Girl	Boy	Girl
Khatlon-Kulob	69.14	73.10		
Dushanbe	78.83	73.55		
DRS	59.70	65.80	75.20	62.81
Sughd	72.05	77.63	55.82	72.70

Note: Highlighted cells show means that are below the ORF benchmark for the corresponding grade level and language. In addition, the stronger the color, the lower the score.

Analyses of factors associated with lower performance of Russian grade 4 rural boys in Sughd revealed two types of findings. First, as displayed in Figure 24, these boys are of higher SES, have mothers who can read, and have reading books at school to take home, which are factors normally associated with stronger performance. These results do not mean that higher SES, having mothers who can read, or having books at school to take home can cause lower performance, but rather the contrary—despite these contextual characteristics, these students are at risk. In particular, as with their Russian grade 2 female counterparts in DRS, these boys also live in homes where Russian is not spoken, suggesting that they, too, would benefit from Russian language support tailored to the needs of non-Russian speakers, who potentially come from linguistic minorities.

⁴⁰ 33.6% of Russian Grade 2 students did not list fathers as members of their household who can read.

⁴¹ Khatlon-Bokhtar was removed due to small sample size.

Figure 24. Factors associated with group with lowest scores: Russian grade 4

Language at Home	Family Literacy
Do not speak Russian	Mothers can read
Socioeconomic Status	Homework / Books
Higher socio-economic status	Reading books at school to take home

Evaluation Question 3: Do materials provided by RWM or teacher instructional practices supported by RWM serve as predictors of Tajik and Russian reading outcomes of students in Grade 2 and Grade 4 in schools supported by RWM? If yes, which materials or practices are predictors?

This section answers the third evaluation question by reporting results of the analyses measuring whether RWM-specific materials and teachers’ reading instructional practices serve as predictors of student reading performance as measured by ORF. To get a project-level picture, results are presented across languages and grade levels. Results are presented by topic. Each topic discussion includes a table with coefficients, which represents the associations between the variables that represent RWM materials or instructional practices, and ORF scores. While all relevant items and composites were analyzed, only those that showed a significant relationship to ORF were kept in the corresponding tables. Coefficients indicate the average change in ORF that would take place by increasing each of the variables by one value. For example, if a binary variable is associated to a coefficient of 5.00 it means that whenever that variable takes the value of 1, students show an increase in ORF performance by 5.00 CWPM. Similarly, if a table shows a coefficient of -4.00 for a categorical variable, it means that moving up one category is associated with a decrease in performance by 4.00 CWPM.

ASSESSMENT AND EVALUATION PRACTICES

Teachers’ self-reported assessment practices, including their evaluation methods and use of assessment results, were among the most relevant variables associated with ORF, as displayed in Table 23. Most of these practices were associated with improvements in ORF scores, ranging from an additional 4.65 to 12.97 CWPM depending on the language and grade. In other words, using different types of evaluation methods, as well as using assessment results for multiple purposes, tended to correlate positively with students’ reading achievement. Results for Russian grade 4 stood out as the only exception, as using oral evaluations was associated with lower ORF scores.

Table 23. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language

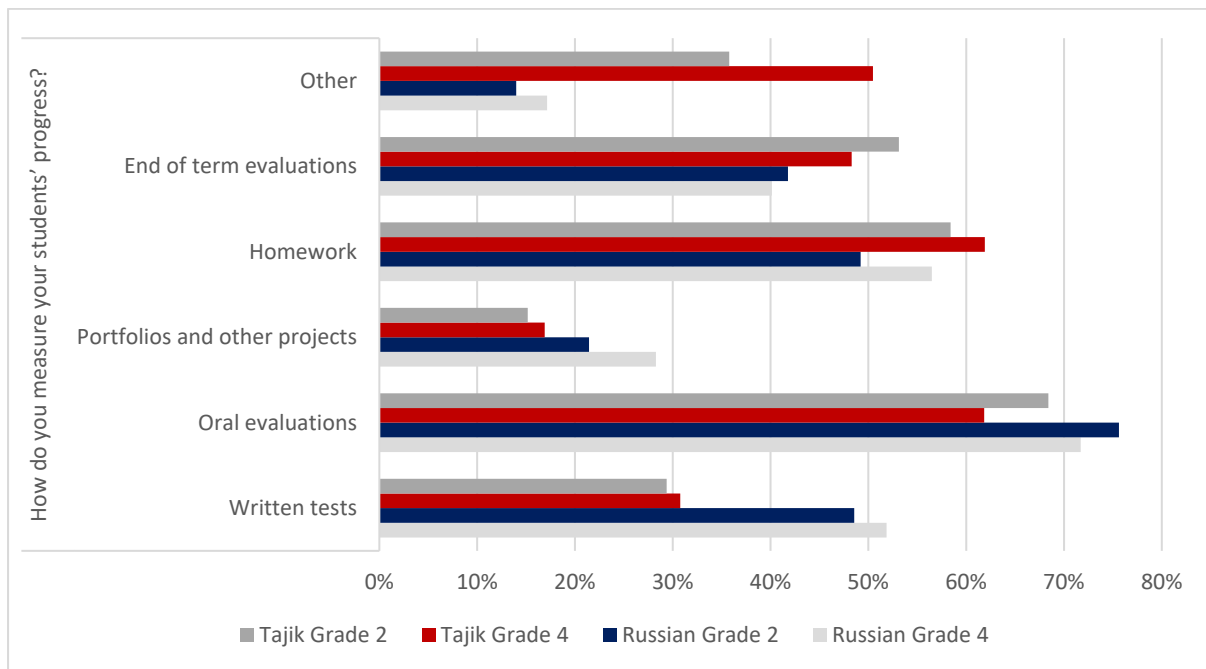
Materials or Instructional Practices		Associated increase/decrease in ORF score (CWPM)			
		Tajik		Russian	
		Grade 2	Grade 4	Grade 2	Grade 4
Assessment and evaluation practices	Teacher measures students' progress using written tests (I=Yes)	8.83			8.32
	Teacher measures students' progress using oral evaluations (I=Yes)		12.97		-9.11
	Teacher measures students' progress using portfolios and other projects (I=Yes)	7.48		9.98	
	Teacher measures students' progress using homework (I=Yes)				7.23
	Teacher measures students' progress using end of term evaluations (I=Yes)	6.01			

Use results of students' oral and written assessments to grade students (I = Yes)		8.47		5.74
Use results of students' oral and written assessments to evaluate students' understanding of subject matter (I = Yes)	4.72	12.64	7.93	
Use results of students' oral and written assessments to adapt teaching to better suit students' needs (I = Yes)	4.65			10.22

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. In particular, green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

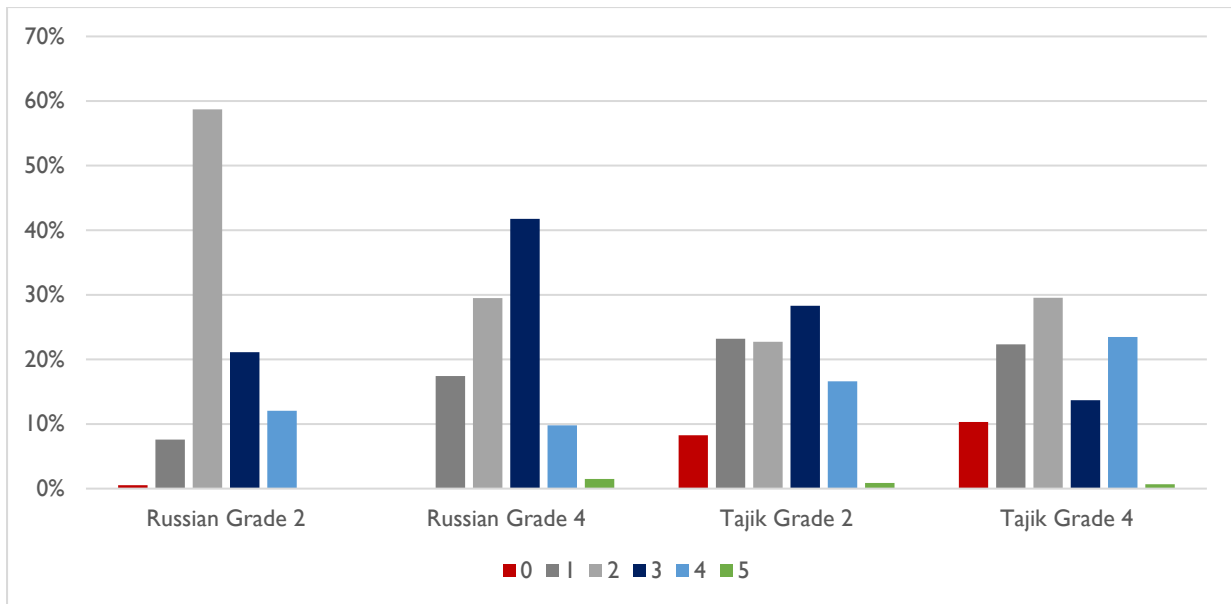
Figure 25 shows a variety of evaluation methods used by teachers based on their self-reported responses, with a preference for more traditional forms of evaluation. Overall, teachers reported oral evaluation as their most preferred method of evaluation to measure students' progress. Less preferred methods of evaluation included end-of-term evaluations and written tests. These two findings might reflect RWM's emphasis on formative assessment. Fewer teachers reported using portfolios and other projects to evaluate students' performance and progress.

Figure 25. Teachers Preferred Evaluation Methods, by Language and Grade Level



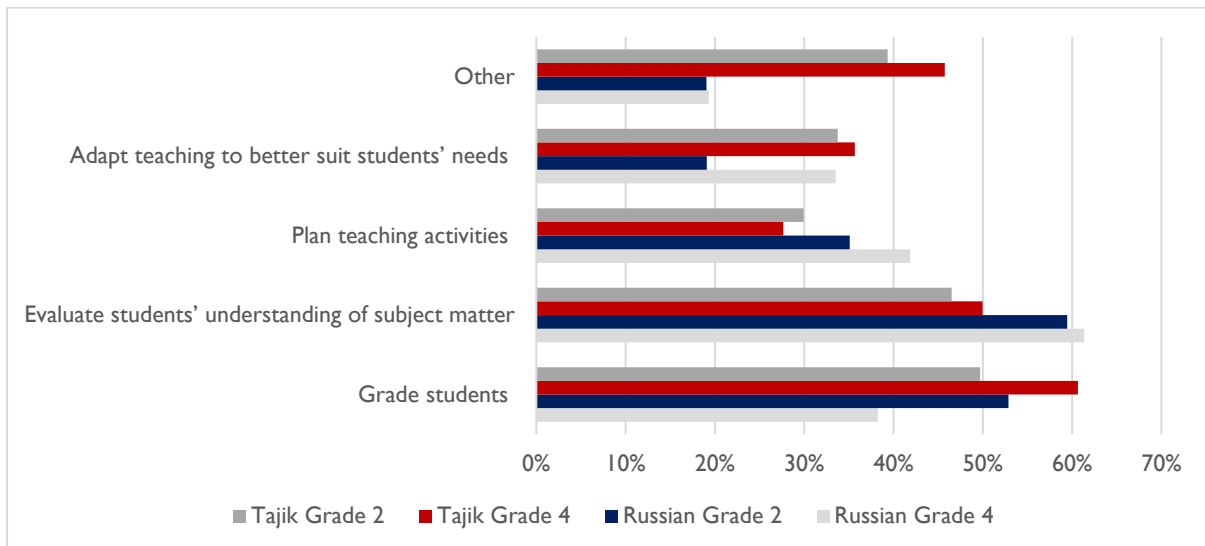
Further analysis revealed that most teachers used either two or three evaluation methods to measure student progress, as illustrated in Figure 26. Extremely few teachers reported using all types of evaluation methods, suggesting that there is room for further diversifying the use of assessments. The category of "other" was excluded because teachers did not specify the additional methods they used.

Figure 26. Evaluation Methods used by Teachers (%), by Language and Grade Level⁴²



Overall, most teachers reported using assessment results either to evaluate students’ understanding or grade students, as displayed in Figure 27. Fewer teachers reported using assessment results to plan teaching activities or adapting teaching to better suit students’ needs. While teachers reported using assessment results in a variety of ways, results clearly highlight room for increasing formative uses of assessments, such as planning activities or adapting teaching.

Figure 27. Use of Assessment Results, by Language and Grade Level



Further analysis revealed that most teachers reported using assessment results in one or two ways, as illustrated in Figure 28. Although results varied by language and grade, teachers in Russian schools seemed to use assessment in more diverse ways. Extremely few teachers reported using assessment results in all possible ways, suggesting room for improvement is possible. The category of “other” (selected by 19.06 to 45.74 percent of the teachers, depending on the language and grade level) was excluded because teachers did not specify the additional ways they used assessment results. See

⁴² Excluding the category “other.”

Figure 29. Change in Use of Assessment Results Across Time, Tajik and Figure 30. Evolution in Use of Assessment Results Across Time, Russian.

Figure 28. Total Uses of Assessment Results, by Language and Grade Level⁴³

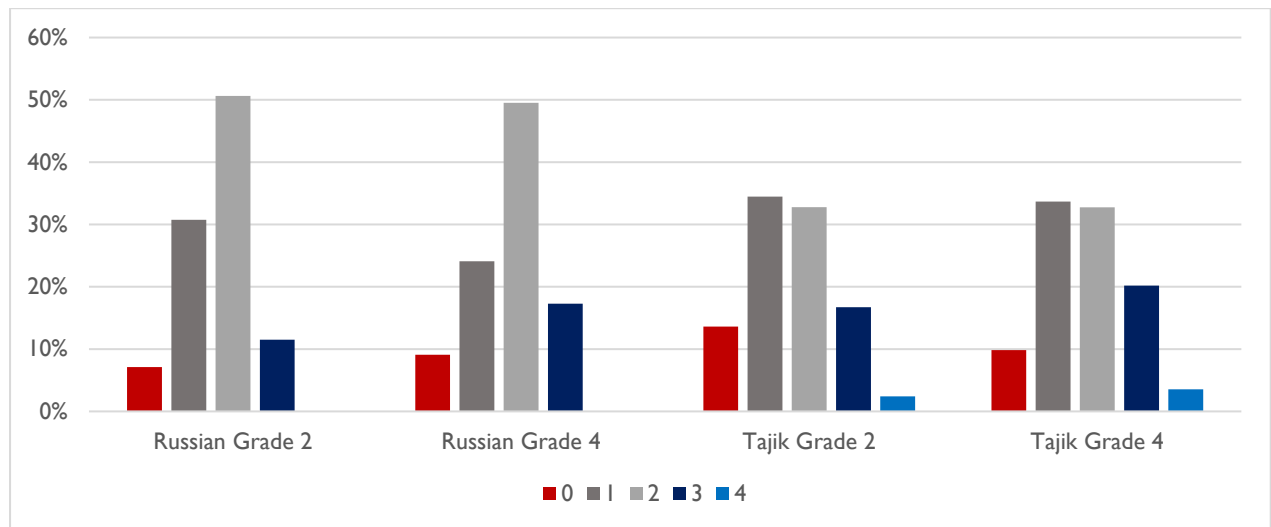
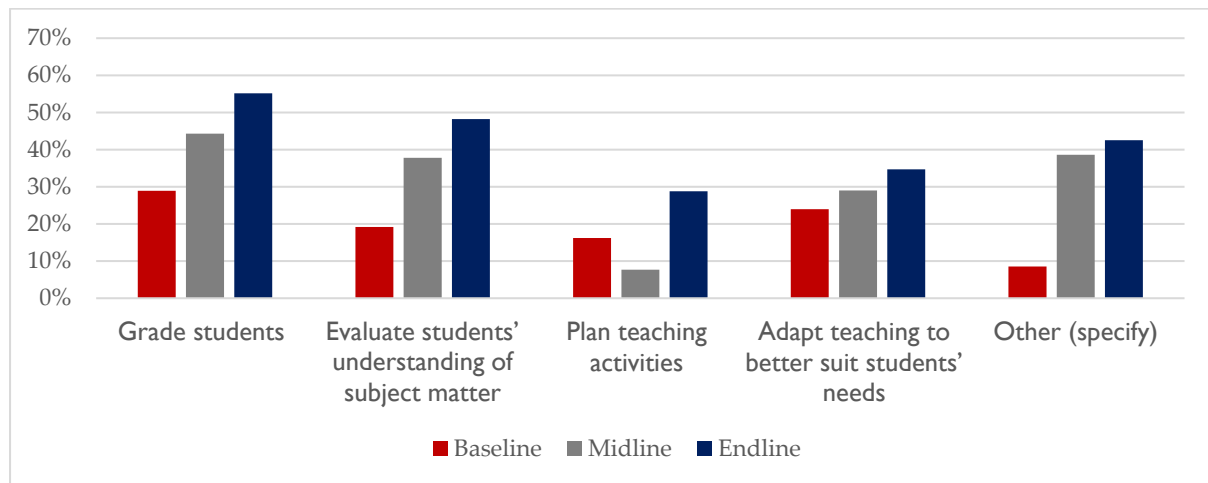


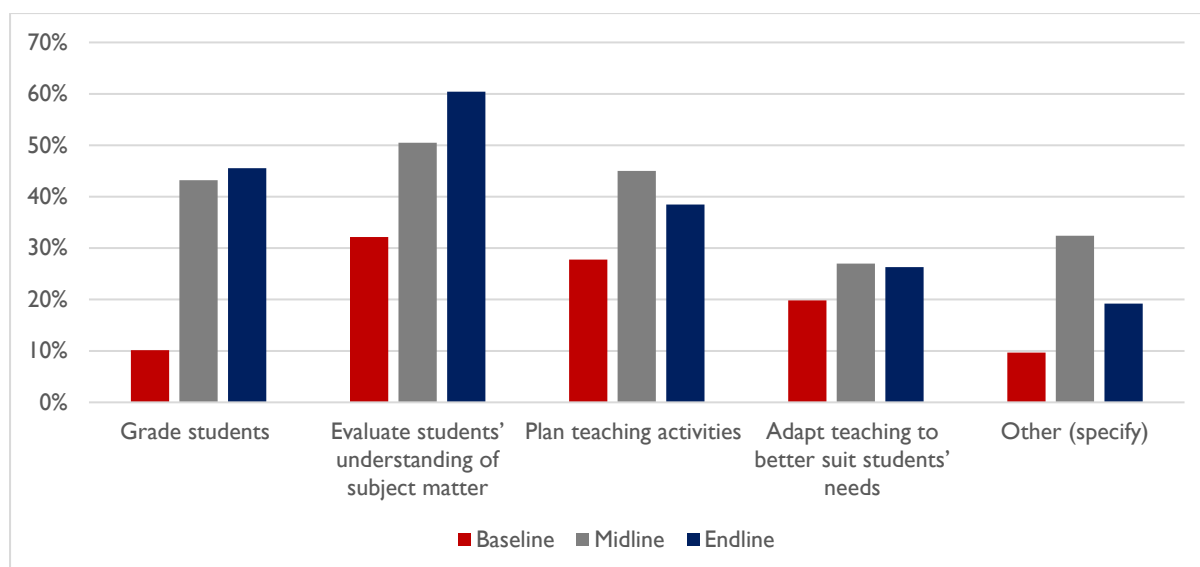
Figure 29 and Figure 30 show how teachers’ use of assessment has changed over time. Overall, Tajik teachers having increasingly used assessments in diverse ways over time. The trend is similar for Russian teachers, but less obvious, with declines from midline to endline in planning teaching activities and adapting teaching to better suit students’ needs.

Figure 29. Change in Use of Assessment Results Across Time, Tajik



⁴³ Excluding the category “other.”

Figure 30. Change in Use of Assessment Results Across Time, Russian



With few exceptions, more evaluation methods and more uses of assessment results were consistently associated with an increase in ORF, as displayed in Table 24. This important result was detected when replicating the analysis between assessment and evaluation methods and ORF, but instead including the number of evaluation methods used, as well as the number of ways teachers used assessment results. For example, using one more evaluation method was associated with an increase in performance between 3.40 CWPM and 3.96 CWPM, depending on the language and grade level. Similarly, using assessment and evaluation results in one more way was associated with an increase in performance between 3.48 CWPM and 8.46 CWPM, depending on the language and grade level.

Table 24. Associations between Total Evaluation Methods and Uses of Assessment Results and ORF by Grade and Language

Assessment and evaluation practices - Composites	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
Total evaluation methods used (0-5)	3.96		3.68	3.40
Diverse use of assessment and evaluation results (0-4)	4.25	8.46	3.48	

INSTRUCTIONAL PRACTICES

The quality and intensity of teaching were associated with improvements in ORF scores, ranging from an additional 1.72 and 10.24 CWPM, as displayed in Table 25. In other words, good teaching correlated positively with students' reading achievement. The only negative association was found in the Russian grade 4 results, for which making connections between lessons and other content knowledge, or students' daily life, was associated with a decrease of 7.30 CWPM in ORF. In addition, teaching practices had a higher impact on Tajik schools, particularly for grade 2 students, than Russian schools.

Table 25. Associations between RWM Materials or Observed Teaching Practices and Oral Reading Fluency by Grade and Language

		Associated increase/decrease in ORF score (CWPM)			
Materials or Instructional Practices		Tajik		Russian	
		Grade 2	Grade 4	Grade 2	Grade 4
Instructional Practices	The teacher explicitly articulates the objectives of the lesson and relates classroom activities to the objectives	4.92		5.95	
	Teacher's explanation of content is clear	6.23	14.92		
	Teacher makes connections in the lesson that relate to other content knowledge or students' daily lives	5.33			-7.30
	Lesson plans are developed by teacher him/herself (1=Yes)				10.24
	Number of prereading activities (0-7)	2.28	3.67		1.83
	Number of while-reading activities (0-5)	2.88			
	Number of reading skills taught (0-5) ⁴⁴	3.34			
	Number of post-reading activities (0-6)	1.72	4.08		
	Number of teaching strategies used (0-4)	3.18	6.69		

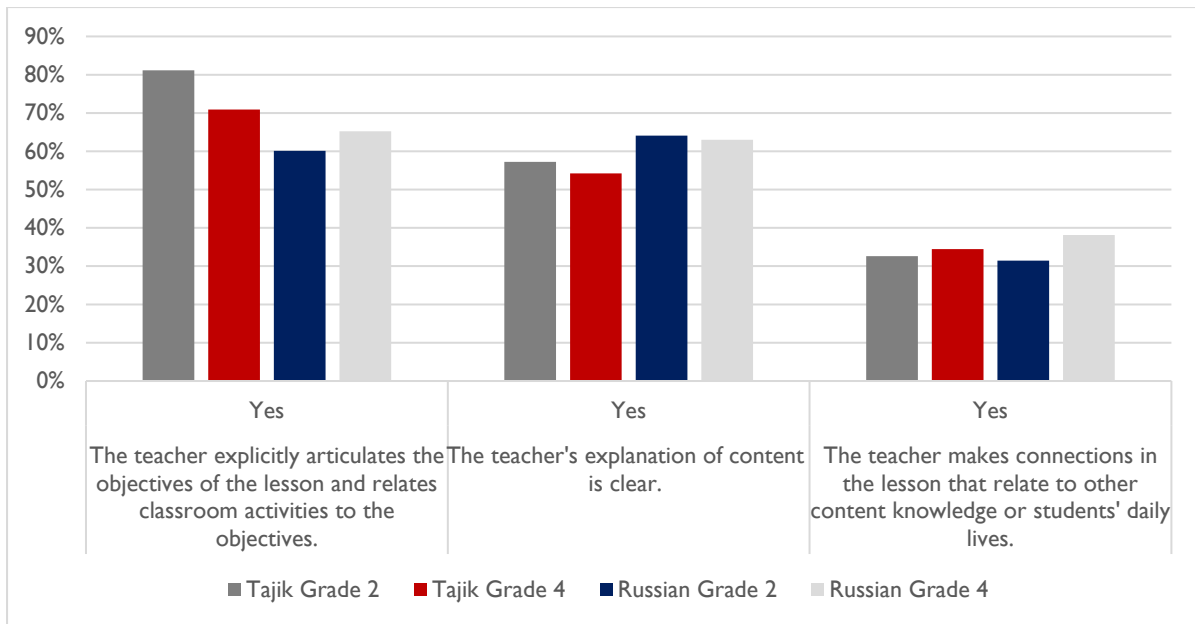
Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. In particular, green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

Data on teachers' instructional practices were collected from a classroom observation tool, including the good teaching practices that teachers employed, as well as the activities they did before, while, and after their students read in class. These data provide further insight into the observed relationships and teaching practices more generally, as displayed in Figures 29 to 35 in this section. External enumerators who were knowledgeable with RWM materials and techniques were trained to administer the classroom observation. Inter-rater reliability scores are presented in Annex C.

The good teaching practices employed by teachers were generally similar across languages and grades, as illustrated in Figure 31. Most teachers explicitly articulated objectives and related classroom activities to these objectives. Enumerators also observed that a majority of teachers explained content clearly, but that fewer teachers made connections between the lesson and other content knowledge and/or students' daily lives.

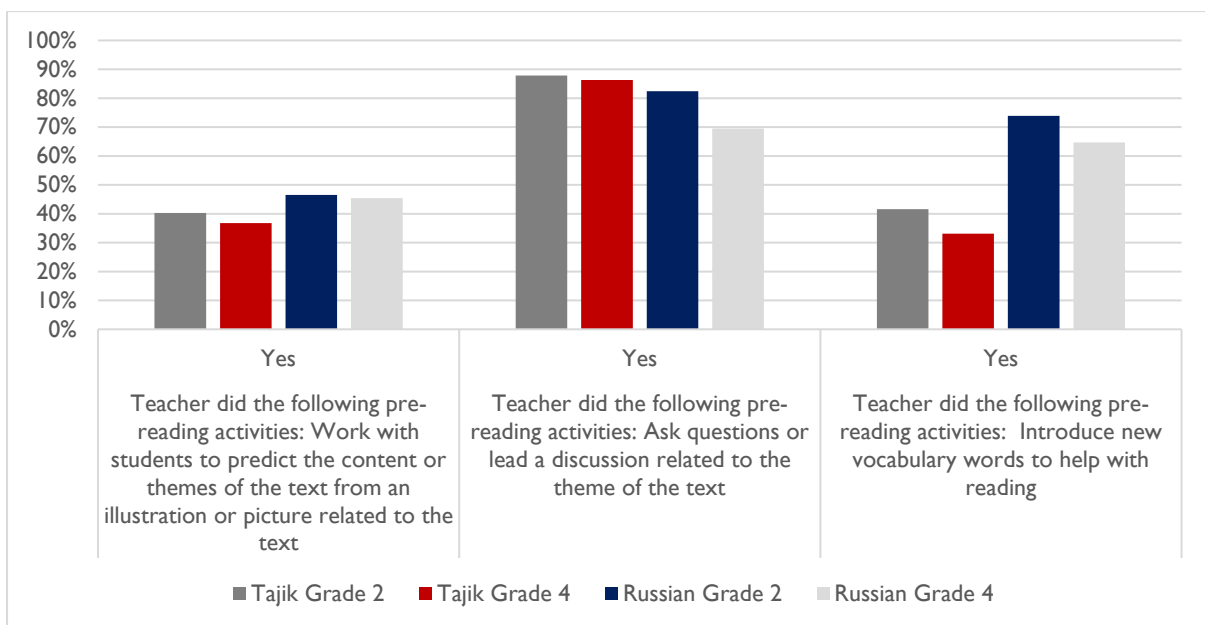
⁴⁴ While additional reading skills exist, this lesson observation tool looked for five specific skills identified as priorities by RWM: a) looking for text conventions, b) phonemic consciousness, c) fluent reading, d) vocabulary, and e) reading comprehension.

Figure 31. Overall Good Teaching Practices, by Language and Grade Level



More variation was evident across languages and grades in teachers' pre-reading activities, as displayed in Figure 32. Faced with different choices, most teachers preferred to ask questions or lead a discussion related to the theme of the text. Fewer teachers worked with students to predict the content or themes of the text from an illustration or picture or introduced new vocabulary words than asked questions related to the theme of the text.

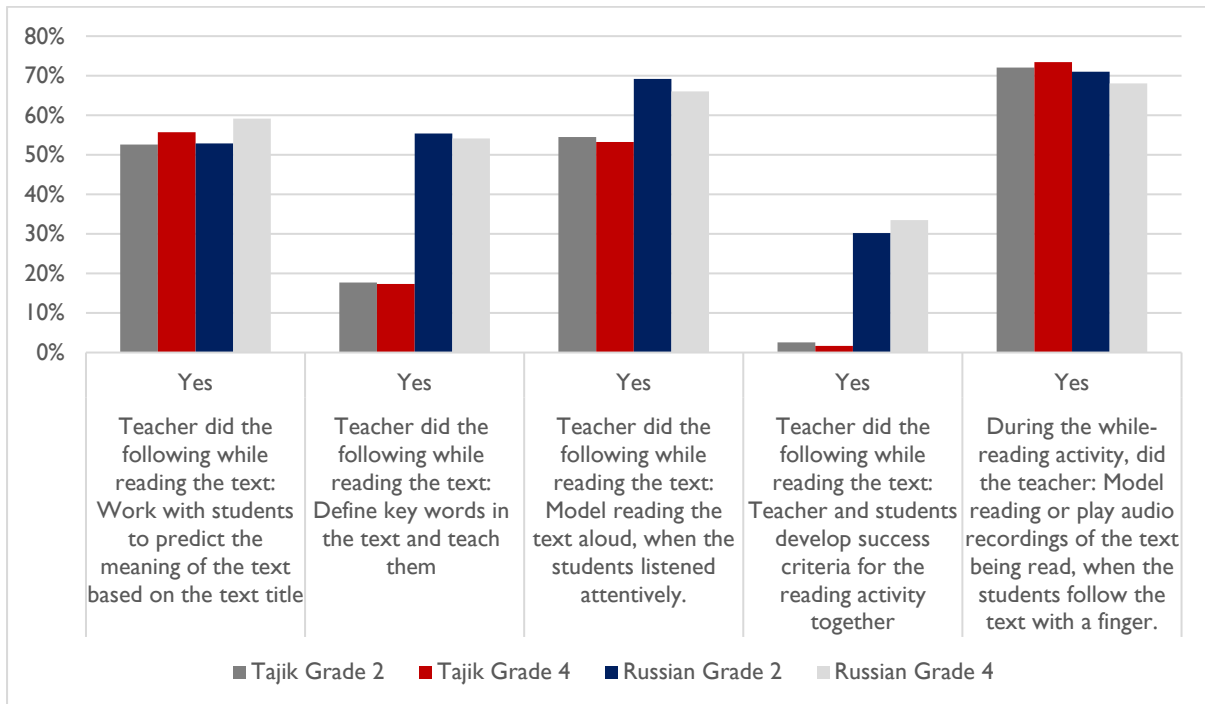
Figure 32. Teachers' Prereading Activities, by Language and Grade Level



Although some teachers' activities while reading did not vary across languages and grades, some important variations were found, as shown in Figure 33. Most teachers in both Tajik and Russian classrooms worked with students to predict the meaning of the text based on the title and modeled reading or played audio recordings of the text. Stark differences were observed between Russian and Tajik classrooms in other activities, however. A much higher proportion of teachers in Russian schools defined key words in the text and taught them than those in Tajik classrooms. In addition, a

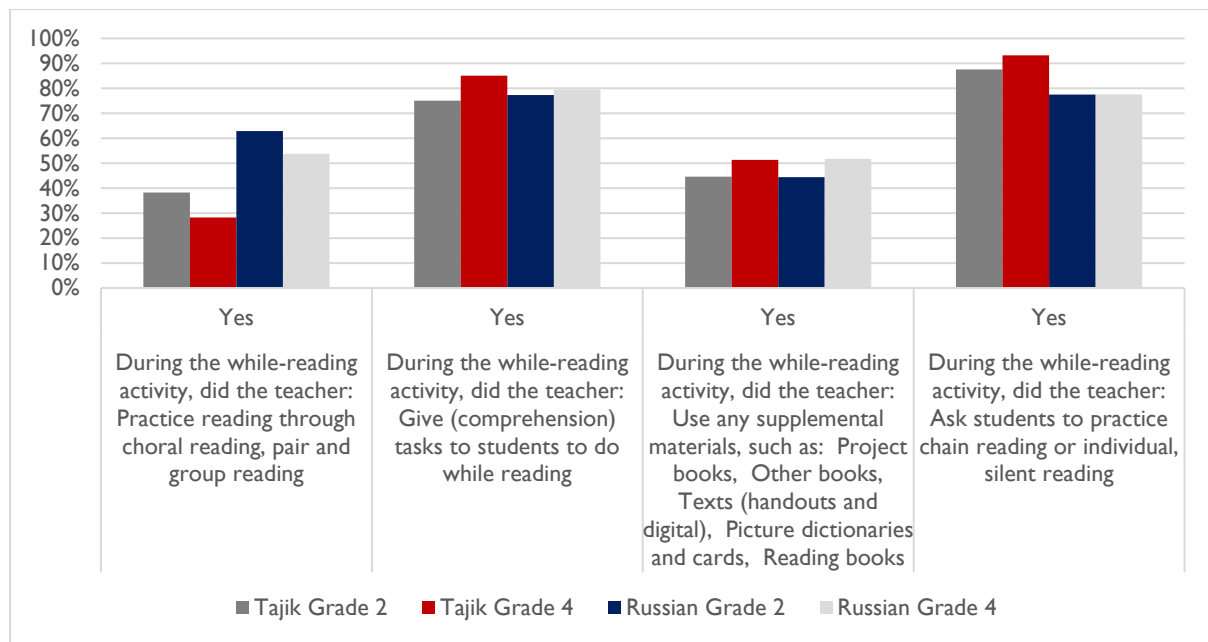
higher percentage of teachers in Russian classrooms modeled reading the text aloud than their Tajik counterparts. While some teachers in Russian classrooms developed success criteria for the reading activity, extremely few teachers in Tajik classrooms did so. Overall, there is variation in preferred “while reading activities” among teachers. There is also room for increasing the use of specific practices such as developing success criteria for the reading activity.

Figure 33. Teachers’ While-Reading Activities – Part I, by Language and Grade Level



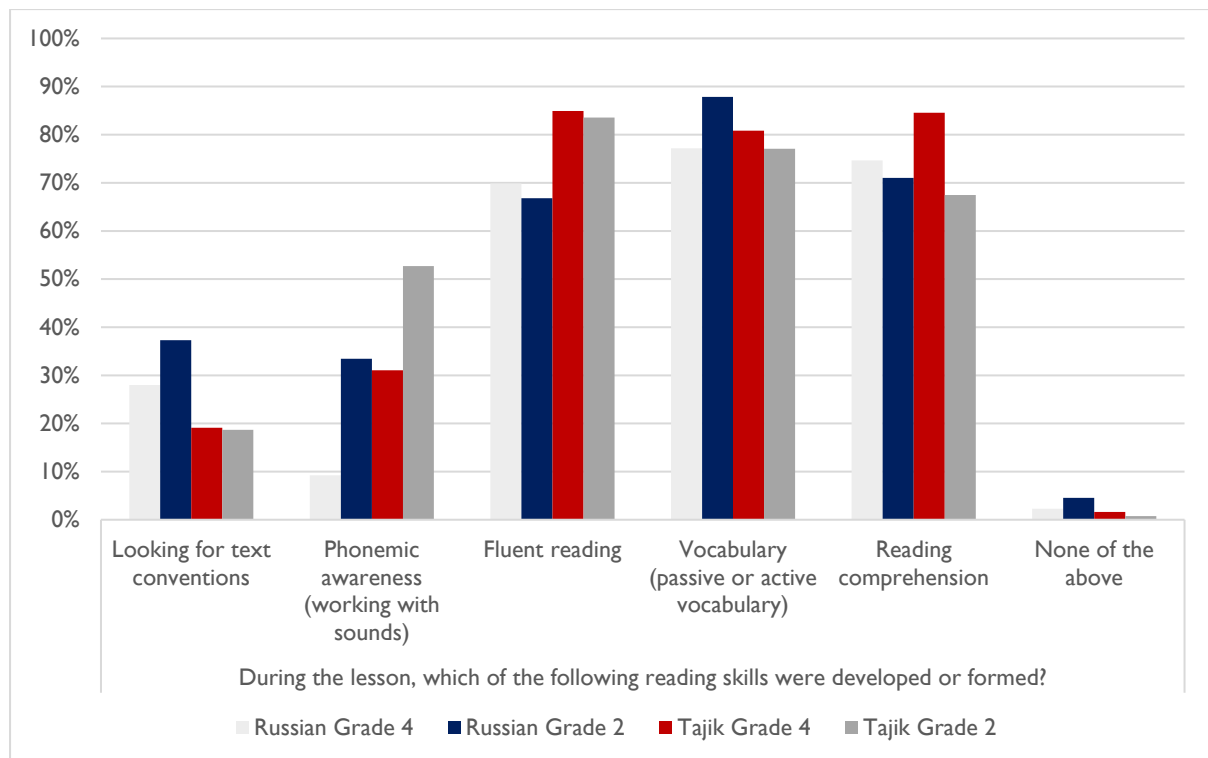
Enumerators noted less variation across languages and grades for other activities teachers did while reading, as displayed in Figure 34. While most teachers gave comprehension tasks to students or asked them to practice chain reading or silent reading, fewer teachers practiced reading through choral reading or used supplemental materials. It is possible that the observed emphasis on silent reading and comprehension drove most of the measured gains on the EGRA reading comprehension and silent reading comprehension subtasks.

Figure 34. Teachers' While-Reading Activities – Part II, by Language and Grade Level



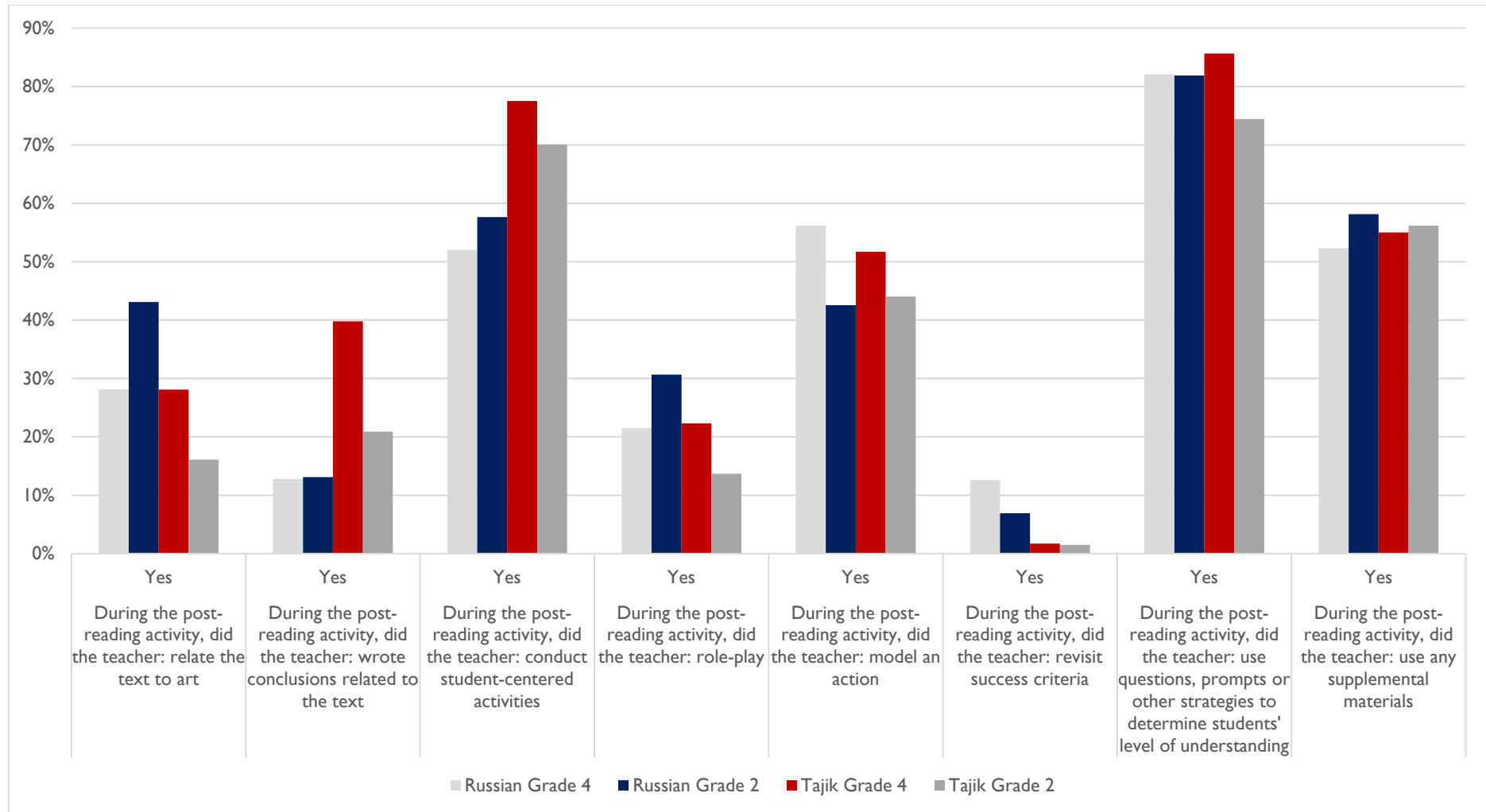
Enumerators observed an emphasis on higher-order skills being developed in classrooms, including reading comprehension, vocabulary, and fluent reading, as shown in Figure 35, rather than lower-order skills, including developing phonemic awareness or looking for text conventions. Students' high levels of literacy in Tajikistan, even in grade 2, likely accounted for this observed difference. Interestingly, no clear pattern by grade level was evident.

Figure 35. Reading Skills Developed, by Language and Grade Level



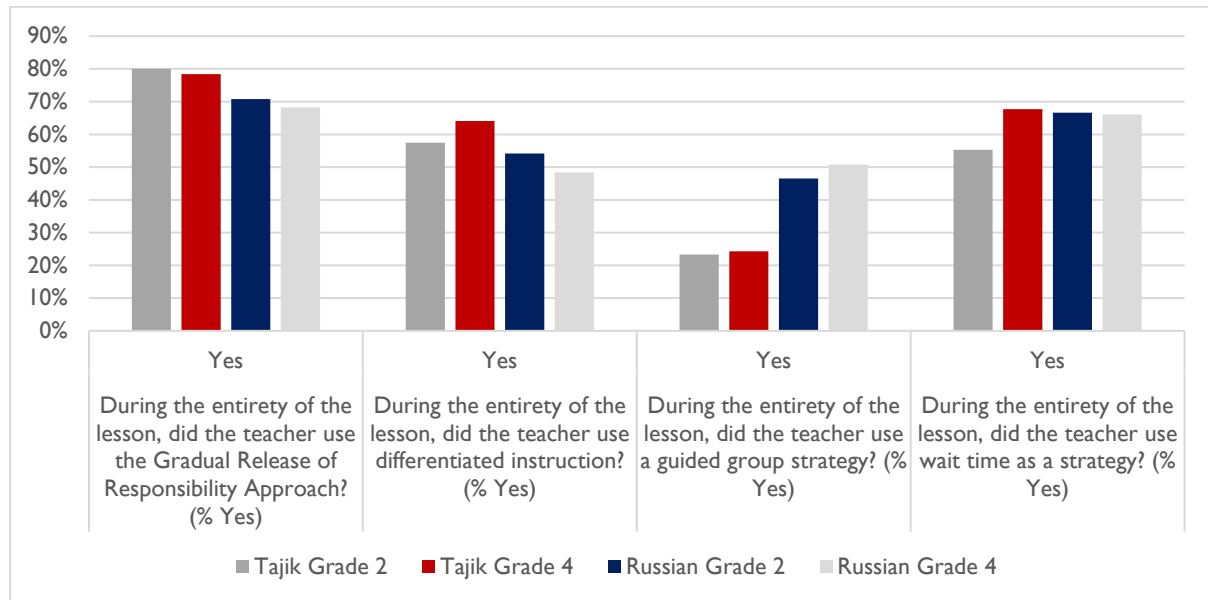
Some language trends were noted in teachers' post-reading activities, as illustrated in Figure 36, but no clear grade level trends were identified. The most notable results include that most teachers used questions or similar strategies to check students' level of understanding, as well as conducted student-centered activities, particularly in Tajik schools. By contrast, very few teachers revisited success criteria, especially in Tajik schools, which was consistent with the behaviors observed while reading.

Figure 36. Teachers' Post-Reading Activities, by Language and Grade Level



Teachers used a variety of strategies during instruction, as displayed in Figure 37. Teachers most frequently used the gradual release of responsibility approach. Most used wait time as a strategy, and slightly fewer used differentiated instruction. While approximately half of the teachers in Russian classrooms used guided group strategies, only about a quarter of the teachers in Tajik classrooms did so.

Figure 37. Teacher Strategies Used, by Language and Grade Level



CLASSROOM OBSERVATION LONGITUDINAL ANALYSIS

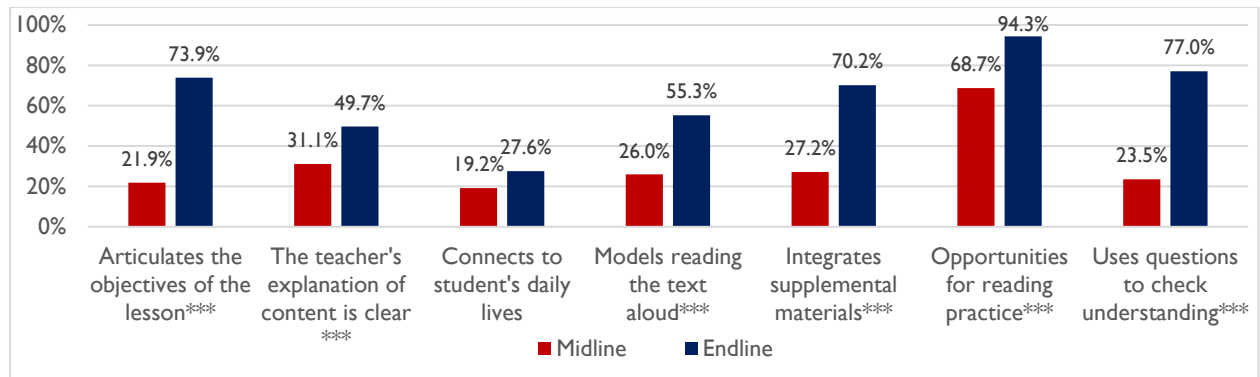
While the midline classroom observation tool was modified for the endline and the methodologies for observation were not the same⁴⁵, several elements were retained in order to conduct longitudinal comparisons on the application of certain RWM activities in the classroom. These items fall into two categories: 1) general teaching techniques, and 2) an emphasis on a specific reading skill during the lesson. Weighted results from this analysis are presented below, and a description of the items used for analysis can be found in Annex B.

TAJIK GRADE 2 TEACHING TECHNIQUES

Overall, the proportion of teachers using specific RWM teaching techniques increased. A significantly higher proportion of teachers at endline articulated the objectives of the lesson (73.9 percent) and clearly explained the lesson’s content (49.7 percent) compared with teachers at midline (21.9 percent and 31.1 percent, respectively). A significantly higher proportion of teachers also modeled reading out loud (55.3 percent at endline, compared with 26.0 percent at midline), integrated supplemental materials (70.2 percent to 27.2 percent), gave opportunities for reading practice (94.3 percent to 68.7 percent) and used questions to check understanding (77.0 percent to 23.5 percent). While a higher proportion of teachers at endline than midline connected lesson content to students’ knowledge and daily lives, the difference was not significant.

⁴⁵ At midline, observers scanned the classroom twice over a period of 15 minutes to observe and take notes, using the remaining time to write up their observations. In contrast, the endline tool was an observation for the entire duration of a class period.

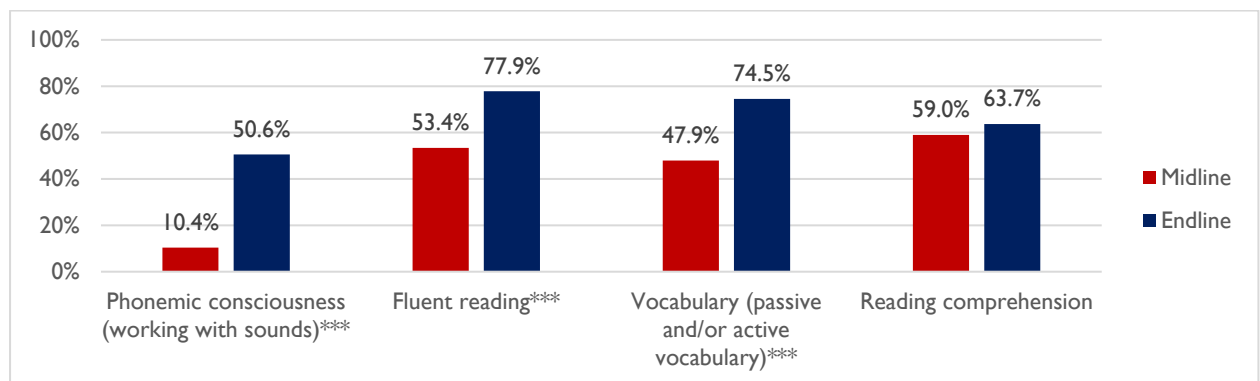
Figure 38. Percentage of Tajik Grade 2 Teachers Using Specific Teaching Techniques, Midline and Endline



TAJIK GRADE 2 READING SKILLS

A higher proportion of teachers at endline than midline emphasized the fundamental reading skills of phonemic consciousness (50.6 percent at endline, compared with 10.4 percent at midline), fluent reading (77.9 percent to 53.4 percent), vocabulary (74.5 percent to 47.9 percent), and reading comprehension (63.7 percent to 59.0 percent). The difference in proportions was significant for every skill except for reading comprehension.

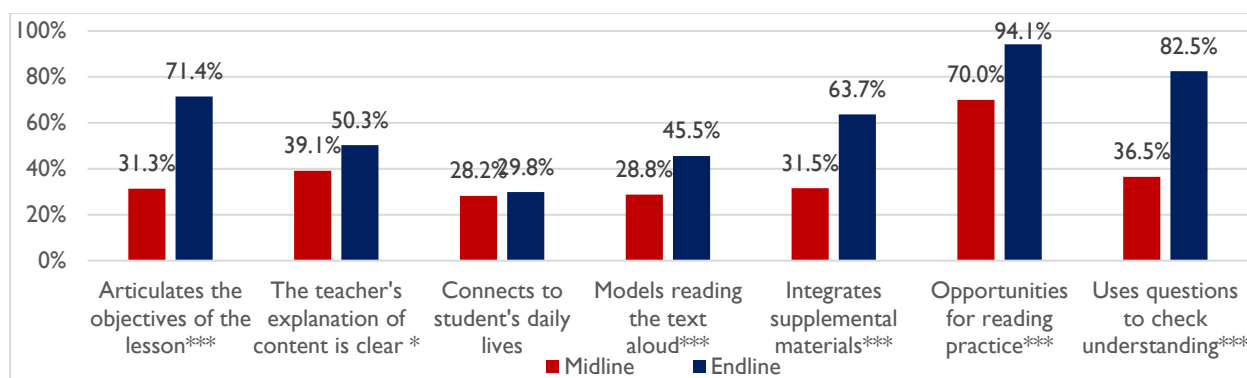
Figure 39. Percentage of Tajik Grade 2 Teachers Focusing on Specific Reading Skills, Midline and Endline



TAJIK GRADE 4 TEACHING TECHNIQUES

Overall, the proportion of teachers using specific RWM teaching techniques increased. A significantly higher proportion of teachers at endline articulated the objectives of the lesson (71.4 percent) and clearly explained the lesson's content (50.3 percent) than midline (31.3 percent and 39.1 percent, respectively). A significantly higher proportion of teachers also modeled reading out loud (45.5 percent at endline, compared with 28.8 percent at midline), integrated supplemental materials (63.7 percent to 31.5 percent), gave opportunities for reading practice (94.1 percent to 70.0 percent) and used questions to check understanding (82.5 percent to 36.5 percent). While a higher proportion of teachers at endline than midline connected lesson content to students' knowledge and daily lives, the difference was not significant.

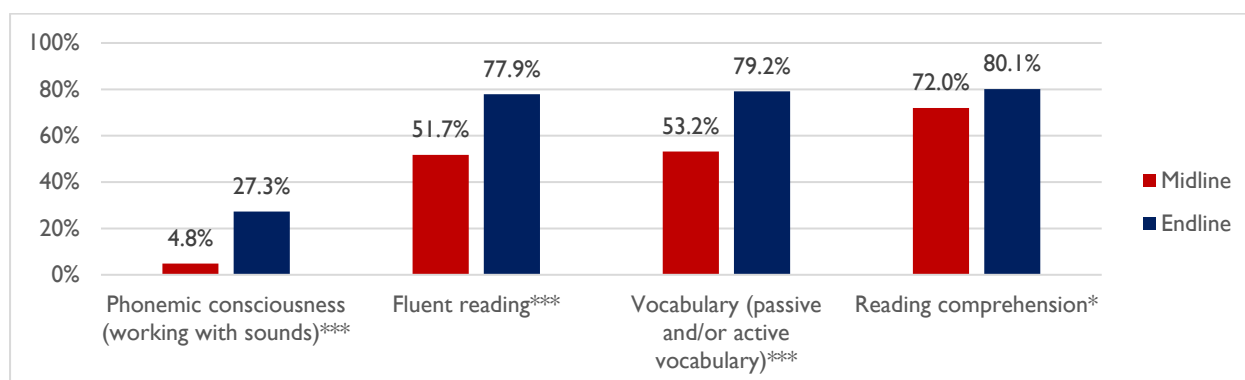
Figure 40. Percentage of Tajik Grade 4 Teachers Using Specific Teaching Techniques, Midline and Endline



TAJIK GRADE 4 READING SKILLS

At endline, a significantly higher proportion of teachers emphasized the fundamental reading skills of phonemic consciousness (27.3 percent at endline, compared with 4.8 percent at midline), fluent reading (77.9 percent to 51.7 percent), vocabulary (79.2 percent to 53.2 percent), and reading comprehension (80.1 percent to 72.0 percent). A relatively low proportion of teachers focused on phonemic consciousness compared to other skills, likely because this fundamental skill is included in curricula in earlier grades.

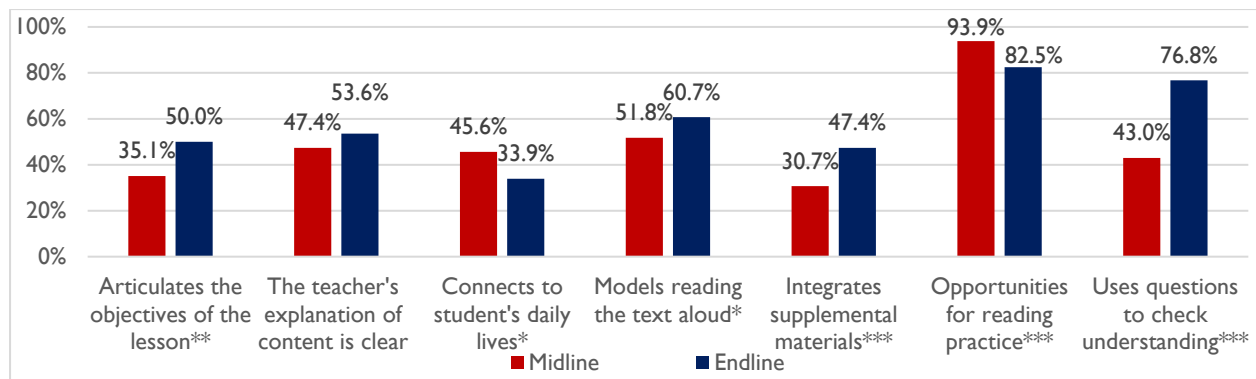
Figure 41. Percentage of Tajik Grade 4 Teachers Focusing on Specific Reading Skills, Midline and Endline



RUSSIAN GRADE 2 TEACHING TECHNIQUES

Overall, the proportion of teachers using specific RWM teaching techniques generally increased. A significantly higher proportion of teachers at endline articulated the objectives of the lesson (50.0 percent) and modeled reading out loud (60.7 percent) than midline (35.1 percent and 51.8 percent, respectively). A significantly higher proportion of teachers also used questions to check understanding (76.8 percent at endline, compared with 43.0 percent at midline) and integrated supplemental materials (47.4 percent to 30.7 percent). However, the proportion of teachers using certain teaching techniques significantly decreased—connecting the lesson to students' knowledge and daily lives (33.9 percent at endline, compared with 45.6 percent at midline) and including opportunities to practice reading (82.5 percent to 93.9 percent).

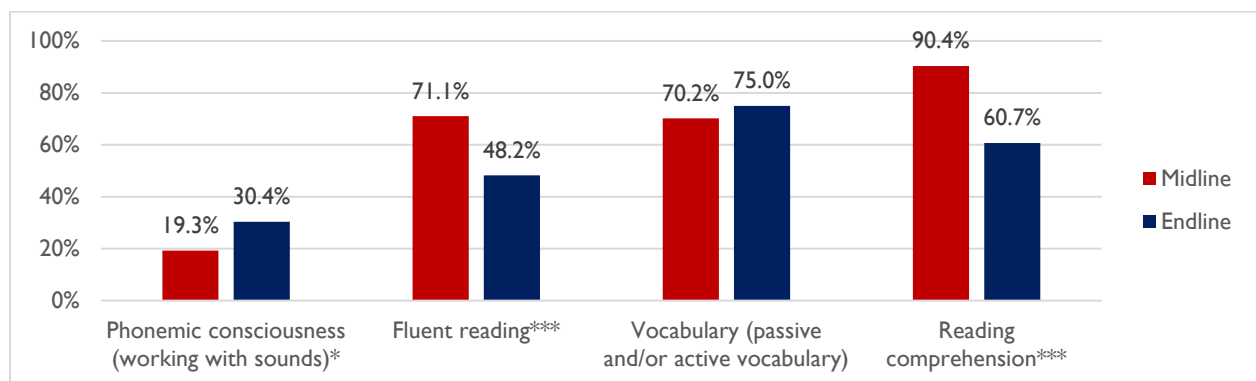
Figure 42. Percentage of Russian Grade 2 Teachers Using Specific Teaching Techniques, Midline and Endline



RUSSIAN GRADE 2 READING SKILLS

At endline, a higher proportion of teachers emphasized phonemic consciousness (30.4 percent at endline, compared with 19.3 percent at midline), and vocabulary (75.0 percent to 70.2 percent), while a lower proportion at endline included fluent reading (48.2 percent to 71.1 percent) and reading comprehension (60.7 percent to 90.4 percent). All differences were statistically significant, except for vocabulary.

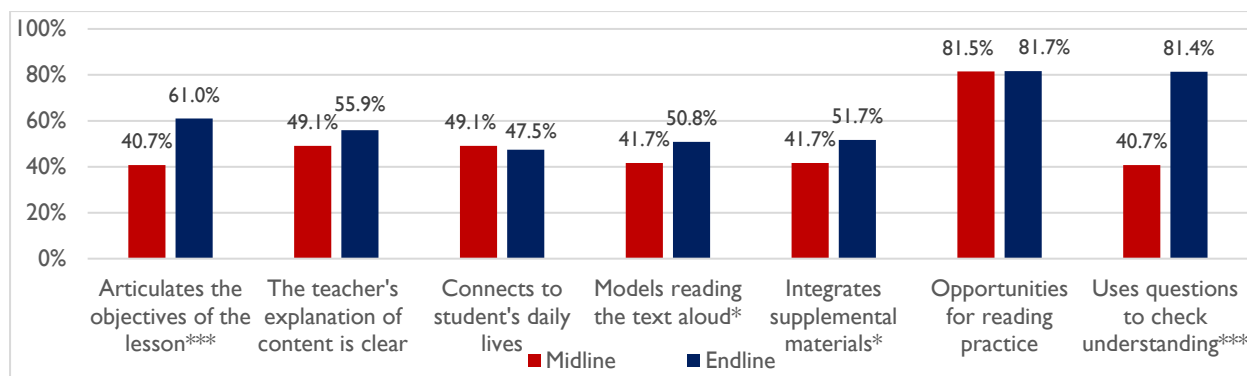
Figure 43. Percentage of Russian Grade 2 Teachers Focusing on Specific Reading Skills, Midline and Endline



RUSSIAN GRADE 4 TEACHING TECHNIQUES

Overall, the proportion of teachers using specific RWM teaching techniques increased at endline. A significantly higher proportion of teachers articulated the objectives of the lesson at endline (61.0 percent) than midline (40.7 percent). A significantly higher proportion of teachers also modeled reading out loud (50.8 percent at endline, compared with 41.7 percent at midline), integrated supplemental materials (51.7 percent to 41.7 percent), and used questions to check understanding (81.4 percent compared to 40.7 percent). While a higher proportion of teachers clearly explained content at endline than midline, the difference was not significant. At endline, the proportion of teachers connecting content to students' lives and providing opportunities for reading practice was relatively comparable to midline.

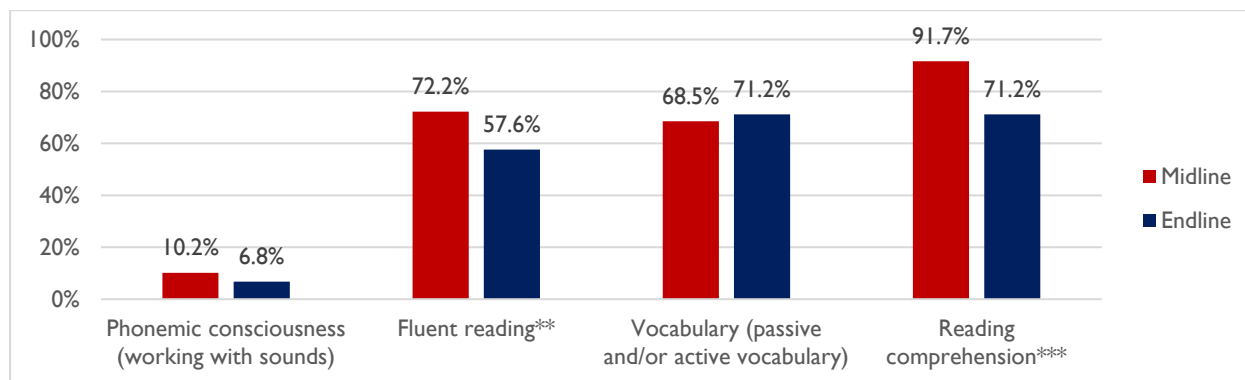
Figure 44. Percentage of Russian Grade 4 Teachers Using Specific Teaching Techniques, Midline and Endline



RUSSIAN GRADE 4 READING SKILLS

A lower proportion of Russian Grade 4 teachers focused on most fundamental reading skills at endline compared to midline. The proportion of teachers focusing on certain reading skills significantly decreased from midline to endline—emphasizing fluent reading (57.6 percent at endline, compared with 72.2 percent at midline), and reading comprehension (71.2 percent to 91.7 percent). A slightly higher percentage of teachers focused on vocabulary at endline than midline (71.2 percent to 68.5 percent), but the difference was not significant. As with Grade 4 classrooms in Tajik, a relatively low proportion of teachers focused on phonemic consciousness compared with other skills, likely because this fundamental skill is emphasized in earlier grades.

Figure 45. Percentage of Russian Grade 4 Teachers Focusing on Specific Reading Skills, Midline and Endline



RESOURCES

The level of resources available was associated with improvements in ORF, ranging from 3.48 to 13.96 CWPM, as shown in Table 20. Subgroup performance for Oral Reading Fluency – Tajik Grade 4. The only negative association was observed in the Russian grade 2 results, for which the availability of project books was associated with a decrease of 5.34 CWPM in ORF scores. Since most schools had books available, this one exception is likely spurious, resulting from the few schools without books available.

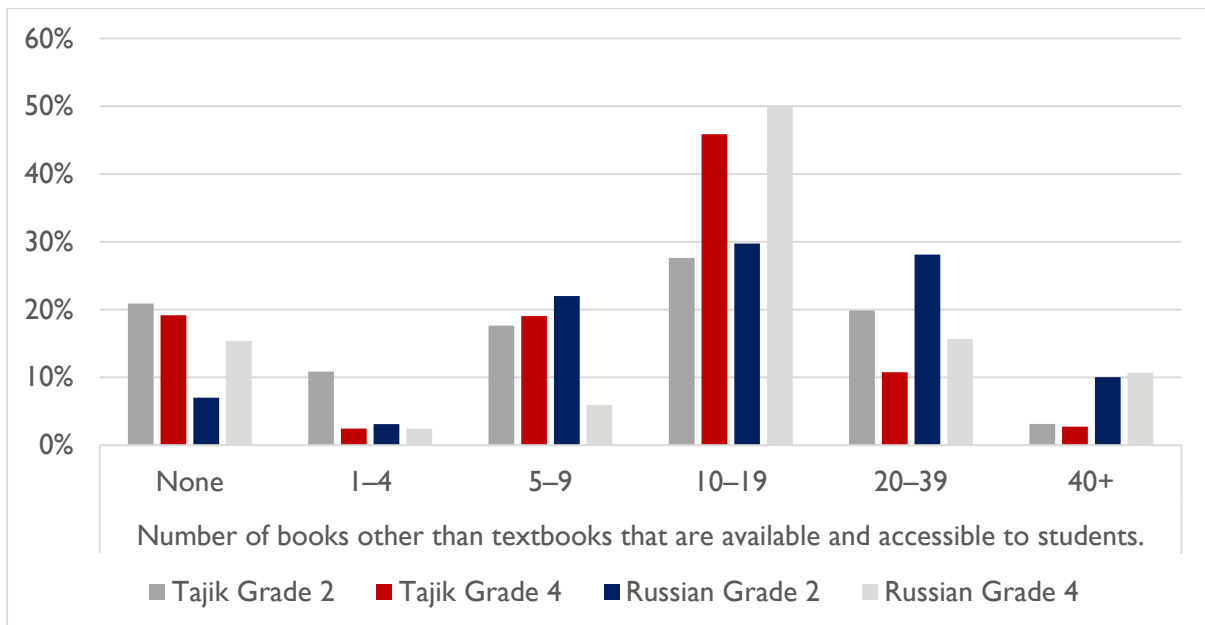
Table 26. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language

		Associated increase/decrease in ORF score (CWPM)			
Materials or Instructional Practices		Tajik		Russian	
		Grade 2	Grade 4	Grade 2	Grade 4
Resources	Class has books other than textbooks			13.96	
	Student resources index (0-3)	5.90	11.20	3.48	
	Teacher resources index (0-5)	4.89		4.01	
	Books provided by the project are available (1=Yes)		7.20	-5.34	

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. In particular, green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

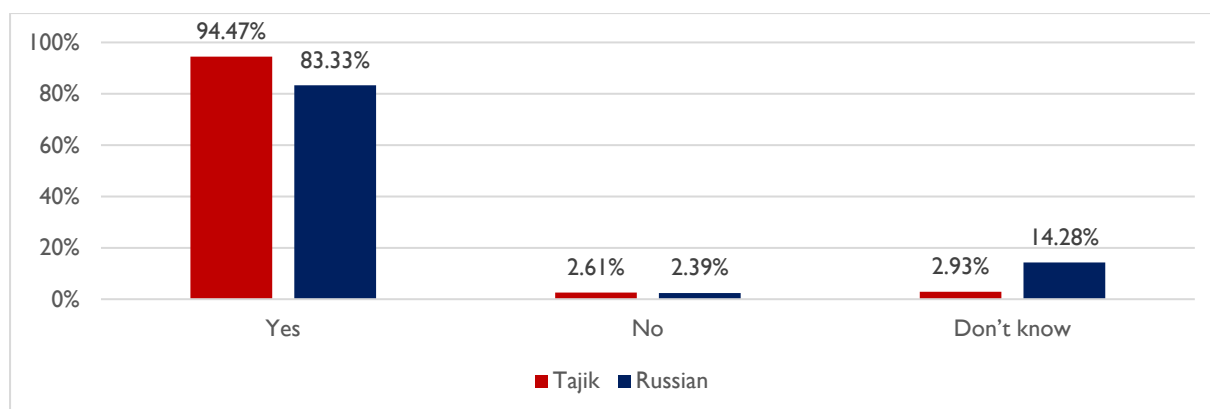
Most classrooms had 10 or more books other than textbooks available, as displayed in Figure 46. Books Other than Textbooks, by Language and Grade Level. The positive connection between ORF scores and having more books suggests that the greater availability of books may encourage more student reading or enable more extracurricular reading-based activities, leading to improved reading outcomes.

Figure 46. Books Other than Textbooks, by Language and Grade Level



As noted, books provided by the project were available in most classrooms, especially Tajik schools, as shown Figure 47.

Figure 47. Books Provided by the Project are Available, by Language and Grade Level



As Table 26 above shows, a positive relationship was found between ORF scores and students' resources, as well as teachers' resources, which were both estimated using indexes. With a maximum score of three, the student resources index was calculated based on the presence of three items—appropriate textbooks, language exercise books, and vocabulary books. At least 90 percent of students present in the classroom had to possess each item. Higher scores indicated that students were better prepared for reading lessons. With a maximum score of five, the teacher resources index was calculated based on the presence of five items—a board, chalk or markers, an interactive board, other visual aids, and a developed lesson plan or lesson summary. Higher scores indicated that teachers had access to more resources for instructing students in reading. The index scores differed somewhat across languages and grades, but on average, students and teachers at all schools reported indexes in the upper 50th percentile, as shown in Table 27. Therefore, the positive relationship between these indexes and ORF scores suggests that proficiency is resource sensitive, indicating a path to further improvement in reading outcomes.

Table 27. Student and Teacher Resources Indexes, Means by Language and Grade Level

	Tajik Grade 2	Tajik Grade 4	Russian Grade 2	Russian Grade 4
Student resources index (0-3)	2.31	2.42	1.98	2.24
Teacher resources index (0-5)	3.65	3.70	3.85	3.75

The lack of certain items in classrooms resulted in differences in the index scores, as displayed in Figure 48 to Figure 50. Although most students had textbooks for reading⁴⁶ in the LOI as well as language exercise books, fewer had vocabulary exercise books, especially in Russian schools. As for teachers, most had developed lessons plans, visual aids, and chalk or markers, but many lacked interactive boards in their classrooms, particularly teachers in Tajik schools.

⁴⁶ In Russian, students have textbooks devoted solely to reading. In Tajik, students have textbooks for both reading and Tajik language.

Figure 48. Percentage of Students with Textbooks, Language Exercise and Vocabulary Exercise Books, by Language and Grade Level

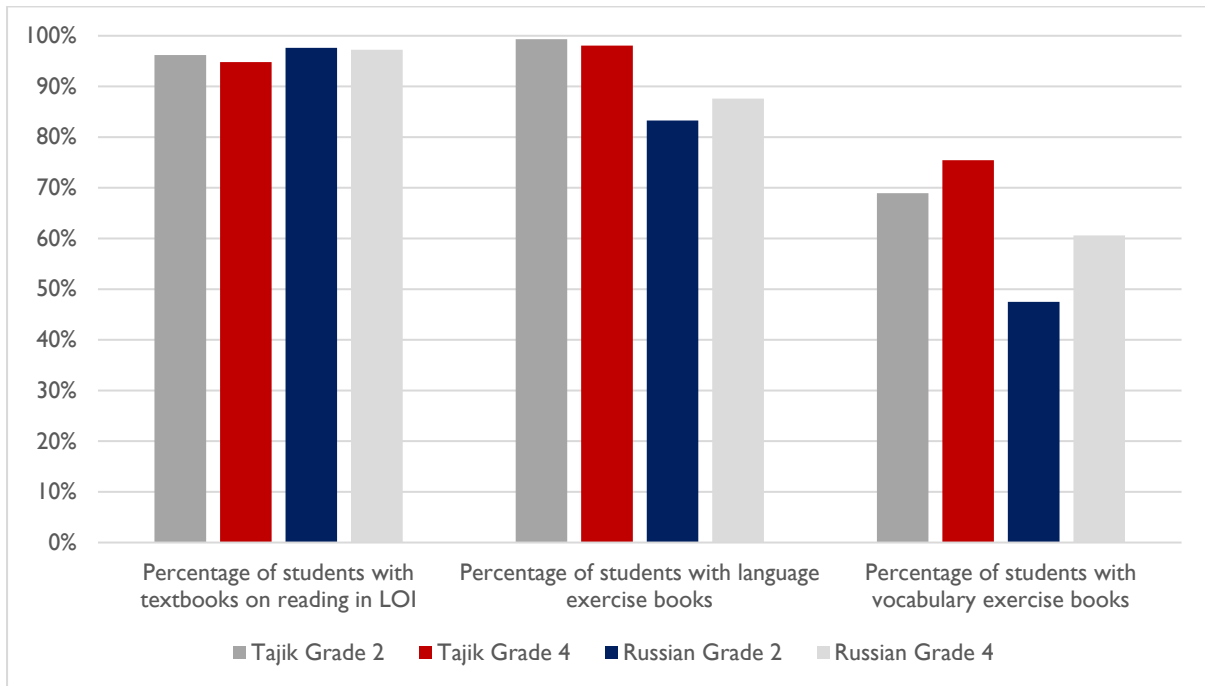


Figure 49. Materials Available to Teachers, by Language and Grade Level

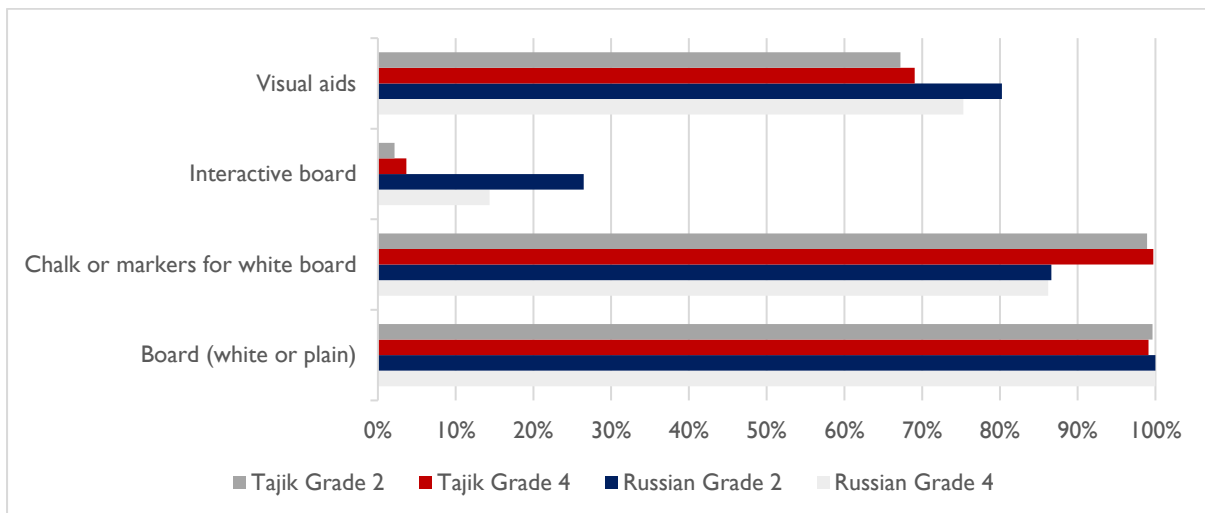
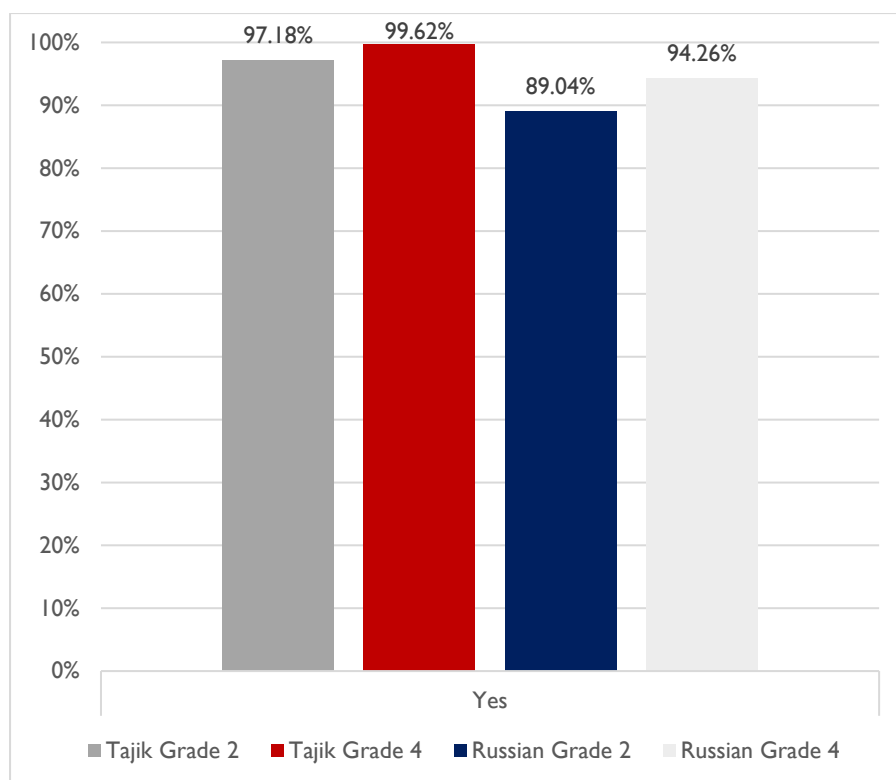


Figure 50. Teacher with Developed Lesson Plan, by Language and Grade Level



OTHER RWM EFFORTS

Other RWM efforts, including reading corners and the use of logbooks, were associated with improvements in ORF scores from 5.99 to 11.59 CWPM. None of these efforts, however, were associated with improvement in Russian grade 4 ORF scores.

Table 28. Associations between RWM Materials or RWM Instructional Practices and Oral Reading Fluency by Grade and Language

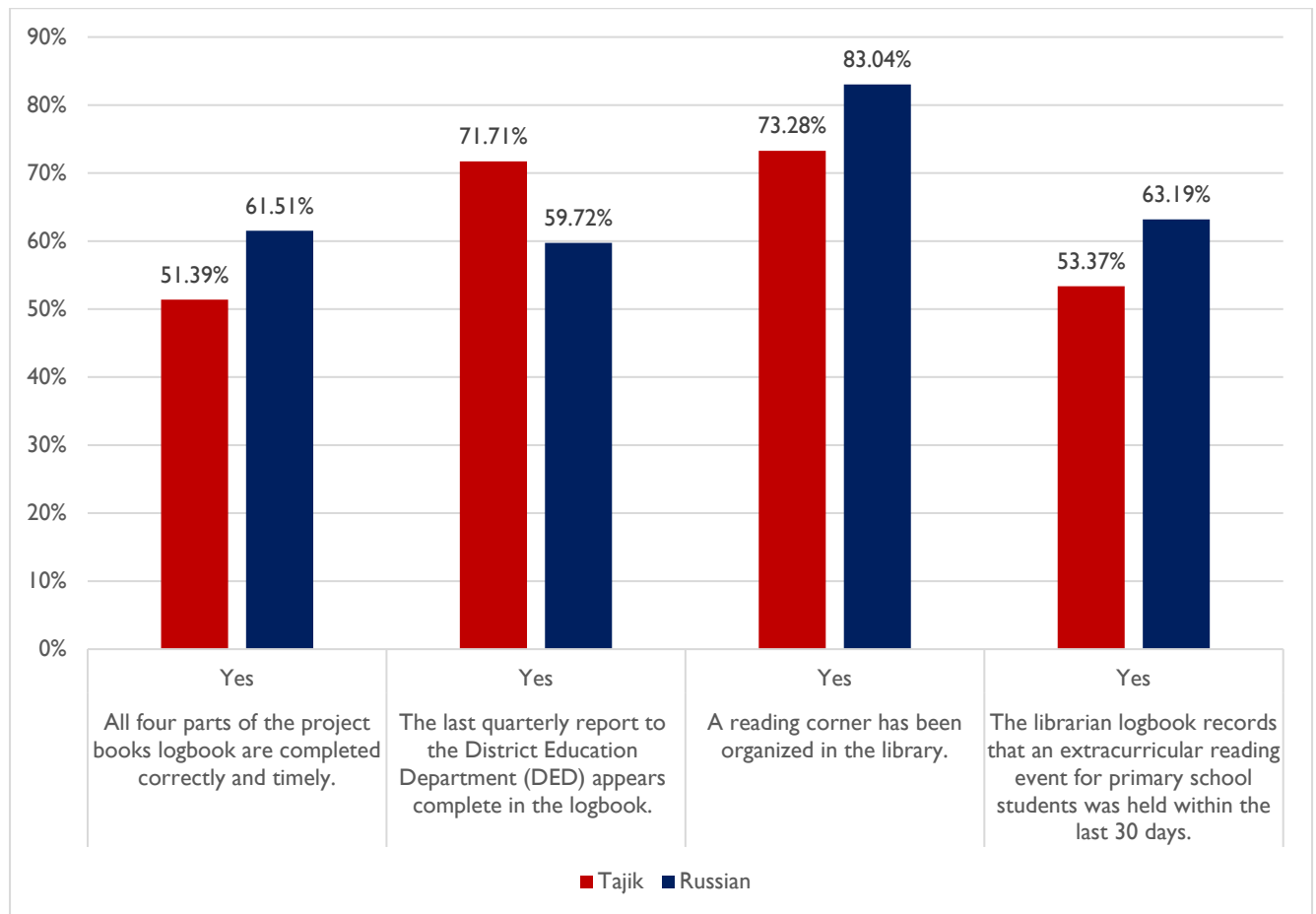
Materials or Instructional Practices		Associated increase/decrease in ORF score (CWPM)			
		Tajik		Russian	
		Grade 2	Grade 4	Grade 2	Grade 4
Other RWM Efforts	All four parts of the project books logbook are completed correctly and timely (I=Yes)			6.58	
	The last quarterly report to the DED appears complete in the logbook (I=Yes)		11.59		
	The librarian logbook records that an extracurricular reading event for primary school students was held within the last 30 days (I=Yes)	4.77		5.99	
	A reading corner has been organized in the library (I=Yes)		6.06	7.96	

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. In particular, green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

While most Tajik and Russian schools had a reading corner for primary grades organized in the library (defined as having books for early grade students and a place to sit and read), as shown in Figure 51, the use of logbooks varied. While 71.1 percent of Tajik schools had a complete quarterly report to the District Education Department, only slightly more than half had all parts of the

projects book logbook completed, as well as logbook records showing that an extracurricular reading event for primary school students had been held within the last 30 days. By contrast, the proportions for these three logbook-related efforts were similar for Russian schools.

Figure 51. Other RWM Efforts, by Language and Grade Level



Evaluation Question 4: Which contextual factors or other classroom measures are predictors of Tajik and Russian reading outcomes of students in Grade 2 and Grade 4 in schools supported by RWM?

This section responds to the fourth evaluation question by reporting results of the analyses that measure whether contextual factors other than RWM-specific materials and teachers’ reading instructional practices serve as predictors of student reading performance as measured by ORF. To get a project-level picture, results are presented together across languages and grade levels. However, for simplicity, results are presented by survey/questionnaire.

Survey items with sufficient variance were included in a hierarchical⁴⁷ regression analysis to determine their relationship to ORF scores. Each factor was analyzed individually and after controlling for the location and region of the schools.

STUDENT-LEVEL PREDICTORS OF READING OUTCOMES

For Tajik grade 2 students, the following factors were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

⁴⁷ Except for Tajik G4

- Being female, +4.54 CWPM
- Speaking Tajik at home, +8.98 CWPM
- Speaking Pamiri dialects at home, +15.14 CWPM
- Having reading books at home, +4.91 CWPM
- Having a father who can read, +3.52 CWPM
- Having a teacher assign homework more frequently,⁴⁸ +1.71 CWPM
- Having a teacher who rephrases questions, +5.47 CWPM higher
- Having a teacher encourage students to try harder if they cannot respond to a question, +8.45 CWPM
- Studying after school for one additional day per week, +1.21 CWPM
- Having reading books at school that students can take home, +3.27 CWPM

On the other hand, some factors for Tajik grade 2 students were negatively associated with ORF, with each listed relationship being statistically significant and the decrease in ORF scores associated with each factor noted:

- Speaking Uzbek at home, -7.94 CWPM
- Having a teacher who hits students if they are unable to answer a question, -4.7 CWPM
- Having a teacher who sends students to the corner of the classroom if they are unable to answer a question, -6.55 CWPM
- Receiving help from someone at home when doing homework, -4.10 CWPM

For Tajik grade 4 students, the following factors were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Speaking Pamiri dialects at home, +28.08 CWPM
- Having reading books at home, +14.47 CWPM
- Having a teacher who rephrases questions, +7.63 CWPM
- Having a teacher who encourages students to try harder, +7.99 CWPM
- Having reading books at school that can be taken home, +8.87 CWPM

On the other hand, some factors for Tajik grade 4 students were negatively associated with ORF, with each listed relationship being statistically significant and the decrease in ORF scores associated with each factor noted:

- Having a teacher who scolds students if they are unable to answer a question, -11.40 CWPM
- Having a teacher who hits students if they are unable to answer a question, -10.91 CWPM

For Russian grade 2 students, the following factors were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Being female, +6.50 CWPM
- Speaking Russian at home, +3.97 CWPM
- Having reading books at home, +3.70 CWPM
- Having teachers who rephrases questions, +4.18 CWPM
- Having a teacher who encourage students to try harder when unable to answer a question, +6.26 CWPM

On the other hand, some factors for Russian grade 2 students were negatively associated with ORF, with each listed relationship being statistically significant and the decrease in ORF scores associated

⁴⁸ The positive association between ORF and teachers' frequency of assigning homework was measured by increasing frequency based on the following scale—never to rarely, rarely to once a week, once a week to once every other lesson, or once every other lesson to once a lesson.

with each factor noted:

- Speaking Pamiri dialects at home, -11.22 CWPM
- Having a teacher who puts marks if students are unable to answer a question, -2.64 CWPM
- Having a teacher who scolds students if they are unable to answer a question, -5.84 CWPM
- Receiving help with homework from someone at home, -6.54 CWPM

For Russian grade 4 students, the following factors were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Speaking Russian at home, +5.71 CWPM
- Having books at home, +14.94 CWPM
- Having a mother who can read, +7.76 CWPM
- Having a father who can read, +5.71 CWPM
- Practicing reading aloud at home with someone, +4.88 CWPM
- Having a teacher assign homework more frequently,⁴⁹ +5.07 CWPM
- Having a teacher who rephrases a question, +5.50 CWPM
- Having a teacher ask a question again if students are unable to answer it, +9.35 CWPM

On the other hand, some factors for Russian grade 4 students were negatively associated with ORF, with each listed relationship being statistically significant and the decrease in ORF scores associated with factors noted if enough responses were given to compute meaningful coefficients:

- Speaking Tajik at home, -3.88 CWPM
- Having a teacher who puts marks if they are unable to answer a question, -5.33 CWPM

The most salient consistencies were observed for three variables. These variables had statistically significant relationships for all four groups—Tajik grades 2 and 4 and Russian grades 2 and 4:

- A positive association with having reading books at home
- A positive association with having a teacher who encourages students to try harder when they do not respond to a question
- A negative association with having a teacher who hits students when unable to answer a question

Overall, literacy resources and good teaching practices were positively associated with ORF scores, whereas negative teaching practices were associated with declines in ORF scores. In addition, some results were counterintuitive. For example, receiving help at home with homework could be thought of as positively associated with ORF scores, but it could be that students who receive more help rely more on others and learn less.

Table 29. Associations between Student Variables and Oral Reading Fluency by Grade and Language

Student Survey Variable	Associated increase/decrease in ORF score (CWPM)			
	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
Student is female (I=Yes)	4.54		6.50	
Language spoken at home - Tajik (I =Yes)	8.98			-3.88
Language spoken at home - Russian (I =Yes)			3.97	5.71

⁴⁹ The positive association between ORF and teachers' frequency of assigning homework was measured by increasing frequency based on the following scale—never to rarely, rarely to once a week, once a week to once every other lesson, or once every other lesson to once a lesson.

Associated increase/decrease in ORF score (CWPM)				
Student Survey Variable	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
Language spoken at home - Pamiri dialects (I =Yes)	15.14	28.08	-11.22	
Language spoken at home - Uzbek (I =Yes)	-7.94			
Has reading books at home (I=Yes)	4.91	14.47	3.70	14.94
Mother can read (I=Yes)				7.76
Father can read (I=Yes)	3.52			5.71
Practices reading aloud to someone at home (I=Yes)				4.88
Frequency with which teacher assigns homework (Categorical)	1.71			5.07
If unable to answer a question - teacher puts mark (I=Yes)			-2.64	-5.33
If unable to answer a question - teacher rephrases, explains (I=Yes)	5.47	7.63	4.18	
If unable to answer a question - teacher encourages student to try harder (I=Yes)	8.45	7.99	6.26	5.50
If unable to answer a question - teacher asks again (I=Yes)				9.35
If unable to answer a question - teacher corrects the student but does not scold him/her (I=Yes)				
If unable to answer a question - teacher scolds student (I=Yes)		-11.40	-5.84	
If unable to answer a question - teacher sends student outside of classroom (I=Yes)				
If unable to answer a question - teacher hits student (I=Yes)	-4.70	-10.91	*	*
If unable to answer a question - teacher sends student to the corner of the classroom (I=Yes) ⁵⁰	-6.55		*	*
Number of days after school studying (0-7)	1.21			
When doing homework, receives help from someone at home (I=Yes)	-4.10		-6.54	
Has reading books at school that can be taken home (I=Yes)	3.27	8.87		

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. Green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

* Results are omitted due to extremely small number of affirmative responses

TEACHER-LEVEL PREDICTORS OF READING OUTCOMES

For Tajik grade 2 students, the following factors related to their teachers were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

⁵⁰ Reports of teachers hitting students or sending them out of the classroom were rare compared to other responses. The low N of responses for these values might explain the large associated change in ORF scores. Among students, 11.32% in Tajik Grade 2, 3.23% in Tajik Grade 4, 0.79% in Russian Grade 2, and 0.23% in Russian Grade 4 reported that the teacher hits students if they are unable to answer a question.

- Teachers who speak Pamiri dialects at home, +7.04 CWPM
- Teachers with more experience, with each additional year of experience associated with +0.3 CWPM
- Teachers who have attended in-service training in the last year, +5.17 CWPM
- Teachers who have attended in-service training on how to teach reading, +5.51 CWPM
- Teachers who have received methodological support in the past year, +5.22 CWPM
- Teachers who believe that children should understand stories by the end of the year, +7.56 CWPM
- Teachers who know that parents review students' homework, +4.46 CWPM

For Tajik grade 4 students, the following factors related to their teachers were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Female teachers, +8.40 CWPM
- Teachers who speak Pamiri dialects at home, +26.86 CWPM
- Teachers who advanced further in school,⁵¹ +4.57 CWPM
- Teachers with more experience, with each additional year of experience associated with +0.44 CPWM
- Teachers who have received in-service training on how to teach reading, +10.55 CWPM
- Teachers with a teacher guide, +27.68 CWPM
- Teachers who do not need help with teaching, +5.49 CWPM
- Teachers who seek advice from education supervisor or subject specialist when needing help, +10.40 CWPM

On the other hand, teachers who believed that students should read fluently later in primary school were associated with a decrease in ORF scores, with each additional grade level for which teachers expected children to read associated with a decrease in ORF of 7.53 CWPM.

For Russian grade 2 students, the following factors related to their teachers were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Teachers who advanced further in school,⁵² +3.04 CWPM
- Teachers with more experience, with each additional year of experience associated with +0.42 CWPM
- Teachers who expect children to read grade level stories, +6.24 CWPM
- Teachers who expect students to sound out words they do not know by the end of the year, +6.03 CWPM
- Teachers who expect children to write fluently later in primary school, +4.80 CWPM
- Teachers who discuss teaching practice casually with other teachers when needing help, +4.06 CWPM
- Teachers who seek advice from the education supervisor or subject specialist when needing help with teaching, +4.07 CWPM
- Teachers who believe that more parents review students' homework, +5.32 CWPM
- Teachers satisfied with parental involvement in the classroom, +6.94 CWPM
- Teachers who teach in larger classrooms, +0.54 CWPM

⁵¹ The positive association between ORF and having a teacher who advanced further in school was measured based on the following scale—from secondary to incomplete higher education, or from incomplete higher education to complete higher education.

⁵² The positive association between ORF and having a teacher who advanced further in school was measured based on the following scale—from secondary to incomplete higher education, or from incomplete higher education to complete higher education.

On the other hand, some factors related to Russian grade 2 students' teachers were negatively associated with ORF, with each listed relationship being statistically significant and the decrease in ORF scores associated with each factor noted:

- Teachers who speak Pamiri dialects at home, -8.23 CWPM
- Teachers who assign students to read during school time, -2.85 CWPM
- Teachers who have assigned students to read at home in the past five school days, -4.87 CWPM
- Teachers who have been observed more frequently by the deputy director,⁵³ -2.86 CWPM

For Russian grade 4 students, the following factors related to their teachers were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Teachers who speak Russian at home, +19.04 CWPM
- Teachers with more experience, with each additional year of experience associated with +0.29 CWPM
- Teachers who have received methodological support in the past year, +7.75 CWPM
- Teachers who expect children to sound out words they do not know by the end of the year, +10.98 CWPM
- Teachers who have the impression that more parents review students' homework, +6.97 CWPM
- Teachers satisfied with parental involvement in the classroom, +6.34 CWPM
- Teachers who teach larger classes, +0.48 CWPM

On the other hand, some factors related to Russian grade 4 students' teachers were negatively associated with ORF, with each listed relationship being statistically significant and the decrease in ORF scores associated with each factors noted:

- Teachers who speak Tajik at home, -5.98 CWPM
- Teachers who speak Pamiri dialects at home, -18 CWPM
- Teachers with a teacher guide, -7.68 CWPM
- Teachers who discuss teaching practice casually with other teachers when needing help, -5.24 CWPM
- Teachers who seek support from the education advisor or subject specialist when needing help, -11.47 CWPM

Teachers' levels of experience were the most consistent result across grade levels and languages. More experienced teachers were associated with small but consistent improvements in ORF scores. Some factors were inconsistent, however. While teachers who speak Pamiri dialects at home were related to gains in ORF for students at Tajik-speaking schools, it was related to declines in ORF at Russian-speaking schools. Other inconsistent factors included teachers with a teacher guide and teachers who seek advice from the education supervisor or subject specialist when needing help. Teachers who speak Uzbek at home were not significantly related to achievement either positively or negatively.

⁵³ The negative association between ORF and having a teacher who was observed more frequently by the deputy director was measured based on the following scale—from once every 2-3 months to once every month, from once every month to once every two weeks, from once every two weeks to once every week, or from once every week to daily.

Table 30. Associations between Teacher Variables and Oral Reading Fluency by Grade and Language

Teacher Survey Variable	Associated increase/decrease in ORF score (CWPM)			
	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
Teacher is female (I=Yes)		8.40	N.A. ⁵⁴	N.A. ⁴⁵
Native language - Tajik (I=Yes)				-5.98
Native language - Russian (I=Yes)				19.04
Native language - Pamiri dialects (I=Yes)	7.04	26.86	-8.23	-18.00
Highest level of education (Categorical)		4.57	3.04	
Years of experience	0.30	0.44	0.42	0.29
Attended in-service training or professional development sessions such as workshops in the last year (I=Yes)	5.17			
Attended in-service training on how to teach reading (I=Yes)	5.51	10.55		
Received methodological support or assistance at school this past year (I=Yes)	5.22			7.75
Last 5 school days - students were assigned to do reading on their own in school time (I=Yes)			-2.85	
Last 5 school days - students were assigned to do reading at home (I = Yes)			-4.87	
Has teacher guides (I=Yes)		27.68		-7.68
Reading skills children should have at the end of the school year - Read grade level stories (I = Yes)			6.24	
Reading skills children should have at the end of the school year - Sound out words they don't know (I = Yes)			6.03	10.98
Reading skills children should have at the end of the school year - Understand stories that they read (I = Yes)	7.56			
Grade level at which students should read fluently (Categorical)		-7.53		
Grade level at which students should write (Categorical)			4.80	
Frequency with which deputy director observes classes (Categorical)			-2.86	
Never need help with their teaching (I=Yes)		5.49		
Discuss casually with other teachers when need help with their teaching (I=Yes)			4.06	-5.24
Seek advice from education supervisor or subject specialist when need help with their teaching (I=Yes)		10.40	4.07	-11.47
Number of parents who review students' homework (Categorical)	4.46		5.32	6.97
Satisfied with parental involvement in the classroom (I=Yes)			6.94	6.34
Total class enrollment			0.54	0.48

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. Green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

⁵⁴ There were only male teachers in Russian schools.

SCHOOL DIRECTOR-LEVEL PREDICTORS OF READING OUTCOMES

For Tajik grade 2 students, two director-level factors were positively associated with ORF. Directors who reported that teachers have received training on how to teach reading outside of in-service teacher training were associated with an increase of 15.45 CWPM, while holding parent-teacher association (PTA) meetings regularly was associated with an increase of 9.99 CWPM. Both relationships were statistically significant.

For Tajik grade 4 students, the following factors related to their school directors were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Directors who reported that teachers have received training on how to teach reading outside of in-service training, +13.76 CWPM
- Directors reporting higher levels of classroom observation,⁵⁵ +5.33 CWPM
- Having a library or reading room at school, +13.01 CWPM
- Holding regular PTA meetings, +31.07 CWPM

On the other hand, one director-level factor was negatively associated with ORF scores for Tajik grade 4 students. Classrooms in which twice as many girls were present than boys were associated with a decline of 10.12 CWPM. The relationship was statistically significant.

For Russian grade 2 students, the following factors related to their school directors were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- Directors who receive support to be more effective in teaching students with disabilities, +7.26 CWPM
- Directors who reported that teachers have received training on how to teach reading outside of in-service training, +2.47 CWPM
- Directors satisfied with the level of support from the PTA, +7.13 CWPM
- Student-class ratio for grade 2, +0.58 CWPM
- Student-class ratio for grade 4, +0.39 CWPM

On the other hand, two director-level factors were negatively associated with ORF scores for Russian grade 2 students. Students at schools with female directors were associated with a decline of 8.73 CWPM, and directors reporting having sufficient resources materials and textbooks were associated with a decline of 4.60 CWPM. Both relationships were statistically significant.

For Russian grade 4 students, one director-level factor was positively associated with ORF, with the relationship being statistically significant. Directors who reported that teachers have received training on how to teach reading outside of in-service training were associated with an increase of 10.40 CWPM. On the other hand, one director-level factor was negatively associated with ORF scores, with the relationship being statistically significant. Each additional year of experience for a school director was associated with a decline of 0.31 CWPM.

The only consistent finding across languages and grade levels was with directors who reported that teachers have received training on how to teach reading outside of the in-service teacher training. This finding partially mirrored the analysis around teacher-level predictors.

⁵⁵ The positive association between ORF and having a director who reported observing classes more frequently was measured based on the following scale—from once every 2-3 months to once every month, from once every month to once every two weeks, from once every two weeks to once every week, or from once every week to daily.

Table 31. Associations between Director Variables and Oral Reading Fluency by Grade and Language

School Director Variable ⁵⁶	Associated increase/decrease in ORF score (CWPM)			
	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
Director is female (I=Yes)			-8.73	
Years of experience				-0.31
School has a program to support director to be more effective in teaching students with disabilities (I=Yes)			7.26	
Primary school teachers have received training on how to teach reading, outside of in-service teacher training (I=Yes)	15.45	13.76	2.47	10.40
Frequency of classroom observation (Categorical)		5.33		
Sufficient resource materials/textbooks (I=Yes)			-4.60	
Library or reading room (I=Yes)		13.01		
Holds regular parents-teachers association meetings (I=Yes)	9.99	31.07		
Satisfied with the level of support the PTA provides to the school (I=Yes)			7.13	
Girl to boy ratio ⁵⁷ - G2			-10.12	
Student - class ratio G2			0.58	
Student - class ratio G4			0.39	

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. Green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

SCHOOL-LEVEL PREDICTORS OF READING OUTCOMES

For Tajik grade 2 students, the only school-level factor related to ORF scores was the number of library books for primary students,⁵⁸ which was associated with an increase of 10.18 CWPM.

For Tajik grade 4 students, the following factors related to their school were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- The number of books for primary students in the library,⁵⁹ +11.16 CWPM
- Clean and tidy school building and school grounds, +10.16 CWPM
- School environment index,⁶⁰ with each increase of one point in the four-point index associated with +5.48 CWPM

⁵⁶ A few relationships were omitted due to the small number of relevant cases (< 2%) or due to a coefficient smaller than 0.01 CWPM.

⁵⁷ Total number of girls/total number of boys

⁵⁸ The positive association between ORF and the number of books was measured based on the following scale—from no books to 1–50 books, from 1–50 books to 51–100 books, and from 51–100 books to more than 100 books.

⁵⁹ Ibid.

⁶⁰ The school environment index reflected whether a school had clean grounds, a library, 100 or more reading books available for primary grade students in the library, and a reading corner established in the library. The maximum possible score for this index was four, and higher scores indicate a more conducive environment for learning at the school.

For Russian grade 2 students, the following factors related to their school were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- The number of books for primary students in the library,⁶¹ +3.91 CWPM
- Clean and tidy school building and school grounds, +10.05 CWPM
- School environment index, with each increase of one point in the four-point index associated with +8.79 CWPM

For Russian grade 4 students, the following factors related to their school were positively associated with ORF, with each listed relationship being statistically significant and the increase in ORF scores associated with each factor noted:

- The number of books for primary students in the library,⁶² +9.33 CWPM
- School environment index, with each increase of one point in the four-point index associated with +8.66 CWPM

On the other hand, three school-level factors were negatively associated with ORF scores for Russian grade 4 students. Schools that used Tajik as well as Russian as an LOI were associated with a decrease of 15.56 CWPM, while school that used Uzbek as well as Russian as an LOI were associated with a decrease of 17.09 CWPM. Students attending schools with additional shifts were associated with a decrease of 8.70 CWPM.

The only consistent finding across languages and grade levels had to do with having more library books for students, which was positively associated with students' ORF scores.

Table 32. Associations between School Inventory Variables and Oral Reading Fluency by Grade and Language

School Inventory Variable	Associated increase/decrease in ORF score (CWPM)			
	Tajik		Russian	
	Grade 2	Grade 4	Grade 2	Grade 4
Language of Instruction includes Tajik (1=Yes)				-15.56
Language of Instruction includes Uzbek (1=Yes)				-17.09
Number of shifts for primary students (1-3)				-8.70
Number of books for primary students in the library (0= No books, 1 = 1–50 books, 2 = 51–100 books, 3 = More than 100 books)	10.18	11.16	3.91	9.33
The school building and the school grounds are clean and tidy (1=Yes)		10.16	10.05	
School environment index (0-4)		5.48	8.79	8.66

Note: Highlighted cells show variables that have a significant association with ORF for the given language and grade level, after controlling for region and urbanicity. In particular, green cells show positive associations, and pink cells show negative associates. Variables not included in this table were not significantly related to ORF across languages and grade levels.

⁶¹ The positive association between ORF and the number of books was measured based on the following scale—from no books to 1–50 books, from 1–50 books to 51–100 books, and from 51–100 books to more than 100 books.

⁶² Ibid.

Evaluation Question 5: What proportion of students can read and understand the meaning of grade level text (Standard Foreign Assistance (F) Indicators ES.1-1 and ES.1-2) at each time point in schools served by RWM in Grade 2 and Grade 4 in Tajik and in Russian?

To answer the final evaluation question, this section presents the percentage of students attaining reading proficiency benchmarks for both grades and for each language. These results correspond to USAID’s Foreign Assistance F-indicators ES.1-1 and ES.1-2: the proportion of student in RWM schools who demonstrated that they could read and understand the meaning of the grade-level text. Student performance on the ORF and reading comprehension subtasks provide the data for these indicators. The reading and comprehension benchmarks were established in cooperation with the MoES—40 CWPM for grade 2 students and 80 CWPM for grade 4 students in both languages—and 80 percent of reading comprehension questions answered correctly—or four out of five correct answers.

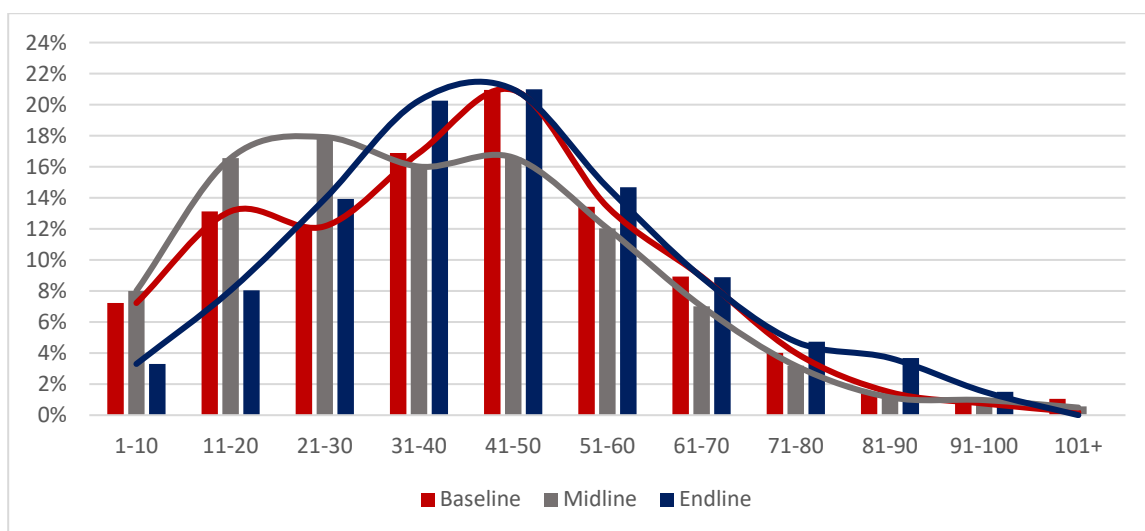
As noted in the midline report, comparisons of the reading comprehension benchmark findings should be made with caution. With only five items for this subtask, one additional correct answer improves a student’s score by 20 percentage points, which is a large gain. Such few items implies lower levels of reliability on the measure. In addition, comparisons between languages should be avoided due to the sociolinguistic differences across contexts.

Benchmarking results disaggregated by sex, urbanicity, and region appear in Annex F.

TAJIK GRADE 2

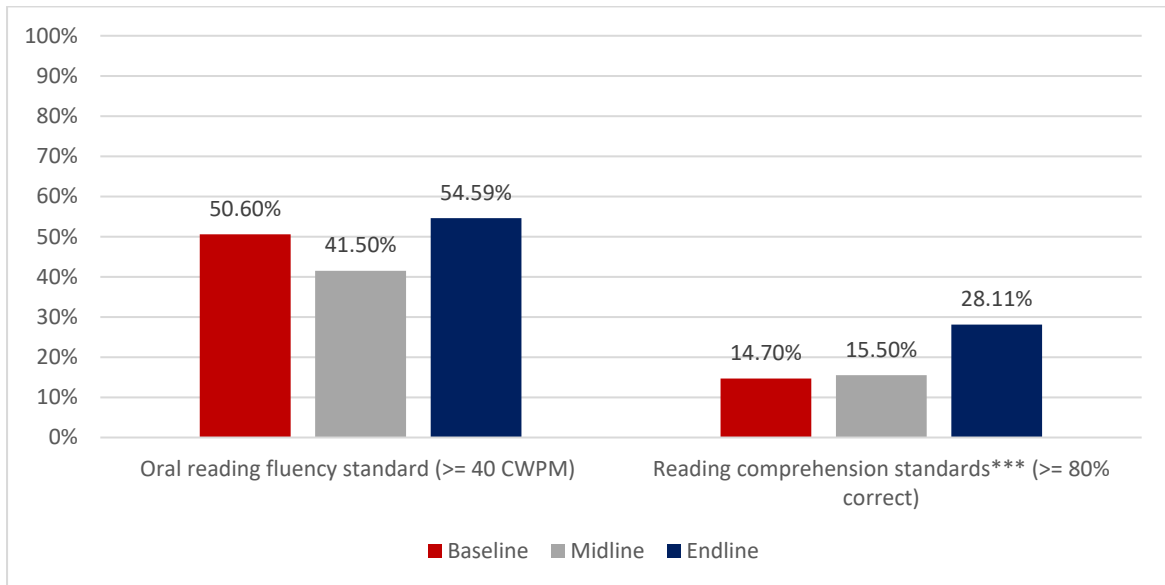
The overall distribution of ORF scores for Tajik grade 2 students illustrates how reading proficiency changed across time points, as shown in Figure 52. At endline, represented in blue, considerably fewer students had ORF scores below 30 CWPM, with more students included in the distribution’s center and right tail. The reduction of low achievers at endline primarily drove the overall improvement of students’ ORF scores.

Figure 52. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Tajik Grade 2



The percentage of Tajik grade 2 students who reached reading proficiency benchmarks at each time point is displayed in Figure 53. At endline, 54.59 percent of students attained the ORF benchmark. Although this percentage was slightly higher than at baseline, the difference was not statistically significant. At endline, 28.11 percent of students met the reading comprehension benchmark, which was statistically significantly higher than at baseline.

Figure 53. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Tajik Grade 2

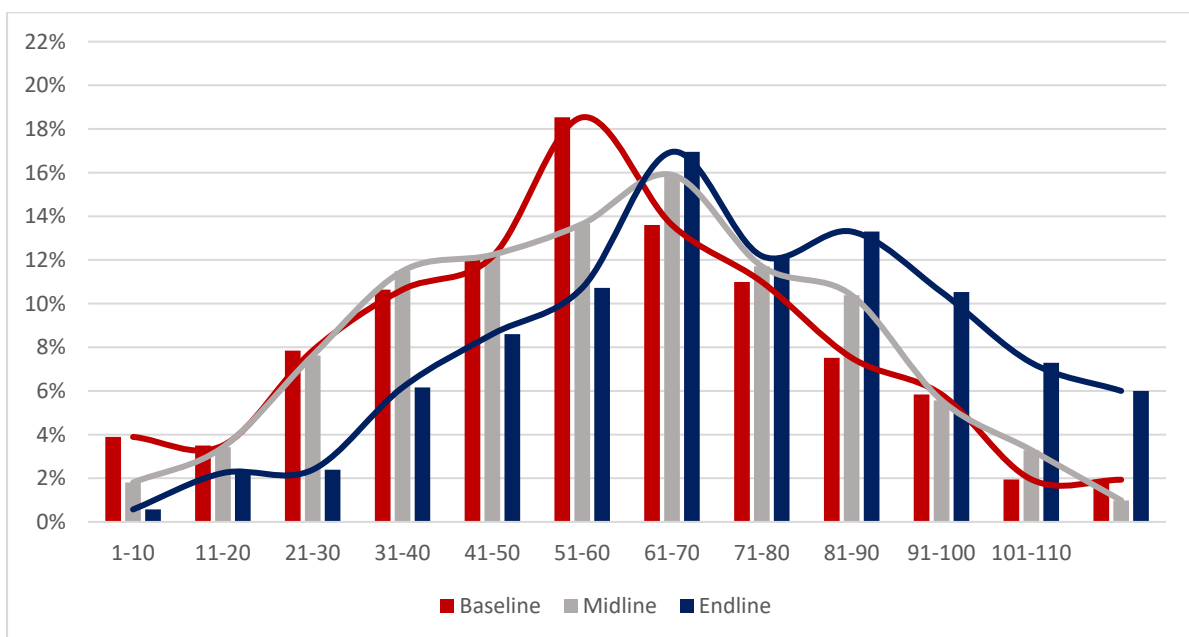


Note: Three asterisks (***) denotes differences between endline and baseline that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between endline and baseline that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between endline and baseline that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between baseline and midline was not statistically significant.

TAJIK GRADE 4

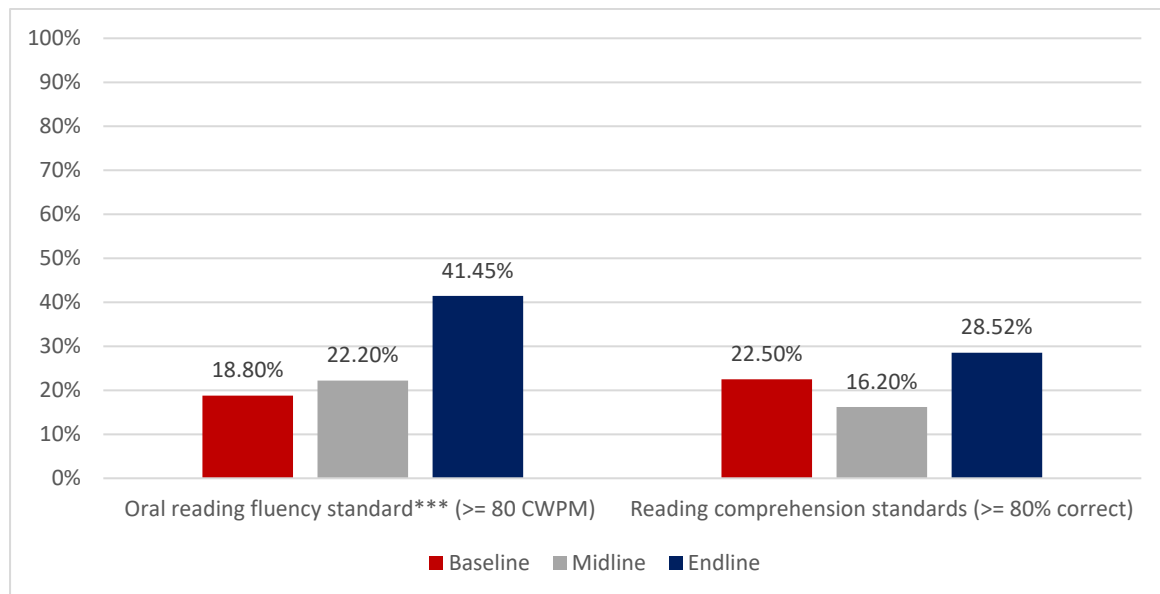
The overall distribution of ORF scores for Tajik grade 4 students, as displayed in Figure 54, shows how reading proficiency improved across evaluation points. At endline, represented in blue, considerably fewer students scored below 60 CWPM, and considerably more scored over 81 CWPM. Therefore, the overall improvement of RWM students' ORF scores was driven by both the reduction of low achievers and the increase in high achievers' proficiency.

Figure 54. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Tajik Grade 4



The percentage of Tajik grade 4 students who achieved reading proficiency benchmarks at each time point is shown in Figure 55. At endline, 41.45 percent of students met the ORF benchmark, which was statistically significantly higher than at baseline, and 28.52 percent of students met the reading comprehension benchmark. Although the proportion of students attaining the reading comprehension benchmark at endline was higher than at baseline, the difference was not statistically significant.

Figure 55. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Tajik Grade 4

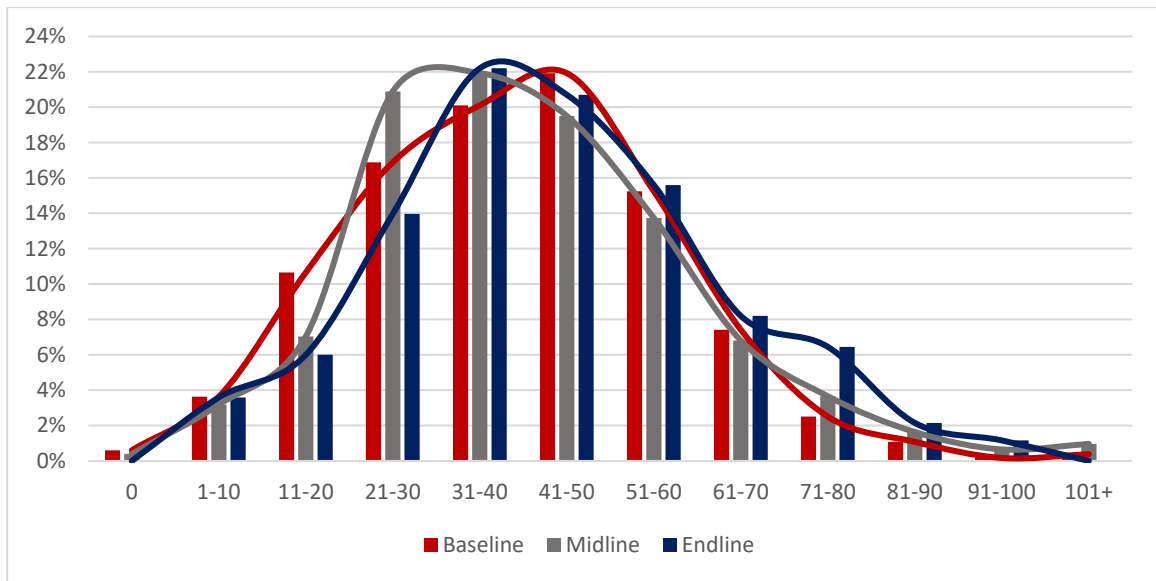


Note: Three asterisks (***) denotes differences between endline and baseline that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between endline and baseline that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between endline and baseline that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between baseline and midline was not statistically significant.

RUSSIAN GRADE 2

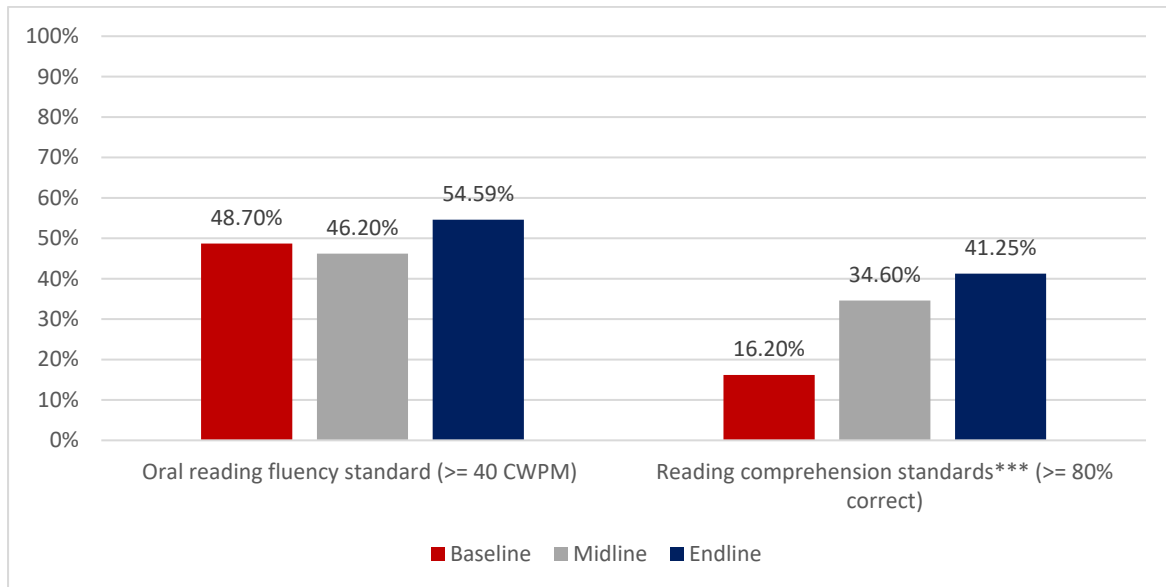
The overall distribution of ORF scores for Russian grade 2 students remained similar across time points, as illustrated in Figure 56, despite some changes in the distribution. At endline, represented in blue, considerably fewer students were part of the 21–30 CWPM interval, and considerably more were included in the 71–80 CWPM interval. These changes, however, did not result in notable changes in the distribution curves across time points.

Figure 56. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Russian Grade 2



The percentage of Russian grade 2 students who achieved reading proficiency benchmarks at each time point is illustrated in Figure 57. At endline, 54.59 percent⁶³ of students met the ORF benchmark on the ORF subtask, and 41.25 percent attained the reading comprehension benchmark. Although both percentages were higher than at baseline, the difference was only statistically significant for reading comprehension.

Figure 57. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Russian Grade 2



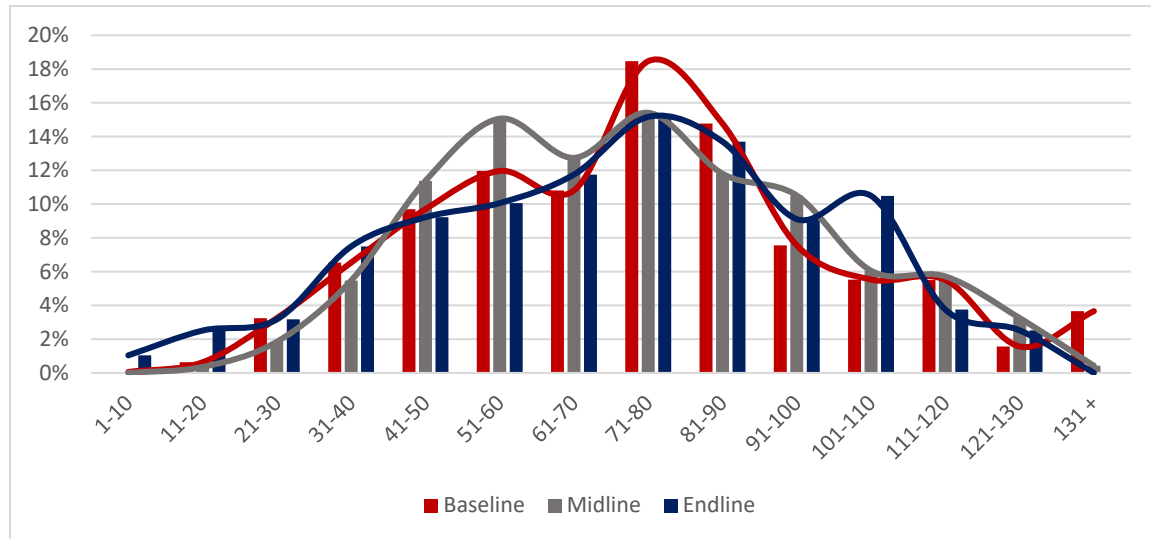
Note: Three asterisks (***) denotes differences between endline and baseline that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between endline and baseline that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between endline and baseline that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between baseline and midline was not statistically significant.

RUSSIAN GRADE 4

⁶³ Identical to Tajik grade 2 students

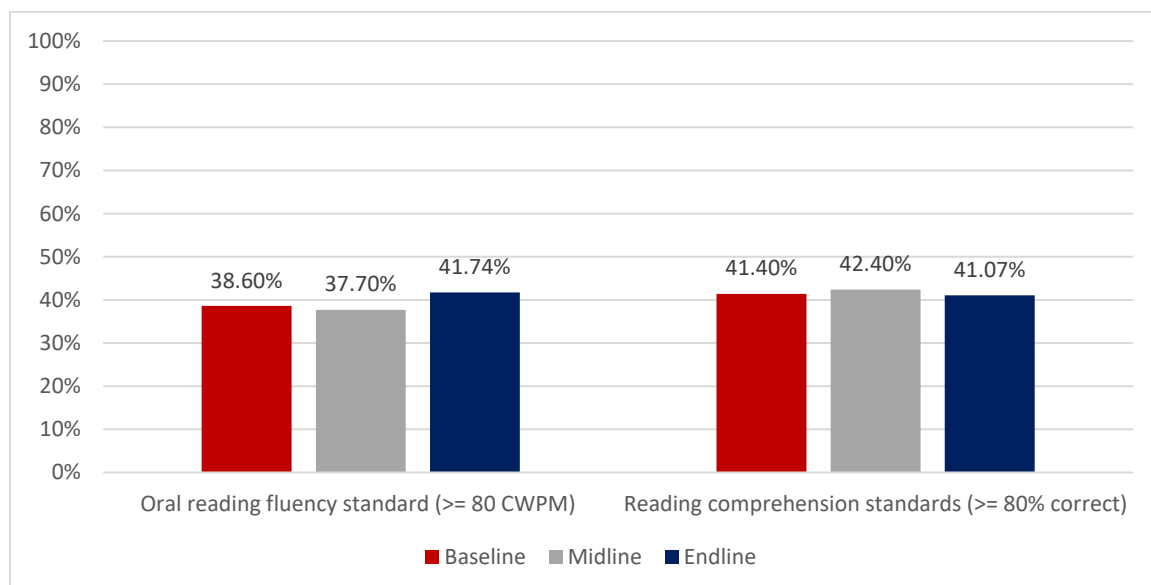
The overall distribution of ORF scores for Russian grade 4 students flattened across time points, as illustrated in Figure 58. At endline, represented in blue, shifts occurred across the distribution, with the right side being slightly more pronounced. In particular, considerably more students at endline were part of at the 101–110 CWPM interval. Overall, the endline score distribution is flatter. In other words, although scores may be distributed more evenly across the scale, no significant changes in measures of central tendency, such as the mean or median, may have occurred.

Figure 58. Distribution of Oral Reading Fluency Scores for All Students by Data Collection Point, Russian Grade 4



The percentage of Russian grade 4 students who achieved reading proficiency benchmarks at each time point is shown in Figure 59. At endline, 41.74 percent of students attained the ORF benchmark, and 41.07 percent of students did so for reading comprehension. There were no statistically significant changes over time.

Figure 59. Percentage of RWM-intervention Students Reaching Reading Proficiency Benchmarks by Grade at Baseline (2018), Midline (2019) and Endline (2021) – Russian Grade 4



Note: Three asterisks (***) denotes differences between endline and baseline that are statistically significant at $p < 0.001$. Two asterisks (**) denotes differences between endline and baseline that are statistically significant at $p < 0.01$. One asterisk (*) denotes differences between endline and baseline that are statistically significant at $p < 0.05$. No asterisks indicates that the difference between baseline and midline was not statistically significant.

DISCUSSION

All the data collected from the three EGRAs administered during the five-year RWM project provide a snapshot of reading outcomes over time and trends in reading performance among students receiving the RWM intervention. Several key findings emerge from the results at the national level, as well as among subgroups such as sex, urbanicity, region, and language spoken at home. Note that this section discusses results that were found to be statistically significant.

National-level Results Snapshot

All four groups of students—Tajik grade 2 and 4 and Russian grade 2 and 4—showed statistically significant improvements from baseline to endline. All groups except for Russian grade 4 students made significant gains on at least five subtasks. Notably, significant gains were made from baseline to endline in reading comprehension by grade 2 students in Tajik and Russian, and on the silent reading comprehension subtask by grade 2 and grade 4 students in both languages. Importantly, no group saw statistically significant declines on any subtask from baseline to endline.

The proportion of students meeting benchmarks increased significantly from baseline to endline in three of the four groups of students assessed. The proportion of students performing at or above benchmarks significantly increased from baseline to endline in three groups—grade 2 Tajik students, grade 4 Tajik students, and grade 2 Russian students. The proportion of grade 4 Russian students meeting benchmarks remained statistically unchanged from baseline to endline.

Subgroup Results by Sex, Location, Region, and Home Language

While girls generally outperformed boys at midline in Tajik, by endline, grade 4 girls and boys were performing comparably. For most of the subtasks, boys' and girls' scores improved similarly from baseline to endline. Significant gains were observed on the silent reading comprehension task for both girls and boys in Tajik and Russian at both grade levels. While both Tajik and Russian grade 2 girls scored significantly higher on ORF than their male peers at midline, this gap disappeared at endline in both languages in grade 4. Nevertheless, for both girls and boys in grade 4 Russian classrooms, scores tended to stagnate and on two subtasks, in grades 2 and 4, Russian students' scores declined.

Tajik and Russian grade 2 rural students saw the greatest proportion of gains on EGRA subtasks over time, while urban students, especially Tajik grade 2 and Russian grade 4, saw the fewest gains. All groups posted gains, and no statistical declines were found by urbanicity. Nevertheless, in grade 2 and 4 Tajik classrooms and grade 2 Russian classrooms, rural students saw more gains on EGRA subtasks than their urban counterparts.

Geographic subgroups require attention. Across regions, the proportion of student groups (disaggregated sex, urbanicity, and region) meeting the ORF benchmark was lowest in DRS, followed by Sughd. All groups highlighted in pink or red require supplementary attention. Of these, Russian grade 2 girls in rural schools in DRS struggled most at 25.54 CWPM. When examining the student group struggling most in each language by grade level by language, three common themes were home language, homework, and having family members who cannot read:

- Grade 2 Tajik boys who live in Uzbek-speaking homes and receive homework with less frequency than their peers.
- Grade 4 Tajik boys who live in Uzbek-speaking homes, receive homework with less frequency than their peers, and have brothers who cannot read
- Grade 2 Russian girls who have fathers who cannot read, did not attend preschool, and have no reading books at school to take home

- Grade 4 Russian boys who live in homes where Russian is not spoken.

Among regions, Dushanbe saw the least improvement. Though the mean score of nearly every student group in Dushanbe met or exceeded the ORF benchmark—the highest proportion of any region in this evaluation (see previous point)—students’ ORF scores in Dushanbe nevertheless stagnated or declined more than any other region in both grades in Tajik and Russian grade 4. This decline may have been due to already-high performance levels, making it more difficult to advance. Two subgroups could benefit from additional support in Dushanbe. The first group, grade 4 students in Tajik schools, could benefit from additional support with comprehension. The choice of best practices should be based on an analysis of how comprehension is currently being taught as well as student factors such as language and culture. For example, the Landscape Report on Early Literacy provides a framework for teaching four comprehension skills: predicting, explaining, summarizing, and evaluating. This report also stresses the importance of linguistic and cultural factors in building comprehension skills and lists other useful considerations, including the need for a solid foundation in word reading and comprehension, directly teaching reading comprehension strategies, ensuring time for daily literacy instruction, and availability and use of appropriate materials.⁶⁴ Intraclass correlation for ORF at endline for this group was low (0.05), indicating high variation amongst scores that suggests a subgroup within this group is struggling; however, the standard error for this group was also large enough to limit drawing firm conclusions. The second group, grade 4 students in Russian schools could benefit from additional phonics instruction, which might help improve fluency as well as increase comprehension scores. The Landscape Report cited above summarizes the research on the importance of phonics instruction (tested, in part, by assessing nonword reading) as a foundational skill for fluency which in turn predicts comprehension. The focus on phonics is strengthened by the decline in nonword reading evinced by this group from baseline to endline.

Predictors and Contextual Factors

In this section, contextual factors are examined in relation to students’ reading performance as measured by ORF rates, or CWPM.

ASSESSMENT AND EVALUATION PRACTICES

Teachers’ assessment practices, including their evaluation methods and use of assessment results, were among the most relevant variables associated with improvements in ORF. For all but Russian grade 4, teachers who used different types of evaluation methods, as well as those who used assessment results for multiple purposes, tended to correlate positively with students’ reading achievement. Still, teachers could diversify their assessment practices, as they indicated a preference for simpler and more traditional forms of evaluation. They reported oral evaluation as their most preferred method of evaluation to measure students’ progress, while less frequently reporting using end-of-term evaluations and written tests. Even fewer teachers reported using portfolios and other projects to evaluate students’ performance and progress. Future efforts should focus on encouraging teachers to use a variety of evaluation methods.

Results clearly highlight room for increasing formative uses of assessments, such as planning activities or adapting teaching. While teachers primarily reported using assessment results to either evaluate students’ understanding or grade students, fewer reported using them to plan teaching activities or adapt teaching to better suit students’ needs. Most teachers reported using assessment results in only one or two ways, and very few teachers reported using assessment results in all possible ways. More evaluation methods and more uses of assessment results were

⁶⁴ USAID, *Landscape Report on Early Literacy* (Washington, DC: USAID, 2016), 36.

consistently associated with an increase in ORF, and teachers would likely benefit from future training on how to use assessment results to plan activities or adapt teaching.

INSTRUCTIONAL PRACTICES

The quality and intensity of teaching were associated with improvements in ORF scores.

Unsurprisingly, good teaching correlated positively with improved student reading performance, and when teachers used instructional practices promoted by RWM, students' performance was stronger. For example, in both Russian grade 2 and Tajik grade 2 classes, teachers who were observed explicitly articulating the objectives of the lesson and relating classroom activities to those objectives were associated with increases of nearly 6 and nearly 5 CWPM, respectively. Teachers who included more prereading, while-reading, and post-reading activities, as well as RWM teaching strategies, were associated with modest but statistically significant increases in CWPM in Tajik grade 2. Similar results were seen in Tajik grade 4. In addition, teachers who rephrased and explained a question if a student was unable to answer it correctly—as well as teachers who encouraged such students to try harder—were generally associated with improved ORF scores.

In some instances, negative discipline measures were found to predict substantial negative results. In all grades and languages, teachers who hit students if they were unable to answer a question correctly were associated with decreases in CWPM, generally by very large magnitudes. In Tajik grade 2, teachers who sent students to the corner of the classroom if they were unable to answer a question correctly were also associated with decreases in CWPM.

Other RWM efforts, including reading corners and the use of logbooks, were associated with improvements in ORF scores, with associated improvements ranging from 5.99 to 11.59 CWPM. None of these efforts, however, were associated with improvements in Russian grade 4 ORF scores.

RESOURCES

The level of resources available in classrooms—such as students' reading textbooks and other materials, as well teacher resources such as a board, visual aids, and a lesson plan—was associated with improvements in ORF. In addition, most classrooms had 10 or more books available, and most classrooms had books provided by the project available. As for findings from the school inventory tool, having more library books for students was positively associated with ORF scores for Tajik grade 2 and 4 and Russian grade 2 and 4. The positive relationship between these resource indexes and ORF scores suggests that proficiency is resource sensitive, indicating a path to further improvement in reading outcomes.

Experienced and trained teachers tended to be associated with improved ORF scores. More experienced teachers were associated with improvements in CWPM in Tajik grade 2 and 4 and Russian grade 2 and 4. Other factors associated with increases in ORF, especially in Tajik grade 2 included teachers who advanced further in school, those who had received support at school in the past year, and those who had attended in-service training or professional development in the last year, especially in teaching reading. Similarly, directors who reported that their teachers had received training on how to teach reading outside of in-service teacher training were associated with improvements in ORF in every grade and language. These results are suggestive of a positive relationship between training activities like RWM's and improved student outcomes.

STUDENTS AND PARENTS

Speaking the LOI at home predicted improved ORF scores. For both Russian grades, speaking Russian at home was associated with improved ORF scores (a gain of 3.97 CWPM in grade 2 and 5.71 CWPM in grade 4). Similarly, for Tajik grade 2 students, speaking Tajik at home was associated with improved ORF scores (a gain of 8.98 CWPM), while speaking Uzbek at home was associated with decreased ORF scores (a decline of 7.84 CWPM). For both Tajik grades, speaking

Pamiri dialects at home was associated with large increases in ORF scores (a gain of 15.14 CWPM in grade 2 and 28.08 CWPM in grade 4), but this finding is likely related to a lurking variable uncontrolled for in the analysis, perhaps attending early childhood education.

Parental involvement predicted better reading outcomes. In all but Tajik grade 4, students with parents who reviewed their homework were associated with improvements in ORF. In both Russian grades, teachers who reported they were satisfied with parental involvement in the classroom were also associated with improvements. In both Tajik grades, schools that held regular PTA meetings were associated with improved student performance. In Russian grade 2, directors who were satisfied with the level of support the PTA provided to the school were also associated with improved student performance. In addition, having books at home was positively associated with ORF in all grades and languages.

Proportion of Students Meeting Grade-Level Benchmarks

In perhaps the most significant finding of this endline study, the proportion of students meeting the ORF benchmark improved significantly over the life of the project. For the purposes of reporting to Standard Foreign Assistance (F) Indicators ES.1-1 and ES.1-2—that students should be able to “read and understand the meaning of grade-level text”—RWM calculated the proportion of students able to read 40 CWPM in grade 2 and 80 CWPM in grade 4. The proportion of students meeting the ORF benchmark in each group increased from baseline to endline as follows:

- **Tajik grade 2:** 50.60 percent at baseline, 54.59 percent at endline
- **Tajik grade 4:** 18.80 percent at baseline, 41.45 percent at endline
- **Russian grade 2:** 48.70 percent at baseline, 54.59 percent at endline
- **Russian grade 4:** 38.60 percent at baseline, 41.74 percent at endline

RWM, LTA, and the MoES can build on this growth. Many students in Tajikistan already exceed the benchmark for their grade—and some far exceed it—as the distribution of ORF scores shows. Future efforts to improve students’ reading performance should target the geographic subgroups requiring supplementary attention.

RECOMMENDATIONS

This EGRA endline evaluation found substantial evidence of progress made by RWM over the life of the project. From baseline to endline, all four groups of students—Tajik grade 2 and 4 and Russian grade 2 and 4—showed statistically significant improvements from baseline to endline, and for three of the four groups, the proportion of students meeting ORF or comprehension benchmarks increased during that time. By endline, girls and boys were performing comparably in most cases—effectively equalizing performance among students in Tajik-speaking classrooms, where girls had been performing better at midline. Importantly, boys and girls in grades 2 and 4 in Tajik and grade 2 in Russian improved on several tasks from baseline to endline, and all students, including those in Russian grade 4, improved significantly in silent reading comprehension. Factors associated with reading gains included:

- the use of a variety of assessment types,
- having reading books at home,
- having teachers who use positive discipline strategies such as encouraging students to try harder (punitive practices such as hitting the student were negatively correlated with achievement),
- having experienced and trained teachers,
- having more materials in schools and at home, and
- having parents who review students' homework.

RWM's role in providing some of these types of training and materials suggests that the project played a significant role in improving student outcomes.

The following is a list of key considerations and recommendations.

Assessment practices correlated with performance. While gains were greater when students had teachers who used selected assessment methods, students' reading achievement was also stronger for all but Russian grade 4 when teachers reported using different methods, and were stronger for Tajik students in grade 2 and Russian students in grade 4 when teachers used results of students' oral and written assessments to adapt teaching to better suit their students' needs. (see Findings, EQ 3).

Recommendation 1: Continue to improve teachers' use of a range of formative assessment practices. Future efforts should encourage teachers to use a variety of evaluation methods, as well as strategies for using assessment results to adapt teaching to better suit their students' needs.

Professional development correlated with performance. This EGRA found a positive relationship between training activities like in-service training, especially focused on reading, and improved student outcomes (see Findings, EQ 3).

Recommendation 2: Identify strategies that could ensure ongoing professional development over the near to medium term to sustain and build on the gains realized in RWM, such as providing three to five days per year to teachers in reading strategies.

Classroom management practices correlated with performance. Though the focus of this EGRA was not classroom management, this evaluation found evidence that positive discipline strategies like encouraging students were associated with better learning outcomes, while negative strategies such as hitting had the reverse effect (see Findings, EQ 3).

Recommendation 3: Expand teachers' use and appreciation of positive discipline strategies and identify ways to monitor and correct cases of more punitive approaches.

Support at-risk student groups with targeted interventions. While high proportions of student groups (disaggregated by region, sex, urbanicity, grade and language) were struggling in DRS and Sughd and, therefore, require supplementary attention, particular attention should be paid to the lowest-scoring student groups (see Findings, EQ 2).

Recommendation 4: Support the most at-risk student groups as follows:

- Grade 2 and 4 Tajik boys: Encourage assigning homework more frequently and providing Tajik linguistic support for students who speak Uzbek at home.
- Grade 2 Russian girls: Encourage providing reading books at school to take home, promoting preschool attendance, and providing Tajik linguistic support for girls who speak Uzbek at home.
- Grade 4 Russian: Provide Russian language support tailored to the needs of linguistically diverse groups or linguistic minorities.

Since each of the four groups is characterized by a gap between languages spoken at home and in the classroom, provide linguistic support for these students by encouraging teachers to find out which languages students speak most frequently, determine students' level of fluency in their second language, and identify ways to bring students' language and culture into the classroom, especially in the early grades.⁶⁵

The stronger the language match, the stronger the reading scores. This was true for language matches for both students and teachers. ORF scores were higher for grade 2 and grade 4 students in Tajik-language schools whose native language was Pamiri (speak Pamiri at home), and higher for these students when their teachers also reported speaking Pamiri at home. Conversely, ORF scores were lower for grade 2 students in Russian-language schools whose native language was Pamiri (speak Pamiri at home), and lower for grade 2 and grade 4 students in Russian-language schools when their teachers reported speaking Pamiri at home.

Recommendation 5: Assess the discrepancies in Russian-language schools between language of instruction, home language of students, and home language of teachers. Where discrepancies are high, consider providing additional language and cultural support to those students and teachers.

Conditions at home and in the classroom matter. This evaluation found that some students showed greater improvement when there were books at home, when parents reviewed their homework, and when parents were involved in the school. It also found positive correlations between performance and the presence of materials in schools such as appropriate textbooks, language exercise books, and vocabulary books (see Findings, EQ 4).

Recommendation 6: Improve the provision of resources at school and in the home.

The selection of these resources should be made in reference to the classroom index, school environment index, and lists of reading materials at home identified in this evaluation.

⁶⁵ For additional explanation and strategies, see Save the Children/UK (2009) Steps Toward Learning: A guide to overcoming language barriers in children's education. <https://resourcecentre.savethechildren.net/library/steps-towards-learning-guide-overcoming-language-barriers-childrens-education>

ANNEXES

ANNEX A: OPERATIONAL DATA COLLECTION

This annex details what occurred during endline operational data collection, including enumerator training and selection of enumerators; procedures followed during EGRA administration, including sampling and school replacement; and data entry and cleaning.

ENUMERATOR TRAINING

RWM took precautions to minimize risk of COVID-19 during both training and data collection. All participants were test for COVID before training. No international STTA traveled for the training, and instead STTA helped conduct the supervisors and observers training remotely. Training was led in-person by STS Assessment Specialist Adiba Kosimova, a highly experienced EGRA trainer in Dushanbe.

Data collection teams included a supervisor, a classroom observer and two enumerators. These three groups of people were trained separately over four sets of training as follows. Prior to training, both Tangerine version 3 and Ona/Open Data Kit Collect applications had been installed on each enumerator's tablet. Tablets were checked to ensure that reinstallation did not cause any programming issues. RWM distributed tablets to EGRA and SSME supervisors, including fully programmed back up tablets in case enumerators encountered tablet issues in the field that could not be resolved remotely.

First, a training of EGRA trainers was held March 24-16 in Dushanbe. This training prepared a set of experienced, regionally based trainers who included the RWM Regional M&E specialists and 2 external trainers to lead enumerator trainings in each region of the country. These trained individuals also served as Quality Control Officers (QCO).

Second, the supervisors and QCO training took place March 28-29, also in Dushanbe. Both supervisors and QCOs reviewed, familiarized themselves with, and practiced the SSME tools, student sampling procedures, and team management and reporting practices.

Third, a classroom observation training took place in Dushanbe March 31-April 1 to train one observer per team on administration of the classroom observation procedures.

Finally, the QCOs traveled to their home regions to deliver regionally based EGRA training sessions over four consecutive days, including one school day to practice the tools and procedures in a school environment to provide enumerators with the opportunity to practice in real-world conditions. These regional trainings focused on introducing the EGRA subtasks and their administration on the data collection software Tangerine, as well as familiarizing enumerators and supervisors with their roles, responsibilities, and EGRA protocols. Enumerators were also trained to complete inter-rater reliability (IRR) measures daily during data collection. These regional EGRA enumerator trainings took place between April 2 and 10, depending on the region.

Seventy-three EGRA enumerators, supervisors and observers were trained in total; 64 were involved in data collection and 5 enumerators stayed in reserve. Two trained enumerators could not join data collection for personal reasons and two did not pass the minimum requirements for data collection, as described below.

SELECTION OF ENUMERATORS

RWM selected enumerators based on performance in three scored enumerators accuracy quizzes.

During these enumerator accuracy quizzes, all enumerators listened to a video and audio-recorded role play that was acted out in person using a “gold script” that included planned incorrect and challenging student responses. The percentage of items for which each enumerator’s scoring was correct was calculated. All enumerators selected for data collection scored at least 90.00 percent of items accurately according to the gold script.

DATA COLLECTION

Immediately following the regional endline trainings, STS issued updated to the Tangerine version 3 software on each enumerator’s tablet.

Endline data collection took place in every region of Tajikistan in April and May 2021. Sixteen teams—11 Tajik-speaking and 5 Russian-speaking—collected data from April 7 to May 3. Each team visited one school per day. Each of the 16 teams consisted of one supervisor, who led sampling and administered the teacher interview, director interview, and school inventory; one lesson observer; and two EGRA enumerators. In total, 202 schools were assessed. Table A1 provides the sampled number of students by region.

Table A1. Sample by Region

Region	Tajik Grade 2	Tajik Grade 4	Russian Grade 2	Russian Grade 4
DRS	234	237	70	80
Dushanbe	240	240	129	130
GBAO	102	104	0	0
Khatlon-Bokhtar	296	296	71	70
Khatlon-Kulob	211	214	10	10
Sughd	259	257	289	309

Each supervisor arranged transportation for their team to and from the school. The teams met with the school director at the beginning of the school day, typically by 8:30 a.m. Upon arrival at the school, the supervisor introduced themselves and the purpose of the visit to the school director. They also worked with the school director to identify an area where students could take the assessment. Supervisors then sampled classes and students, as described in the following section. Once the students were selected, supervisors guided them to the space designated for the testing. Each enumerator tested one student at a time.

This process was completed for grade 4 after the grade 2 assessment was complete. In several schools, grade 4 students were drawn from the school’s second shift given the time required to complete the assessments in grade 2.

WITHIN-SCHOOL SAMPLING OF STUDENTS

At each school, 10 students—five girls and five boys—in each grade were assessed per day. The students were randomly selected using the following steps:

- In cases where there were multiple classes per grade, the supervisor randomly selected a class to assess.
- The supervisor visited the randomly selected class and asked the girls present in the class to count numbers in order, from “one” to the number of girls present.
- Using a random number generator application on the tablet, the supervisor input the number of girls in the class and generated five random numbers, with no numbers repeated.
- The supervisor read the numbers aloud, and the girls whose numbers were called were taken to the assessment waiting area.
- The process described above was repeated for boys to select five boys randomly.

In this manner, each team assessed per day 10 students in grade 2 and then 10 students in grade 4, for a total of 20 students per team per day.

If a student declined to participate in the assessment, enumerators were instructed to select the next student. For example, if a boy with the number 5 refuses to participate, they selected the boy with the number 6.

REPLACEMENT PROCEDURES

Prior to endline data collection, seven Russian-language schools could no longer be included because they had closed, ceased teaching in Russian, or did not have grade 2 or 4 students. In addition, six schools that had previously not taught in Russian began using it as the LOI and were added to the endline sample. Only one Tajik-language school, in GBAO, was removed from the endline sample as it had participated in RWM activities.

DATA QUALITY ASSURANCE PROCESSES

RWM implemented a variety of strategies to track the progress of data collection, as well as provide oversight and quality assurance checks on the EGRA and SSME data collection. Each region was assigned a quality control officer (QCO), who visited every team in the assigned region at least once. QCOs reported any issues in the number and type of data collected daily to the assessment specialist, and these reports were cross-referenced against uploaded data. Any discrepancies were noted, and follow-up calls were made by the assessment specialist to the supervisor to resolve and document issues. Issues and discrepancies were addressed during the data cleaning process. STS staff in Tajikistan also visited 14 schools in person to observe data collection. Thus, each data collection team had at least one on-site spot check, and many teams were visited several times.

In addition, supervisors for each team provided on-the-ground oversight of data collection for their teams in the field, including completing reports that were sent daily as part of the SSME. Throughout operational data collection, RWM followed the guidance laid out in the *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*, also known as the EGRA Toolkit 2.0, by regularly uploading and reviewing data in order to better manage and track data collection issues and progress.⁶⁶ QCOs ensured data collection procedures were followed and submitted daily reports that logged any discrepancies in the number and type of data collected that differed from the intended sample.⁶⁷ These reports were later cross-referenced against the uploaded data in Tangerine and Ona.⁶⁸ Disposition codes were applied to categorize the various issues or problems that emerged during the data collection process. These codes were used in determining cleaning rules that were incorporated into the database using syntax to clean the data accordingly. These coding and flagging procedures helped to ensure the various and nuanced contexts of data collection at the school level were sufficiently cataloged and considered during the data cleaning, analysis, and reporting process.

ANNEX B: ANALYTIC METHODS

EQUATING

Equating techniques vary according to their data collection design and to the statistical methods chosen. The three possible data collection designs for equating include

⁶⁶ RTI International. *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*. (Washington, DC: United States Agency for International Development, 2015). p. 103.

⁶⁷ These reports documented the school demographics, type and number of each assessment or questionnaire collected, status of data upload, and any other issues or challenges encountered that day in the school.

⁶⁸ Tangerine is a commonly-used application to collect EGRA data. Ona is a mobile data collection application, built on an Open Data Kit Collect platform, used by RWM to collect SSME data.

- *Single groups design*, where the same group of students takes two forms;
- *Randomly equivalent groups design*, where students are randomly assigned to one of two forms; and
- *Non-equivalent groups design*, where students represent different populations but the forms they take have common items, known as “anchor items,” to establish a relationship between both groups of students.

In terms of the possible methodologies, equating can be *classical-test theory* based or *item-response theory* based, and within each group, there is a gamut of techniques that can be applied depending on the number of students, items, and other considerations around the sample. The EGRA Toolkit 2.0 provides guidance with on which equating methodologies to use under which circumstance. Following those recommendations, linear equating was used for the ORF subtask. The data collection design was a single-group design in which the same students read both the baseline and midline ORF passages, which enabled analysts to directly attribute differences in difficulty to the items included in each form.

Equating was not conducted on the other subtasks. For initial sound identification, the same items that were used at baseline were used at midline and endline. For letter name identification, familiar word reading, and nonword reading, the items used at baseline and midline were re-randomized at endline. For reading comprehension, silent reading comprehension, and listening comprehension, equating was not done because the number of items was few and the small amount of variance among the items made equating scores with precision difficult; this decision is supported by EGRA Toolkit 2.0 guidance. Instead, these subtasks underwent targeted changes in word choice while keeping the overall story structure and difficulty as close to baseline as possible.

WEIGHTING

Sampling weights were computed separately for girls and boys within each school and stratum. The sampling weights were built under the assumption that the sampling design was a three-level stratified clustered sampling:

1. Schools were selected within strata—LOI, type of school, region;
2. Classrooms were selected in each school for each sampled grade level; and
3. Five girls and five boys were selected in each classroom.

To compute the sampling weights, the analysts needed the following information about all the schools in the relevant population:

- Type of school—RWM+QRP, RWM-only
- Region
- LOI used in the classrooms
- Number of grade 2 and grade 4 classrooms in each school, per language
- Number of students in each grade 2 and grade 4 classroom, per language and gender

The computation of the weights for student “i” in classroom “j” in school “k” was as follows:

$$Final\ Weight_{ijk} = School\ Weight_k * Classroom\ Weight_j * Student\ Weight_i$$

STS collected up-to-date information from the project. Most schools had information about the number of classrooms and the number of students per grade level; whenever these were not available, values were derived through mean imputation at the stratum level. Adjustments were made anytime a classroom reported having less than 5 boys and/or 5 girls. While few schools were closed or opened between midline and endline data collection, STS decided to use midline School Weights as the changes were minimal. New schools or replacement schools took the values

corresponding to schools in their corresponding stratum. Weights were computed using Stata 16.

DATA ANALYSIS

The primary objective of the midline analysis was to answer the study's research questions:

1. How do Tajik and Russian reading outcomes of students in grade 2 and grade 4 nationally vary by subgroups and across time points (baseline, midline, endline)?
2. Which RWM geographic subgroups require supplementary attention, and what kind of supplementary attention is required?
3. Do materials provided by RWM or teacher instructional practices supported by RWM serve as predictors of student reading in Tajik and Russian reading outcomes of students in grade 2 and grade 4 in schools supported by RWM? If yes, which materials or practices are predictors?
4. Which contextual factors or other classroom measures are predictors of Tajik and Russian reading outcomes of students in grade 2 and grade 4 in schools supported by RWM?
5. What proportion of students can read and understand the meaning of grade-level text at each time point in schools served by RWM in grade 2 and grade 4 in Tajik and in Russian?

To answer these questions, analysts merged EGRA and SSME data by grade and language and conducted final analysis on the merged datasets. Four merged datasets were used to respond to the different research questions Russian grade 2, Russian grade 4, Tajik grade 2, and Tajik grade 4. Before responding to each research question, three analyses were conducted:

EGRA Scoring and Analysis

The first step of the EGRA analysis was to score EGRA data, and creating different types of scores: total scores, fluency scores, percent correct scores, zero scores, and benchmark scores, when applicable. The second step of the EGRA analysis was to examine EGRA results for the overall sample. For each subtask, mean score (fluency or percent correct, depending on the task) and zero score results were summarized for baseline, midline and endline separately. Benchmarking results for ORF and reading comprehension were also summarized for each data collection point. Mean score, zero score, and benchmarking results for the overall sample were then examined for statistically significant differences at baseline and endline, as well as statistically significant differences in changes from baseline to endline, using t-tests. The third step was to examine EGRA results by sex. Mean score, zero score, and benchmarking results were summarized for baseline, midline and endline by sex. Results were then examined for statistically significant differences by sex at baseline and endline as well as statistically significant differences in differences by sex across timepoints. The fourth step was to examine EGRA results by urbanicity. Mean score, zero score, and benchmarking results were summarized for baseline, midline and endline by urbanicity. Results were then examined for statistically significant differences by urbanicity at baseline and endline as well as statistically significant differences in differences between urban and rural schools across timepoints. The fifth step was to examine EGRA results by region. Mean score, zero score, and benchmarking results were summarized for baseline, midline and endline by region. Results were then examined for statistically significant differences by region at baseline and endline. The results from these analyses were used to respond to research questions 1, 2 and 5.

SSME Analysis

The steps in this subsection were applied separately to Russian-language and Tajik-language schools. The first step of the SSME analysis was to examine the school inventory data at midline. Results from the school inventory were summarized overall. The second step was to examine the classroom observation data at midline. Results from the classroom observation were summarized overall and by grade. Data for each segment of each observation—there were two segments per observation—

were treated as individual observations. The exception to this was the overall scores for lesson facilitation, checking for understanding, and providing feedback. For each of these items, the scores were averaged across segments for each teacher, and the average teacher scores were generated overall and by grade. The third step was to examine the classroom inventory data at midline. Results from the classroom observation were summarized overall. The fourth step was to examine the teacher questionnaire data at midline. Results from the teacher questionnaires were summarized overall. The fifth step was to examine the school director questionnaire data at midline. Results from the head teacher interview were summarized overall.

Predictor Analysis

The steps in this subsection were applied to each of the four datasets—Russian grade 2, Russian grade 4, Tajik grade 2, and Tajik grade 4—separately. Using weighted hierarchical linear regression (Tajik grade 2, Russian grade 2, Russian grade 4), with students nested within schools, or weighted linear regression (Tajik grade 4), each variable from every single survey (SSME surveys and student surveys) were evaluated for their ability to predict ORF scores. After a first round of analysis, a list of relevant variables was identified by survey and grade/language; these are the variables reported in research questions 3 and 4.

Classroom Observation Analysis: Notes on Variables Used for Comparison Between Midline and Endline

Variable	Measures	Midline notes	Endline notes
INTRO1	4a. The teacher explicitly articulates the objectives of the lesson and relates classroom activities to the objectives	Measured throughout whole lesson	Measured during introduction of lesson only
INTRO2	4b. The teacher's explanation of content is clear	Measured throughout whole lesson	Measured during introduction of lesson only
INTRO3	4c. The teacher makes connections in the lesson that relate to other content knowledge or student's daily lives	Measured throughout whole lesson	Measured during introduction of lesson only
PREREADING2_6	4d. The teacher models reading the text aloud	Measured throughout whole lesson	Measured during pre-reading activity only
SKILL_PHONCONSC	5a. Phonemic consciousness (working with sounds)	Measured throughout whole lesson	Measured during while-reading activity only
SKILL_FLUENTREADING	5a. Fluent reading	Measured throughout whole lesson	Measured during while-reading activity only
SKILL_VOCAB	5a. Vocabulary (passive and/or active vocabulary)	Measured throughout whole lesson	Measured during while-reading activity only
SKILL_READINGCOMP	5a. Reading comprehension	Measured throughout whole lesson	Measured during while-reading activity only
PREREADING_COMPOSITE	5b. Teacher did following prereading activities: Work with the text title, Show the pictures in the text, Define	Measured as one question	Measured as individual questions and drafted into an overall indicator using only the items

Variable	Measures	Midline notes	Endline notes
	key words, Play music, Create a riddle, Ask questions related to text		listed in the label from midline (excludes play music and create a riddle)
WHILEREADING_COMPOSITE	5c. Teacher did following while reading text: Modeling reading; Reading with comments; Technique 'Dialogue with the author'; Graphic organizers	Measured as one question	Measured as individual questions and drafted into an overall indicator using only the items listed in the label from midline (excludes technique "dialogue with the author")
POSTREADING_COMPOSITE	5d. Did the teacher following post reading activities and techniques: Drawing; Writing; Student created activities; Role play	Measured as one question	Measured as individual questions and drafted into an overall indicator using only the items listed in the label from midline.
SUPPLEMENTARY_MATS	5e. The teacher integrates supplemental materials into the lesson (Project books, Other books, Question Cards, handouts and digital texts)	Measured as one question for entire lesson	Measured as individual questions during while-reading and post-reading and drafted into an overall indicator
READING_PRACTICE	5f. Teacher provides opportunities for reading practice	Measured as 4 questions to be measured during whole lesson	Measured as 2 questions during while-reading activity
POSTREADING2_7	6a. After reading, the teacher uses questions, prompts or other strategies to determine students' level of understanding	Measured throughout whole lesson	Measured during post-reading activity only

ANNEX C: RELIABILITY MEASURES

CORRELATIONS

Tables C1 through C12 show the Pearson bi-variate correlation results for the baseline (2018), midline (2019) and endline (2021) EGRA subtasks by grade and language.

Table C1: Correlations between Baseline (2018) EGRA Subtasks, Tajik Grade 2

Subtask	Letter name identification	Initial sound identification	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Letter name identification	1.00							
Initial sound identification	0.12***	1.00						
Familiar word reading	0.48***	0.26***	1.00					

Nonword reading	0.46***	0.27***	0.88***	1.00				
Oral reading fluency	0.47***	0.26***	0.95***	0.89***	1.00			
Reading comprehension	0.38***	0.24***	0.63***	0.59***	0.67***	1.00		
Silent reading comprehension	0.32***	0.18***	0.39***	0.38***	0.40***	0.54***	1.00	
Listening comprehension	0.23***	0.15***	0.17***	0.15***	0.18***	0.31***	0.38***	1.00

Note: Three asterisks (***) indicates a statistical significance of $p < 0.001$.

Table C2: Correlations between Midline (2019) EGRA Subtasks, Tajik Grade 2

Subtask	Letter name identification	Initial sound identification	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Letter name identification	1.00							
Initial sound identification	0.08**	1.00						
Familiar word reading	0.50***	0.25***	1.00					
Nonword reading	0.44***	0.26***	0.89***	1.00				
Oral reading fluency	0.47***	0.25***	0.94***	0.91***	1.00			
Reading comprehension	0.36***	0.23***	0.71***	0.66***	0.73***	1.00		
Silent reading comprehension	0.23***	0.16***	0.46***	0.46***	0.48***	0.56***	1.00	
Listening comprehension	0.11***	0.07*	0.17***	0.15***	0.18***	0.34***	0.38***	1.00

Note: Two asterisks (**) indicates a statistical significance of $p < 0.01$. Three asterisks (***) indicates a statistical significance of $p < 0.001$.

Table C3: Correlations between Endline (2021) EGRA Subtasks, Tajik Grade 2

Subtask	Letter name identification	Initial sound identification	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Letter name identification	1							
Initial sound identification	0.25*	1						
Familiar word reading	0.44*	0.20*	1					
Nonword reading	0.44*	0.21*	0.83*	1				
Oral reading fluency	0.43*	0.22*	0.95*	0.82*	1			
Reading comprehension	0.35*	0.25*	0.62*	0.53*	0.68*	1		
Silent reading comprehension	0.35*	0.29*	0.43*	0.44*	0.45*	0.58*	1	
Listening comprehension	0.24*	0.19*	0.18*	0.19*	0.21*	0.42*	0.43*	1

Table C4: Correlations between Baseline (2018) EGRA Subtasks, Tajik Grade 4

Subtask	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Familiar word reading	1.00					
Nonword reading	0.87***	1.00				
Oral reading fluency	0.92***	0.88***	1.00			
Reading comprehension	0.53***	0.53***	0.60***	1.00		
Silent reading comprehension	0.39***	0.40***	0.43***	0.53***	1.00	
Listening comprehension	0.31***	0.32***	0.35***	0.55***	0.51***	1.00

Note: Three asterisks (***) indicates a statistical significance of $p < 0.001$.

Table C5: Correlations between Midline (2019) EGRA Subtasks, Tajik Grade 4

Subtask	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Familiar word reading	1.00					
Nonword reading	0.87***	1.00				
Oral reading fluency	0.92***	0.88***	1.00			
Reading comprehension	0.59***	0.57***	0.67***	1.00		
Silent reading comprehension	0.38***	0.39***	0.41***	0.49***	1.00	
Listening comprehension	0.24***	0.24***	0.29***	0.45***	0.44***	1.00

Note: Three asterisks (***) indicates a statistical significance of $p < 0.001$.

Table C6: Correlations between Endline (2021) EGRA Subtasks, Tajik Grade 4

Subtask	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Familiar word reading	1					
Nonword reading	0.78*	1				
Oral reading fluency	0.88*	0.84*	1			
Reading comprehension	0.45*	0.52*	0.60*	1		
Silent reading comprehension	0.45*	0.44*	0.50*	0.58*	1	
Listening comprehension	0.25*	0.29*	0.33*	0.54*	0.54*	1

Table C7. Correlations between Baseline (2018) EGRA Subtasks, Russian Grade 2

Subtask	Letter name identification	Initial sound identification	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Letter name identification	1.00							
Initial sound identification	0.08*	1.00						
Familiar word reading	0.57***	0.18***	1.00					
Nonword reading	0.52***	0.20***	0.82***	1.00				
Oral reading fluency	0.55***	0.20***	0.91***	0.85***	1.00			
Reading comprehension	0.39***	0.22***	0.60***	0.53***	0.69**	1.00		
Silent reading comprehension	0.26***	0.19***	0.34***	0.31***	0.43***	0.59***	1.00	
Listening comprehension	0.22***	0.15***	0.24***	0.18***	0.29***	0.48***	0.63***	1.00

Note: One asterisk (*) indicates a statistical significance of $p < 0.05$. Two asterisks (**) indicates a statistical significance of $p < 0.01$. Three asterisks (***) indicates a statistical significance of $p < 0.001$.

Table C8. Correlations between Midline (2019) EGRA Subtasks, Russian Grade 2

Subtask	Letter name identification	Initial sound identification	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Letter name identification	1.00							
Initial sound identification	0.09*	1.00						
Familiar word reading	0.56***	0.02	1.00					

Nonword reading	0.55***	0.04	0.85***	1.00				
Oral reading fluency	0.52***	0.02	0.90***	0.85***	1.00			
Reading comprehension	0.44***	0.04	0.59***	0.56***	0.66***	1.00		
Silent reading comprehension	0.27***	0.04	0.42***	0.39***	0.48***	0.63***	1.00	
Listening comprehension	0.22***	0.01	0.23***	0.24***	0.30***	0.54***	0.55***	1.00

Note: One asterisk (*) indicates a statistical significance of $p < 0.05$. Two asterisks (**) indicates a statistical significance of $p < 0.01$. Three asterisks (***) indicates a statistical significance of $p < 0.001$.

Table C9. Correlations between Endline (2021) EGRA Subtasks, Russian Grade 2

Subtask	Letter name identification	Initial sound identification	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Letter name identification	1							
Initial sound identification	0.16*	1						
Familiar word reading	0.55*	0.17*	1					
Nonword reading	0.52*	0.16*	0.84*	1				
Oral reading fluency	0.50*	0.19*	0.93*	0.84*	1			
Reading comprehension	0.35*	0.22*	0.68*	0.60*	0.74*	1		
Silent reading comprehension	0.16*	0.21*	0.42*	0.39*	0.47*	0.64*	1	
Listening comprehension	0.22*	0.25*	0.26*	0.24*	0.30*	0.54*	0.52*	1

Table C10. Correlations between Baseline (2018) EGRA Subtasks, Russian Grade 4

Subtask	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Familiar word reading	1.00					
Nonword reading	0.81***	1.00				
Oral reading fluency	0.88***	0.82***	1.00			
Reading comprehension	0.49***	0.39***	0.58***	1.00		
Silent reading comprehension	0.28***	0.21***	0.39***	0.52***	1.00	
Listening comprehension	0.15***	0.09*	0.25***	0.44***	0.45***	1.00

Note: One asterisk (*) indicates a statistical significance of $p < 0.05$. Three asterisks (***) indicates a statistical significance of $p < 0.001$.

Table C11. Correlations between Midline (2019) EGRA Subtasks, Russian Grade 4

Subtask	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Familiar word reading	1.00					
Nonword reading	0.75***	1.00				
Oral reading fluency	0.85***	0.79***	1.00			
Reading comprehension	0.56***	0.48***	0.66***	1.00		
Silent reading comprehension	0.38***	0.34***	0.45***	0.64***	1.00	
Listening comprehension	0.30***	0.25***	0.39***	0.57***	0.61***	1.00

Note: Three asterisks (***) indicates a statistical significance of $p < 0.001$ and one asterisk (*) indicates a statistical significance of $p < 0.05$

Table C12. Correlations between Endline (2021) EGRA Subtasks, Russian Grade 4

Subtask	Familiar word reading	Nonword reading	Oral reading fluency	Reading comprehension	Silent reading comprehension	Listening comprehension
Familiar word reading	1					
Nonword reading	0.65*	1				

Oral reading fluency	0.85*	0.76*	1			
Reading comprehension	0.30*	0.50*	0.58*	1		
Silent reading comprehension	0.11*	0.27*	0.28*	0.50*	1	
Listening comprehension	0.15*	0.25*	0.27*	0.38*	0.54*	1

EGRA INTER-RATER RELIABILITY

To record the level of enumerator agreement throughout data collection, enumerators undertook daily measures of IRR according to the following protocol. Enumerators worked in pairs to assess the first student of the day. During this first assessment, one enumerator acted as the “main enumerator,” administering the EGRA and scoring the student responses in his or her tablet. The second enumerator simultaneously listened and also scored the student responses on his or her tablet. Once the assessment was completed, and the student had returned to class, the two enumerators compared and discussed their scoring of the student’s responses. Any points of disagreement or difference in marking were brought to the attention of supervisors and discussed during team meetings. Enumerator pairs took turns playing the role of the main enumerator from each day to the next. Results are presented in Tables C13 – C16.

Table C13: Baseline (2018), Midline (2019) and Endline (2021) IRR Results, Tajik Grade 2

Grade 2, Tajik	Baseline (2018) N=141		Midline (2019) N=75		Endline (2021) N=77	
	Percentage agreement	Kappa	Percentage agreement	Kappa	Percentage agreement	Kappa
Letter name identification	95.30%	0.84	97.70%	0.84	98.74%	0.88
Initial sound identification	98.70%	0.97	97.50%	0.94	99.35%	0.96
Familiar word reading	99.00%	0.97	99.10%	0.97	99.82%	0.93
Nonword reading	98.10%	0.94	99.10%	0.98	99.66%	0.97
Oral reading fluency	98.60%	0.96	99.20%	0.98	99.51%	0.90
Reading comprehension	99.40%	0.95	98.90%	0.98	99.48%	0.98
Silent reading comprehension	N.A.	N.A.	N.A.	N.A.	100.00%	1.00

Listening comprehension	98.60%	0.97	100.00%	1.00	99.35%	1.00
-------------------------	--------	------	---------	------	--------	------

Table C14: Baseline (2018), Midline (2019) and Endline (2021) IRR Results, Tajik Grade 4

Grade 4, Tajik	Baseline (2018) N=60		Midline (2019) N=70		Endline (2021) N=61	
	Percentage agreement	Kappa	Percentage agreement	Kappa	Percentage agreement	Kappa
Familiar word reading	98.90%	0.96	99.70%	0.99	99.84%	0.97
Nonword reading	98.40%	0.96	99.20%	0.98	99.41%	0.94
Oral reading fluency	98.80%	0.95	99.50%	0.99	99.26%	0.96
Reading comprehension	99.00%	0.97	100.00%	1.00	98.36%	0.97
Silent reading comprehension	N.A.	N.A.	N.A.	N.A.	99.34%	1.00
Listening comprehension	98.70%	0.96	98.60%	0.97	100.00%	1.00

Table C15: Baseline (2018), Midline (2019) and Endline (2021) IRR Results, Russian Grade 2

Grade 2, Russian	Baseline (2018) N=28		Midline (2019) N=37		Endline (2021) N=28	
	Percentage agreement	Kappa	Percentage agreement	Kappa	Percentage agreement	Kappa
Letter name identification	99.20%	0.96	98.80%	0.91	95.68%	0.72
Initial sound identification	99.30%	0.99	98.60%	0.97	96.79%	0.93
Familiar word reading	99.40%	0.98	99.70%	0.99	100.00%	1.00
Nonword reading	99.60%	0.99	99.20%	0.97	99.86%	0.99
Oral reading fluency	99.70%	0.99	99.90%	1.00	99.86%	1.00
Reading comprehension	100.00%	1.00	99.50%	0.99	100.00%	1.00
Silent reading comprehension	N.A.	N.A.	N.A.	N.A.	100.00%	1.00
Listening comprehension	100.00%	1.00	98.90%	0.98	99.11%	0.98

Table C16: Baseline (2018), Midline (2019) and Endline (2021) IRR Results, Russian Grade 4

Grade 4, Russian	Baseline (2018) N=12		Midline (2019) N=25		Endline (2021) N=24	
	Percentage agreement	Kappa	Percentage agreement	Kappa	Percentage agreement	Kappa
Familiar word reading	100.00%	1.00	99.40%	0.97	99.33%	0.95
Nonword reading	99.30%	0.98	98.50%	0.95	99.75%	0.98
Oral reading fluency	100.00%	1.00	99.10%	0.96	99.84%	0.97
Reading comprehension	100.00%	1.00	100.00%	1.00	98.33%	0.95
Silent reading comprehension	N.A.	N.A.	N.A.	N.A.	100.00%	1.00
Listening comprehension	100.00%	1.00	99.20%	0.99	100.00%	1.00

LESSON OBSERVATION INTER-RATER RELIABILITY

To record the level of lesson observer agreement throughout data collection, STS's assessment specialist visited eleven data collection teams to record a parallel lesson observation. After the lesson was completed, the two data collectors compared and discussed their scoring of the lesson. Any noteworthy points of disagreement or difference in marking were discussed with lesson observers on other teams. Results are presented in Table C17.

Table C17: Endline (2021) Lesson Observation IRR Results

	Endline (2021) N=11	
	Percentage agreement	Kappa
Classroom Observation	73.16%	0.34

ANNEX D: INTRAClass CORRELATIONS

The *intraclass correlation* (ICC), or the ICC coefficient, is a descriptive statistic that describes how strongly units in the same group resemble each other. A value above 0.1 is generally interpreted as indicating enough similarity between values from the same group. This means that analyses need to account for the hierarchical structure of the data to avoid a misspecification of standard errors: instead of using linear regression, analysts need to use hierarchical linear regression, etc. ICC levels based on ORF scores are presented in tables D1-D4.

Table D1. Intraclass Correlation Coefficients for Oral Reading Fluency, Tajik Grade 2

Time period	ICC	SE	95% Confidence Interval	
			Lower	Upper
Baseline (2018)	0.28	0.03	0.22	0.34
Midline (2019)	0.28	0.03	0.22	0.35
Endline (2021)	0.28	0.03	0.23	0.32

Table D2. Intraclass Correlation Coefficients for Oral Reading Fluency, Tajik Grade 4

Time period	ICC	SE	95% Confidence Interval	
			Lower	Upper
Baseline (2018)	0.26	0.03	0.20	0.32
Midline (2019)	0.27	0.03	0.21	0.33
Endline (2021)	0.05	0.02	0.03	0.10

Table. Intraclass Correlation Coefficients for Oral Reading Fluency, Russian Grade 2

Time period	ICC	SE	95% Confidence Interval	
			Lower	Upper
Baseline (2018)	0.21	0.04	0.13	0.29
Midline (2019)	0.14	0.04	0.06	0.22

Endline (2021)	0.29	0.03	0.25	0.34
----------------	------	------	------	------

Table D4. Intraclass Correlation Coefficients for Oral Reading Fluency, Russian Grade 4

Time period	ICC	SE	95% Confidence Interval	
			Lower	Upper
Baseline (2018)	0.19	0.04	0.10	0.28
Midline (2019)	0.18	0.04	0.10	0.27
Endline (2021)	0.24	0.02	0.21	0.28

ANNEX E: DETAILED EGRA TABLES

Table E1: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores Overall, Tajik Grade 2

Subtask	Baseline			Midline			Endline		
	N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	1426	53.61	1.30	1385	46.69	0.66	1342	67.30	6.55
Initial sound identification (% Correct)	1426	56.93	1.85	1385	52.11	1.70	1342	70.68	1.92
Familiar word reading (CWPM)	1426	38.89	1.24	1385	35.50	0.98	1341	41.38	1.08
Nonword reading fluency (cnwpm)	1426	22.34	0.84	1385	20.85	0.61	1341	26.34	0.83
Oral reading fluency (equated)	1426	40.06	1.30	1385	36.66	1.07	1342	43.62	1.31
Reading comprehension (% Correct)	1426	36.57	1.60	1385	36.80	1.34	1342	49.36	1.76
Silent Reading comprehension (% Correct)	1426	31.98	1.09	1385	37.66	1.19	1342	51.02	1.73
Listening comprehension (% Correct)	1426	76.51	1.10	1385	66.98	1.18	1342	74.91	1.62

Table E2: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Sex, Tajik Grade 2

Subtask	Sex	Baseline	Midline	Endline
---------	-----	----------	---------	---------

		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	Female	719	54.14	1.22	690	47.49	0.78	670	76.46	12.71
	Male	707	53.10	1.72	695	45.94	0.83	672	57.98	1.54
Initial sound identification (% Correct)	Female	719	58.74	2.27	690	54.68	2.22	670	72.71	2.21
	Male	707	55.20	1.97	695	49.66	2.07	672	68.61	1.97
Familiar word reading (CWPM)	Female	719	40.22	1.28	690	38.60	1.18	669	43.15	1.40
	Male	707	37.62	1.53	695	32.56	1.23	672	39.57	1.10
Nonword reading fluency (cnwpm)	Female	719	23.12	0.87	690	22.64	0.75	669	26.96	1.03
	Male	707	21.60	1.01	695	19.16	0.80	672	25.70	0.84
Oral reading fluency (equated)	Female	719	41.92	1.40	690	40.12	1.36	670	45.95	1.66
	Male	707	38.29	1.47	695	33.36	1.14	672	41.26	1.34
Reading comprehension (% Correct)	Female	719	39.30	1.78	690	41.27	1.71	670	51.66	2.09
	Male	707	33.97	1.94	695	32.54	1.48	672	47.02	1.83
Silent Reading comprehension (% Correct)	Female	719	34.03	1.33	690	40.34	1.37	670	52.72	2.07
	Male	707	30.03	1.64	695	35.11	1.59	672	49.28	1.88
Listening comprehension (% Correct)	Female	719	77.35	1.23	690	69.35	1.46	670	77.52	1.83
	Male	707	75.71	1.36	695	64.73	1.41	672	72.26	1.82

Table E3: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Urbanicity, Tajik Grade 2

Subtask	Urbanicity	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	Urban	399	56.39	2.89	399	48.81	1.33	399	55.27	4.62
	Rural	1027	52.14	1.24	986	45.61	0.87	943	72.84	9.39

Initial sound identification (% Correct)	Urban	399	59.27	4.41	399	55.98	3.53	399	65.80	5.72
	Rural	1027	55.68	2.09	986	50.13	2.18	943	72.92	1.96
Familiar word reading (CWPM)	Urban	399	44.92	1.52	399	41.97	2.79	399	45.23	2.11
	Rural	1027	35.69	1.54	986	32.20	1.01	943	39.61	1.14
Nonword reading fluency (cnwpm)	Urban	399	26.82	1.33	399	24.53	1.74	398	28.21	1.90
	Rural	1027	19.97	0.76	986	18.97	0.60	942	25.48	0.79
Oral reading fluency (equated)	Urban	399	46.78	1.81	399	43.45	2.87	399	48.66	2.51
	Rural	1027	36.51	1.46	986	33.18	1.10	986	41.31	1.39
Reading comprehension (% Correct)	Urban	399	44.53	1.85	399	45.76	2.66	789	52.70	3.89
	Rural	1027	32.35	1.94	986	32.21	1.59	943	47.82	1.87
Silent Reading comprehension (% Correct)	Urban	399	39.15	1.68	399	46.78	2.72	399	50.62	4.49
	Rural	1027	28.18	1.50	986	33.00	1.43	943	51.20	1.63
Listening comprehension (% Correct)	Urban	399	78.45	1.48	399	71.51	1.61	399	76.72	3.13
	Rural	1027	75.48	1.55	986	64.67	1.58	943	74.08	2.07

Table E4: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Region, Tajik Grade 2

Subtask	Region	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	DRS	258	50.67	1.81	252	51.26	1.53	234	101.76	31.36
	Dushanbe	259	59.17	1.50	259	53.67	0.83	240	53.66	1.24
	GBAO	132	43.61	2.73	106	46.55	2.78	102	54.47	1.29
	Khatlon-Bokhtar	260	60.17	2.91	257	42.34	1.11	296	63.79	4.47
	Khatlon-Kulob	257	43.21	3.98	251	36.14	1.55	211	56.39	1.70
	Sughd	260	50.38	1.49	260	48.26	1.06	259	54.61	2.05

Initial sound identification (% Correct)	DRS	258	49.83	2.53	252	34.08	2.27	234	52.15	3.39
	Dushanbe	259	64.94	3.39	259	47.31	2.56	240	44.55	2.07
	GBAO	132	88.13	1.97	106	63.50	11.02	102	94.03	1.93
	Khatlon-Bokhtar	260	40.39	2.55	257	45.38	3.03	296	76.04	4.51
	Khatlon-Kulob	257	58.63	5.95	251	42.86	4.47	211	72.71	4.86
	Sughd	260	76.78	3.99	260	84.14	3.28	259	91.92	2.04
Familiar word reading (CWPM)	DRS	258	38.16	2.29	252	36.27	1.97	234	40.81	1.73
	Dushanbe	259	46.49	1.66	259	46.95	1.61	239	45.35	1.56
	GBAO	132	47.54	1.95	106	44.69	2.07	102	44.25	2.66
	Khatlon-Bokhtar	260	36.02	3.38	257	27.01	1.26	296	41.62	2.64
	Khatlon-Kulob	257	35.84	2.01	251	36.31	2.66	211	41.21	2.20
	Sughd	260	39.07	2.74	260	36.12	2.34	259	39.27	1.93
Nonword reading fluency (cnwpm)	DRS	258	21.05	1.20	252	20.56	1.33	233	26.48	1.64
	Dushanbe	259	27.85	1.09	259	27.39	1.07	240	26.43	1.17
	GBAO	132	24.90	0.98	106	25.15	2.08	102	26.79	1.31
	Khatlon-Bokhtar	260	21.51	2.55	257	16.62	0.83	296	27.56	2.00
	Khatlon-Kulob	257	20.92	1.25	251	21.46	1.77	211	26.30	1.86
	Sughd	260	21.55	1.47	260	21.23	1.44	259	24.42	1.12
Oral reading fluency (equated)	DRS	258	39.30	2.16	252	36.97	2.44	234	41.96	2.15
	Dushanbe	259	48.42	1.86	259	47.93	1.84	240	48.68	2.06
	GBAO	132	49.33	1.84	106	45.64	2.81	102	48.32	3.50
	Khatlon-Bokhtar	260	37.35	3.63	257	28.63	1.45	296	44.73	3.17
	Khatlon-Kulob	257	36.84	2.04	251	37.06	2.39	211	45.19	2.50
	Sughd	260	39.61	2.88	260	37.46	2.72	259	39.82	2.24
	DRS	258	39.35	3.10	252	37.40	2.92	234	43.87	3.34

Reading comprehension (% Correct)	Dushanbe	259	48.61	2.28	259	49.30	2.35	240	47.15	2.06
	GBAO	132	42.45	3.19	106	45.93	5.90	102	60.78	3.72
	Khatlon-Bokhtar	260	34.41	4.10	257	30.07	2.23	296	57.89	3.72
	Khatlon-Kulob	257	31.62	2.66	251	42.70	4.14	211	52.77	3.57
	Sughd	260	30.43	3.45	260	32.19	3.45	259	41.30	3.01
Silent Reading comprehension (% Correct)	DRS	258	31.01	2.52	252	35.69	2.39	234	49.66	3.21
	Dushanbe	259	48.90	1.97	259	52.87	1.64	240	45.45	2.20
	GBAO	132	37.89	2.80	106	51.63	5.19	102	60.48	3.33
	Khatlon-Bokhtar	260	33.88	2.43	257	34.97	2.12	296	57.68	3.98
	Khatlon-Kulob	257	34.42	2.32	251	43.63	3.51	211	47.26	3.69
	Sughd	260	17.46	1.56	260	27.85	3.04	259	46.96	2.74
Listening comprehension (% Correct)	DRS	258	74.78	2.60	252	62.21	2.10	234	65.92	2.66
	Dushanbe	259	82.57	1.37	259	70.73	1.99	240	71.90	1.94
	GBAO	132	79.36	1.92	106	69.28	6.14	102	67.00	2.45
	Khatlon-Bokhtar	260	82.18	1.74	257	75.95	2.05	296	87.45	2.56
	Khatlon-Kulob	257	74.33	3.28	251	65.70	2.88	211	74.97	1.87
	Sughd	260	67.74	2.53	260	58.08	3.62	259	67.10	4.50

Table E5: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores Overall, Tajik Grade 4

Subtask	Baseline			Midline			Endline		
	N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Familiar word reading (CWPM)	1393	49.70	1.11	1376	49.94	1.15	1348	57.07	2.10
Nonword reading fluency (cnwpm)	1393	26.10	0.58	1376	26.41	0.67	1348	31.52	1.23
Oral reading fluency (equated)	1393	56.72	1.40	1376	59.86	1.41	1348	75.97	3.16

Reading comprehension (% Correct)	1393	44.07	1.53	1376	42.15	1.68	1348	48.71	2.81
Silent Reading comprehension (% Correct)	1393	52.77	1.17	1376	71.61	1.38	1348	79.30	1.26
Listening comprehension (% Correct)	1393	52.71	1.40	1376	61.83	1.08	1348	76.25	1.76

Table E6: Baseline (2018), Midline (2019) and Endline (2021) Oral Reading Fluency and Reading Comprehension Standard Results Overall, Tajik Grade 4

Subtask	Baseline			Midline			Endline		
	N	Proportion	Standard Error	N	Proportion	Standard Error	N	Proportion	Standard Error
Achieved oral reading fluency standard (\geq 80 CWPM)	1393	18.8%	1.8%	1376	22.2%	1.7%	1348	41.4%	3.9%
Achieved reading comprehension standards (\geq 80% correct)	1393	22.5%	2.1%	1376	16.2%	2.2%	1348	28.6%	4.4%

Table E7: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Sex, Tajik Grade 4

Subtask	Sex	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Familiar word reading (CWPM)	Female	697	52.34	1.19	690	53.10	1.22	677	58.79	2.31
	Male	696	47.14	1.36	686	46.86	1.57	671	55.36	2.21
Nonword reading fluency (cnwpm)	Female	697	27.83	0.71	690	28.11	0.77	677	31.96	1.27
	Male	696	24.42	0.67	686	24.76	0.84	671	31.09	1.29
Oral reading fluency (equated)	Female	697	60.28	1.51	690	64.13	1.57	677	77.61	2.69
	Male	696	53.27	1.69	686	55.69	1.76	671	74.35	4.34
Reading comprehension (% Correct)	Female	697	42.96	1.84	690	43.11	1.91	677	49.10	2.68
	Male	696	45.14	1.87	686	41.20	1.78	671	48.32	3.24
	Female	697	53.50	1.48	690	74.96	1.35	677	80.54	1.60

Silent Reading comprehension (% Correct)	Male	696	52.07	1.23	686	68.33	1.96	671	78.08	1.51
Listening comprehension (% Correct)	Female	697	53.98	1.53	690	64.17	1.44	677	78.99	1.72
	Male	696	51.48	1.83	686	59.54	1.26	671	73.55	2.11

Table E8: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Urbanicity, Tajik Grade 4

Subtask	Urbanicity	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Familiar word reading (CWPM)	Urban	394	55.43	2.01	400	57.11	1.76	398	63.79	4.88
	Rural	999	46.73	1.39	976	46.27	1.34	950	53.51	1.35
Nonword reading fluency (cnwpm)	Urban	394	29.07	1.11	400	30.62	0.98	398	33.71	3.11
	Rural	999	24.56	0.76	976	24.25	0.71	950	30.36	0.83
Oral reading fluency (equated)	Urban	394	63.36	2.28	400	67.90	2.12	398	87.13	7.91
	Rural	999	53.27	1.92	976	55.74	1.71	950	70.05	1.66
Reading comprehension (% Correct)	Urban	394	50.17	2.09	400	50.17	2.71	398	53.33	7.21
	Rural	999	40.91	2.19	976	38.03	1.82	950	46.26	1.81
Silent Reading comprehension (% Correct)	Urban	394	58.40	1.54	400	77.71	1.60	398	80.75	3.07
	Rural	999	49.86	1.51	976	68.47	1.80	950	78.54	1.22
Listening comprehension (% Correct)	Urban	394	54.91	2.22	400	65.45	1.51	398	78.89	4.19
	Rural	999	51.57	1.90	976	59.97	1.47	950	74.86	1.56

Table E9: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Region, Tajik Grade 4

Subtask	Region	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error

Familiar word reading (CWPM)	DRS	249	51.88	2.02	248	49.63	2.07	237	56.88	2.22
	Dushanbe	260	64.19	1.67	260	61.24	1.93	240	65.70	2.46
	GBAO	112	54.01	1.32	107	54.51	2.11	104	56.26	2.23
	Khatlon-Bokhtar	260	42.64	3.22	254	43.33	3.51	296	57.26	5.69
	Khatlon-Kulob	253	46.16	1.85	247	46.82	2.70	214	52.37	2.33
	Sughd	259	48.04	1.85	260	51.50	2.43	257	53.70	2.37
Nonword reading fluency (cnwpm)	DRS	249	27.34	1.34	248	25.29	1.05	237	31.90	1.54
	Dushanbe	260	34.52	0.84	260	32.63	0.96	240	33.62	1.29
	GBAO	112	29.68	1.26	107	31.94	1.40	104	32.59	1.24
	Khatlon-Bokhtar	260	22.90	1.46	254	23.86	2.23	296	32.91	3.24
	Khatlon-Kulob	253	24.16	0.94	247	24.03	1.04	214	29.46	1.75
	Sughd	259	23.96	1.09	260	26.79	1.24	257	28.66	1.25
Oral reading fluency (equated)	DRS	249	60.24	3.98	248	60.53	3.00	237	72.27	2.49
	Dushanbe	260	73.66	1.72	260	73.62	2.07	240	95.56	12.19
	GBAO	112	66.26	2.30	107	68.86	2.19	104	82.74	3.69
	Khatlon-Bokhtar	260	47.99	3.14	254	53.00	4.03	296	76.75	7.39
	Khatlon-Kulob	253	55.31	2.19	247	55.50	2.46	214	69.31	2.97
	Sughd	259	52.80	2.52	260	59.25	2.99	257	68.30	2.87
Reading comprehension (% Correct)	DRS	249	49.21	4.67	248	41.41	3.42	237	39.40	3.14
	Dushanbe	260	59.91	1.91	260	53.04	1.86	240	44.22	3.72
	GBAO	112	48.00	2.54	107	47.41	3.33	104	65.29	7.80
	Khatlon-Bokhtar	260	41.26	2.63	254	43.33	4.05	296	59.47	6.59
	Khatlon-Kulob	253	46.41	2.76	247	39.03	2.28	214	39.82	3.27
	Sughd	259	30.98	2.74	260	35.31	3.90	257	45.47	3.31

Silent Reading comprehension (% Correct)	DRS	249	53.47	2.97	248	68.84	2.95	237	73.53	2.68
	Dushanbe	260	61.86	1.89	260	78.33	1.33	240	76.03	1.72
	GBAO	112	55.17	2.75	107	75.10	2.62	104	88.45	4.74
	Khatlon-Bokhtar	260	53.69	2.65	254	72.21	2.78	296	86.19	2.60
	Khatlon-Kulob	253	52.55	2.48	247	76.70	2.43	214	73.64	2.73
	Sughd	259	45.15	1.84	260	66.25	3.89	257	77.39	2.24
Listening comprehension (% Correct)	DRS	249	52.27	4.02	248	60.32	2.81	237	67.12	2.12
	Dushanbe	260	64.57	1.78	260	67.94	1.41	240	72.77	1.72
	GBAO	112	54.89	3.23	107	64.70	3.60	104	82.56	5.88
	Khatlon-Bokhtar	260	53.50	2.98	254	65.35	1.83	296	84.26	3.64
	Khatlon-Kulob	253	56.69	2.33	247	59.99	1.89	214	77.22	3.09
	Sughd	259	42.44	2.49	260	55.76	2.94	257	72.87	3.39

Table E10: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores Overall, Russian Grade 2

Subtask	Baseline			Midline			Endline		
	N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	658	55.10	0.54	563	52.87	0.62	569	54.53	1.27
Initial letter sound identification (% Correct)	658	58.91	1.20	563	57.79	1.43	569	66.17	1.91
Familiar word reading fluency (CWPM)	657	45.18	0.85	563	45.36	0.95	569	47.58	1.27
Nonword reading fluency (cnwpm)	658	24.63	0.39	563	25.60	0.42	569	26.97	0.60
Oral reading fluency - equated (CWPM)	658	39.60	0.75	563	39.81	0.96	569	43.26	1.25
Reading comprehension (% Correct)	658	38.02	1.04	563	56.79	1.42	569	59.84	1.80

Silent reading comprehension (% Correct)	658	37.42	1.21	563	31.84	1.34	569	46.42	2.10
Listening comprehension (% Correct)	658	49.07	1.30	563	45.25	1.50	569	51.73	1.27

Table E11: Baseline (2018), Midline (2019) and Endline (2021) Oral Reading Fluency and Reading Comprehension Standard Results Overall, Russian Grade 2

Subtask	Baseline			Midline			Endline		
	N	Proportion	Standard Error	N	Proportion	Standard Error	N	Proportion	Standard Error
Achieved oral reading fluency standard (≥ 40 CWPM)	658	48.7%	2.0%	563	46.2%	2.2%	569	54.6%	3.0%
Achieved reading comprehension standards ($\geq 80\%$ correct)	658	16.2%	1.6%	563	34.6%	2.3%	569	41.3%	3.1%

Table E12: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Sex, Russian Grade 2

Subtask	Sex	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	Female	320	56.36	0.74	279	54.42	1.04	283	56.37	1.42
	Male	338	60.20	1.52	284	51.59	0.75	286	53.23	1.55
Initial letter sound identification (% Correct)	Female	320	60.20	1.52	279	56.81	2.19	283	65.81	2.79
	Male	338	57.88	1.66	284	58.60	1.84	286	66.43	2.41
Familiar word reading fluency (CWPM)	Female	319	45.98	1.12	279	45.20	1.17	283	51.94	1.69
	Male	338	44.54	1.20	284	45.48	1.31	286	44.53	1.52
Nonword reading fluency (cnwpm)	Female	320	25.15	0.55	279	25.30	0.60	283	28.93	0.85
	Male	338	24.22	0.53	284	25.85	0.59	286	25.59	0.70
Oral reading fluency - equated (CWPM)	Female	320	40.77	1.01	279	40.08	1.26	283	47.47	1.96
	Male	338	38.66	1.03	284	39.59	1.27	286	40.31	1.26

Reading comprehension (% Correct)	Female	320	41.82	1.62	279	57.31	1.88	283	64.46	2.37
	Male	338	34.96	1.32	284	56.36	1.88	286	56.60	2.17
Silent reading comprehension (% Correct)	Female	320	39.43	1.60	279	32.49	1.71	283	48.27	2.64
	Male	338	35.80	1.73	284	31.30	1.80	286	45.13	2.45
Listening comprehension (% Correct)	Female	320	47.86	1.87	279	44.85	2.04	283	52.74	1.84
	Male	338	50.04	1.81	284	45.58	2.03	286	51.02	1.51

Table E13: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Urbanicity, Russian Grade 2

Subtask	Urbanicity	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	Urban	107	55.64	0.58	484	53.12	0.67	509	54.41	1.34
	Rural	551	49.72	1.10	79	49.85	1.68	60	56.48	2.48
Initial letter sound identification (% Correct)	Urban	107	58.60	1.27	484	57.11	1.56	509	65.95	2.02
	Rural	551	62.00	2.67	79	66.00	3.37	60	70.04	2.92
Familiar word reading fluency (CWPM)	Urban	107	46.35	0.88	484	45.93	1.04	509	48.08	1.33
	Rural	550	33.40	1.61	79	38.50	2.12	60	39.04	2.82
Nonword reading fluency (cnwpm)	Urban	107	25.13	0.40	484	25.85	0.45	509	27.19	0.62
	Rural	551	19.62	0.82	79	22.62	1.11	60	23.15	1.43
Oral reading fluency - equated (CWPM)	Urban	107	40.72	0.78	484	40.53	1.05	509	43.77	1.31
	Rural	551	28.33	1.38	79	31.13	1.77	79	34.49	2.45
Reading comprehension (% Correct)	Urban	107	39.41	1.10	484	58.52	1.54	509	60.52	1.88
	Rural	551	24.02	2.57	79	35.91	3.02	60	48.10	5.32
Silent reading comprehension (% Correct)	Urban	107	38.67	1.27	484	33.17	1.47	509	46.95	2.20
	Rural	551	24.83	3.56	79	15.79	2.91	60	37.24	3.46

Listening comprehension (% Correct)	Urban	107	50.11	1.36	484	46.31	1.60	509	51.93	1.33
	Rural	551	38.54	4.85	79	32.54	3.98	60	48.32	3.46

Table E14: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Region, Russian Grade 2

Subtask	Region	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Letter naming fluency (clpm)	DRS	109	52.60	1.24	97	51.37	1.64	70	52.38	2.66
	Dushanbe	149	55.48	1.03	130	54.59	1.15	129	52.33	2.25
	Khatlon-Bokhtar	80	55.23	1.48	60	50.81	2.08	71	62.44	4.76
	Khatlon-Kulob	10	52.54	4.55	10	46.18	4.07	10	53.29	4.12
	Sughd	310	55.85	0.71	266	52.50	0.86	289	56.11	1.27
Initial letter sound identification (% Correct)	DRS	109	48.52	2.78	97	70.41	2.75	70	72.56	3.68
	Dushanbe	149	61.53	2.42	130	54.91	2.91	129	54.53	3.44
	Khatlon-Bokhtar	80	57.14	2.81	60	54.77	3.82	71	80.82	4.05
	Khatlon-Kulob	10	63.85	4.71	10	38.07	6.19	10	77.72	7.72
	Sughd	310	60.73	1.59	266	57.76	1.84	289	76.49	1.84
Familiar word reading fluency (CWPM)	DRS	109	37.77	2.21	97	38.42	1.83	70	35.51	4.56
	Dushanbe	149	48.77	1.63	130	46.47	1.95	129	51.37	2.03
	Khatlon-Bokhtar	80	45.54	1.92	60	47.14	1.98	71	42.28	3.83
	Khatlon-Kulob	10	53.32	3.51	10	52.56	7.97	10	36.96	3.71
	Sughd	310	43.46	1.17	266	46.23	1.29	289	46.70	1.59
Nonword reading fluency (cnwpm)	DRS	109	20.35	1.15	97	23.03	0.93	70	22.18	1.97
	Dushanbe	149	26.20	0.70	130	26.04	0.77		28.03	0.97
	Khatlon-Bokhtar	80	26.85	0.93	60	25.90	1.11	71	27.01	1.74
	Khatlon-Kulob	10	30.16	1.74	10	28.19	3.63	10	24.90	1.86

	Sughd	310	23.38	0.55	266	26.01	0.66	289	26.62	0.78
Oral reading fluency - equated (CWPM)	DRS	109	31.63	1.61	97	32.23	1.43	70	32.16	3.83
	Dushanbe	149	43.01	1.41	130	42.85	2.07	129	46.90	2.08
	Khatlon-Bokhtar	80	41.41	1.99	60	38.42	1.72	71	35.92	2.82
	Khatlon-Kulob	10	47.21	3.60	10	44.76	6.42	10	37.77	3.83
	Sughd	310	37.78	1.07	266	39.64	1.37	289	42.66	1.53
Reading comprehension (% Correct)	DRS	109	26.84	2.34	97	42.90	2.76	70	38.03	6.16
	Dushanbe	149	43.49	1.92	130	60.40	2.83	129	61.42	2.76
	Khatlon-Bokhtar	80	38.37	3.02	60	55.31	3.32	71	55.51	5.51
	Khatlon-Kulob	10	55.85	5.25	10	58.37	7.13	10	63.45	6.09
	Sughd	310	35.11	1.52	266	59.28	2.24	289	63.23	2.53
Silent reading comprehension (% Correct)	DRS	109	26.21	2.81	97	15.54	2.07	70	31.59	4.17
	Dushanbe	149	40.56	2.31	130	32.86	2.66	129	49.47	3.76
	Khatlon-Bokhtar	80	38.31	2.92	60	29.38	2.69	71	44.03	4.72
	Khatlon-Kulob	10	68.15	7.83	10	44.30	7.47	10	60.97	6.72
	Sughd	310	36.26	1.82	266	38.49	2.19	289	45.56	2.23
Listening comprehension (% Correct)	DRS	109	38.48	3.23	97	37.89	3.58	70	46.58	3.58
	Dushanbe	149	50.56	2.44	130	43.98	2.69	129	51.16	2.15
	Khatlon-Bokhtar	80	45.13	2.99	60	47.12	3.88	71	52.67	3.74
	Khatlon-Kulob	10	67.59	5.19	10	48.52	11.34	10	67.03	2.58
	Sughd	310	52.34	2.08	266	49.78	2.42	289	52.84	1.82

Table E15: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores Overall, Russian Grade 4

Subtask	Baseline			Midline			Endline		
	N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error

Familiar word reading (CWPM)	562	65.51	1.10	536	66.46	1.14	599	63.26	1.89
Nonword reading fluency (cnwpm)	562	34.93	0.59	536	35.35	0.63	599	34.97	0.76
Oral reading fluency (equated)	562	73.73	1.34	536	73.24	1.36	599	74.56	2.29
Reading comprehension (% Correct)	562	57.71	1.49	536	57.97	1.65	594	60.54	1.21
Silent Reading comprehension (% Correct)	562	37.02	1.52	536	44.07	1.83	599	46.86	1.62
Listening comprehension (% Correct)	562	39.65	1.26	536	59.58	1.46	599	44.47	1.22

Table E16: Baseline (2018), Midline (2019) and Endline (2021) Oral Reading Fluency and Reading Comprehension Standard Results Overall, Russian Grade 4

Subtask	Baseline			Midline			Endline		
	N	Proportion	Standard Error	N	Proportion	Standard Error	N	Proportion	Standard Error
Achieved oral reading fluency standard (≥ 80 CWPM)	562	38.6%	2.3%	536	37.7%	2.6%	599	41.7%	3.3%
Achieved reading comprehension standards ($\geq 80\%$ correct)	562	41.4%	2.3%	536	42.4%	2.5%	599	41.1%	2.4%

Table E17: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Sex, Russian Grade 4

Subtask	Sex	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Familiar word reading (CWPM)	Female	278	66.59	1.37	266	69.33	1.39	300	60.83	2.03
	Male	284	64.69	1.41	270	64.24	1.67	299	64.88	2.33
Nonword reading fluency (cnwpm)	Female	278	35.58	0.79	266	36.69	0.73	300	35.04	0.93
	Male	284	34.43	0.74	270	34.32	0.94	299	34.92	0.87
Oral reading fluency (equated)	Female	278	76.32	1.64	266	77.85	1.45	300	74.23	2.80
	Male	284	71.75	1.66	270	69.68	2.03	299	74.79	2.79

Reading comprehension (% Correct)	Female	278	58.60	2.20	266	61.48	2.16	300	61.71	1.70
	Male	284	57.03	1.78	270	55.26	2.38	299	59.76	1.48
Silent Reading comprehension (% Correct)	Female	278	35.40	1.70	266	45.07	2.36	300	50.18	2.25
	Male	284	38.26	1.97	270	43.30	2.73	299	44.65	1.99
Listening comprehension (% Correct)	Female	278	38.22	1.70	266	57.92	1.82	300	42.28	1.74
	Male	284	40.73	1.54	270	60.86	2.24	299	45.92	1.57

Table E18: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Urbanicity, Russian Grade 4

Subtask	Urbanicity	Baseline			Midline			Endline		
		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Familiar word reading (CWPM)	Urban	508	66.40	1.08	478	67.10	1.26	71	63.50	1.97
	Rural	54	49.59	2.19	58	58.03	2.68	528	59.21	4.48
Nonword reading fluency (cnwpm)	Urban	508	35.46	0.57	478	35.65	0.70	71	35.04	0.78
	Rural	54	25.48	1.33	58	31.35	1.43	528	33.80	2.61
Oral reading fluency (equated)	Urban	508	74.92	1.31	478	74.15	1.53	71	75.12	2.39
	Rural	54	52.47	2.34	58	61.23	2.52	528	65.15	5.09
Reading comprehension (% Correct)	Urban	508	58.41	1.50	478	58.53	1.84	71	61.10	1.25
	Rural	54	45.18	3.94	58	50.50	6.17	528	50.96	3.75
Silent Reading comprehension (% Correct)	Urban	508	36.52	1.50	478	44.37	2.06	71	46.96	1.70
	Rural	54	45.98	5.44	58	40.06	4.62	528	45.11	3.53
Listening comprehension (% Correct)	Urban	508	39.59	1.26	478	59.64	1.60	71	44.62	1.30
	Rural	54	40.58	4.69	58	58.82	5.82	528	41.88	2.89

Table E19: Baseline (2018), Midline (2019) and Endline (2021) Mean EGRA Scores by Region, Russian Grade 4

Subtask	Region	Baseline	Midline	Endline
---------	--------	----------	---------	---------

		N	Mean	Standard Error	N	Mean	Standard Error	N	Mean	Standard Error
Familiar word reading (CWPM)	DRS	62	54.17	2.61	50	63.68	3.43	80	58.74	2.87
	Dushanbe	150	71.67	1.70	150	70.30	2.26	130	64.30	3.85
	Khatlon-Bokhtar	70	59.27	2.70	60	66.66	3.54	70	65.22	2.93
	Khatlon-Kulob	10	54.13	5.40	10	68.32	4.70	10	65.58	4.52
	Sughd	270	64.68	1.80	266	61.95	1.33	309	62.66	1.83
Nonword reading fluency (cnwpm)	DRS	62	28.81	1.29	50	35.13	1.65	80	33.06	1.92
	Dushanbe	150	37.96	0.91	150	37.12	1.20	130	34.31	1.46
	Khatlon-Bokhtar	70	32.87	1.47	60	38.32	1.81	70	38.94	1.59
	Khatlon-Kulob	10	30.43	4.10	10	36.66	4.23	10	39.69	3.25
	Sughd	270	34.23	1.00	266	31.97	0.76	309	35.20	0.87
Oral reading fluency (equated)	DRS	62	58.72	2.48	50	70.02	4.30	80	63.70	3.69
	Dushanbe	150	81.18	2.15	150	79.92	2.84	130	76.55	4.19
	Khatlon-Bokhtar	70	66.97	3.79	60	69.29	2.70	70	70.77	3.33
	Khatlon-Kulob	10	59.32	7.12	10	73.03	6.29	10	151.65	60.36
	Sughd	270	72.92	2.08	266	66.38	1.45	309	72.96	2.41
Reading comprehension (% Correct)	DRS	62	53.04	3.39	50	44.23	7.60	80	55.10	3.68
	Dushanbe	150	60.42	2.66	150	63.32	2.88	130	60.49	1.93
	Khatlon-Bokhtar	70	56.17	3.97	60	43.56	4.18	70	65.37	4.11
	Khatlon-Kulob	10	18.48	6.08	10	59.24	8.69	10	57.60	7.24
	Sughd	270	58.61	1.94	266	59.81	2.30	309	61.04	1.81
	DRS	62	30.01	3.29	50	36.13	5.69	80	38.43	2.87
	Dushanbe	150	37.08	2.75	150	45.79	3.49	130	43.12	2.84

Silent Reading comprehension (% Correct)	Khatlon-Bokhtar	70	32.89	2.76	60	33.86	4.32	70	58.44	4.87
	Khatlon-Kulob	10	22.29	2.62	10	52.19	7.10	10	50.40	8.22
	Sughd	270	42.27	2.20	266	47.22	2.76	309	50.83	2.03
Listening comprehension (% Correct)	DRS	62	29.15	2.95	50	48.11	6.87	80	30.89	2.72
	Dushanbe	150	40.80	2.34	150	64.19	2.43	130	41.34	1.97
	Khatlon-Bokhtar	70	40.23	2.26	60	51.27	4.00	70	55.26	3.50
	Khatlon-Kulob	10	28.00	4.38	10	62.86	7.27	10	51.20	4.75
	Sughd	270	42.56	1.57	266	59.64	2.17	309	49.08	1.75

ANNEX F: DETAILED TEACHER SSME TABLES

Table F1: Endline (2021) Teacher Survey: Frequency of Responses by Question, Tajik Schools

Question	Responses	Tajik Grade 2				Tajik Grade 4			
		Proportion	SE	95% Conf. Interval		Proportion	SE	95% Conf. Interval	
				Lower	Upper			Lower	Upper
Participant gender	Female	85.36%	3.01%	79.41%	91.31%	88.84%	2.93%	83.05%	94.63%
	Male	14.64%	3.01%	8.69%	20.59%	11.16%	2.93%	5.37%	16.95%
What is your native language?	Tajik	80.40%	4.16%	72.17%	88.62%	81.78%	4.04%	73.78%	89.77%
	Russian	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Uzbek	19.57%	4.16%	11.35%	27.79%	18.17%	4.04%	10.18%	26.16%
	Pamiri dialect	0.03%	0.03%	-0.02%	0.09%	0.06%	0.06%	-0.06%	0.17%
	Other (specify):	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
What is your highest level of education?	Secondary vocational education	22.58%	4.43%	13.81%	31.35%	25.66%	5.26%	15.25%	36.07%
	Incomplete high (university) education	7.80%	2.86%	2.14%	13.46%	23.43%	5.05%	13.44%	33.41%

	Complete high (university) education	67.12%	4.99%	57.24%	77.00%	48.15%	5.92%	36.44%	59.87%
	Other (do not specify)	2.50%	1.68%	-0.83%	5.83%	2.76%	1.75%	-0.71%	6.23%
Attended in-service training or professional development sessions such as workshops in the last year.	Yes	60.47%	5.33%	49.92%	71.02%	51.67%	5.75%	40.29%	63.04%
	No	39.53%	5.33%	28.98%	50.08%	48.33%	5.75%	36.96%	59.71%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Have you attended any in-service training on how to teach reading?	Yes	77.43%	4.29%	68.95%	85.91%	78.78%	4.58%	69.72%	87.85%
	No	22.57%	4.29%	14.09%	31.05%	20.72%	4.54%	11.73%	29.71%
	Do not know	0.00%	N.A.	N.A.	N.A.	0.50%	0.50%	-0.50%	1.49%
If yes, did you receive training for this school year?	Yes	50.08%	6.30%	37.57%	62.60%	42.25%	6.99%	28.38%	56.12%
	No	49.92%	6.30%	37.40%	62.43%	56.84%	7.00%	42.95%	70.73%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.90%	0.91%	-0.89%	2.70%
Have you received any methodological support or assistance at school this past year on how to teach reading?	Yes	95.95%	1.05%	93.86%	98.03%	92.80%	2.21%	88.42%	97.18%
	No	4.05%	1.05%	1.97%	6.14%	7.20%	2.21%	2.82%	11.58%
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
If yes, what type of support was received?	External support visit from education officers	18.12%	3.69%	10.81%	25.43%	21.86%	6.21%	9.57%	34.16%
	Support visit from school methodologist	57.06%	5.30%	46.56%	67.56%	56.36%	6.30%	43.87%	68.84%
	Support visit from school level facilitator	80.22%	3.88%	72.54%	87.90%	77.16%	4.68%	67.88%	86.44%
	School of young teachers	7.98%	2.69%	2.65%	13.30%	12.19%	4.14%	3.99%	20.39%
	Mentorship	33.90%	5.17%	23.66%	44.14%	33.80%	5.93%	22.05%	45.54%
	Don't know	1.68%	1.26%	-0.81%	4.17%	0.00%	N.A.	N.A.	N.A.
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
What grade or grades do you teach in this school year?	Grade 1	3.17%	1.20%	0.79%	5.54%	6.48%	3.46%	-0.36%	13.31%
	Grade 2	98.54%	1.29%	95.99%	101.10%	0.00%	N.A.	N.A.	N.A.

	Grade 3	5.18%	2.25%	0.73%	9.63%	13.41%	5.95%	1.64%	25.18%
	Grade 4	11.71%	4.63%	2.55%	20.88%	100.00%	0.00%	100.00%	100.00%
Do you have a written lesson plan for today?	Yes	96.67%	1.67%	93.37%	99.98%	100.00%	0.00%	100.00%	100.00%
	No	3.33%	1.67%	0.02%	6.63%	0.00%	N.A.	N.A.	N.A.
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Last five school days: Students were assigned reading to do on their own in school time.	Never	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Sometimes	15.57%	3.88%	7.90%	23.25%	5.70%	1.76%	2.21%	9.19%
	Frequently	30.62%	5.11%	20.50%	40.74%	40.71%	5.51%	29.81%	51.62%
	Every day	53.81%	4.99%	43.94%	63.67%	53.59%	5.62%	42.46%	64.72%
Last five school days: Students were assigned reading to do at home.	Never	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Sometimes	3.73%	1.99%	-0.21%	7.66%	6.96%	2.46%	2.09%	11.82%
	Frequently	13.92%	3.35%	7.29%	20.56%	15.55%	3.47%	8.68%	22.42%
	Every day	82.35%	3.77%	74.90%	89.80%	77.50%	4.07%	69.45%	85.54%
Do you use the official reading curriculum in your classroom lessons?	Never	7.75%	3.43%	0.95%	14.55%	1.84%	1.42%	-0.97%	4.64%
	Sometimes	16.07%	3.53%	9.07%	23.06%	18.42%	4.35%	9.82%	27.02%
	Frequently	21.93%	3.90%	14.22%	29.65%	27.53%	6.11%	15.45%	39.62%
	Every day	54.25%	5.13%	44.11%	64.39%	52.21%	5.71%	40.91%	63.51%
Do you have teacher guides?	Yes	82.36%	4.19%	74.07%	90.65%	94.26%	3.34%	87.66%	100.87%
	No	15.02%	3.97%	7.17%	22.87%	5.74%	3.34%	-0.87%	12.34%
	Don't know	2.62%	1.76%	-0.86%	6.09%	0.00%	N.A.	N.A.	N.A.
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
If yes, how useful do you find them?	Not very useful	0.67%	0.65%	-0.62%	1.95%	0.00%	N.A.	N.A.	N.A.
	Moderately useful	21.51%	4.76%	12.08%	30.93%	15.54%	4.08%	7.45%	23.62%
	Very Useful	77.47%	4.79%	67.97%	86.97%	83.05%	4.27%	74.60%	91.51%
	Don't know	0.36%	0.26%	-0.15%	0.87%	0.00%	N.A.	N.A.	N.A.

	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	1.41%	1.40%	-1.36%	4.18%
	Skipped	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
How do you measure your students' progress?	Written tests	29.37%	4.93%	19.60%	39.13%	30.76%	4.75%	21.36%	40.16%
	Oral evaluations	68.40%	4.79%	58.91%	77.89%	61.83%	6.33%	49.30%	74.36%
	Portfolios and other projects	15.18%	4.35%	6.57%	23.79%	16.90%	4.98%	7.06%	26.74%
	Homework	58.41%	4.83%	48.86%	67.96%	61.91%	5.43%	51.18%	72.65%
	End of term evaluations	53.11%	5.09%	43.05%	63.18%	48.29%	5.61%	37.18%	59.40%
	Other	35.77%	4.48%	26.91%	44.64%	50.47%	4.88%	40.83%	60.12%
	Don't know	2.34%	1.28%	-0.20%	4.88%	1.80%	1.24%	-0.65%	4.25%
	What reading skills should your students have at the end of the school year?	Read grade level stories	54.29%	5.20%	44.00%	64.58%	65.46%	5.32%	54.92%
Sound out words they don't know		33.80%	5.05%	23.82%	43.79%	22.57%	4.25%	14.16%	30.99%
Understand stories that they read		72.02%	4.79%	62.55%	81.49%	82.31%	4.75%	72.91%	91.70%
Know letter names		22.55%	3.71%	15.21%	29.89%	13.27%	3.50%	6.34%	20.21%
Other (do not specify)		37.35%	4.46%	28.53%	46.18%	45.09%	5.75%	33.71%	56.47%
Don't know		5.56%	2.97%	-0.32%	11.44%	0.48%	0.48%	-0.47%	1.43%
At what grade level do you expect students to be reading the language of instruction fluently?	Grade 1	65.27%	4.62%	56.13%	74.42%	59.74%	5.41%	49.03%	70.45%
	Grade 2	30.19%	4.67%	20.94%	39.43%	38.64%	5.26%	28.23%	49.06%
	Grade 3	3.33%	2.20%	-1.03%	7.68%	1.61%	1.34%	-1.04%	4.27%
	Grade 4 or higher	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Don't know/refuse	1.21%	1.13%	-1.03%	3.46%	0.00%	N.A.	N.A.	N.A.
At what grade level do you expect students to be writing the language of instruction?	Grade 1	93.71%	2.52%	88.72%	98.69%	93.05%	2.91%	87.29%	98.81%
	Grade 2	6.29%	2.52%	1.31%	11.28%	6.95%	2.91%	1.19%	12.71%
	Grade 3	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Grade 4 or higher	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Don't know/refuse	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.

Do students take books from school to read at home?	Never	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Sometimes	24.74%	4.61%	15.61%	33.86%	16.83%	3.68%	9.55%	24.11%
	Frequently	69.58%	4.83%	60.03%	79.12%	76.56%	4.28%	68.09%	85.03%
	Every day	5.68%	2.02%	1.68%	9.69%	6.62%	2.50%	1.68%	11.55%
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Do the teachers at this school work together as teams to solve problems related to teaching?	Yes	100.00%	0.00%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%
	No	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Do not know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
In this classroom in this shift, do you teach students from more than one grade?	Yes	0.00%	N.A.	N.A.	N.A.	2.73%	1.73%	-0.69%	6.14%
	No	100.00%	0.00%	100.00%	100.00%	97.27%	1.73%	93.86%	100.69%
	Do not know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
How frequently does the deputy director observe your classes?	Never	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Once a year	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Once every 6 months	0.00%	N.A.	N.A.	N.A.	0.59%	0.60%	-0.59%	1.78%
	Once every 2–3 months	0.37%	0.34%	-0.30%	1.05%	0.20%	0.20%	-0.20%	0.60%
	Once every month	35.80%	4.74%	26.43%	45.17%	27.17%	4.49%	18.29%	36.06%
	Once every two weeks	39.31%	5.31%	28.80%	49.81%	34.36%	5.21%	24.05%	44.66%
	Once every week	23.50%	4.03%	15.52%	31.47%	37.03%	5.40%	26.34%	47.72%
	Daily	1.02%	0.56%	-0.09%	2.13%	0.65%	0.52%	-0.39%	1.68%
When you need some help with your teaching, who do you consult?	Never need help	0.00%	N.A.	N.A.	N.A.	1.14%	1.14%	-1.11%	3.39%
	No one to ask for help	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Organize meetings with teachers	47.05%	5.09%	36.97%	57.12%	40.18%	5.84%	28.62%	51.74%
	Discuss casually with teachers	59.12%	4.78%	49.66%	68.57%	64.53%	5.19%	54.27%	74.79%
	Head teacher	10.75%	2.96%	4.89%	16.61%	15.29%	3.95%	7.49%	23.10%
	Assistant head teacher	70.70%	4.77%	61.27%	80.14%	71.28%	4.67%	62.05%	80.51%

	Seek advice from education supervisor or subject specialist	63.00%	4.83%	53.45%	72.55%	66.89%	5.15%	56.69%	77.08%
	Other	5.94%	2.60%	0.80%	11.08%	7.44%	2.40%	2.70%	12.18%
	Don't know/refuse	1.21%	1.13%	-1.03%	3.46%	0.00%	N.A.	N.A.	N.A.
How do you use the results of students' oral and written assessments in your teaching?	Grade students	49.70%	4.99%	39.82%	59.58%	60.64%	5.63%	49.51%	71.77%
	Evaluate students' understanding of subject matter	46.51%	5.29%	36.05%	56.98%	49.97%	5.85%	38.39%	61.55%
	Plan teaching activities	29.93%	4.70%	20.62%	39.24%	27.66%	4.81%	18.13%	37.19%
	Adapt teaching to better suit students' needs	33.74%	4.49%	24.85%	42.63%	35.67%	5.27%	25.24%	46.09%
	Other	39.34%	4.48%	30.48%	48.20%	45.74%	5.50%	34.85%	56.63%
	Do not know / refuse to respond	3.16%	2.22%	-1.23%	7.55%	1.45%	1.02%	-0.56%	3.46%
In your class, how many parents or guardians review students' homework?	None	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Some	26.51%	4.53%	17.55%	35.48%	34.19%	5.98%	22.36%	46.03%
	Most	65.82%	4.93%	56.06%	75.58%	59.24%	6.11%	47.15%	71.33%
	All	7.67%	3.12%	1.50%	13.83%	6.56%	2.75%	1.13%	12.00%
	Don't know/refuse	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Are you generally satisfied with parents' involvement in their children's schoolwork?	Yes	76.11%	4.48%	67.25%	84.97%	61.17%	6.09%	49.12%	73.22%
	No	23.89%	4.48%	15.03%	32.75%	38.83%	6.09%	26.78%	50.88%
	Do not know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.

Table F2: Endline (2021) Teacher Survey: Means of Responses by Question, Tajik Schools

Question	Tajik Grade 2				Tajik Grade 4			
	Mean	SE	95% Conf. Interval		Mean	SE	95% Conf. Interval	
			Lower	Upper			Lower	Upper

How many years of teaching experience do you have?	19.56	1.15	17.28	21.83	19.25	1.52	16.25	22.26
How many hours of in-service training on how to teach reading have you received?	43.18	4.98	33.11	53.26	44.34	0.00	44.34	44.34
In this class, how many boys are enrolled?	16.49	0.99	14.52	18.45	15.22	1.00	13.25	17.19
In this class, how many girls are enrolled?	15.66	1.12	13.45	17.87	15.83	1.07	13.71	17.95

Table F3: Endline (2021) Teacher Survey: Frequency of Responses by Question, Russian Schools

Question	Responses	Russian Grade 2				Russian Grade 4			
		Proportion	SE	95% Conf. Interval		Proportion	SE	95% Conf. Interval	
				Lower	Upper			Lower	Upper
Participant gender	Female	100.00%	0.00%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%
	Male	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
What is your native language?	Tajik	50.57%	6.27%	37.94%	63.19%	39.36%	5.43%	28.44%	50.28%
	Russian	24.32%	6.10%	12.04%	36.60%	32.12%	5.25%	21.56%	42.68%
	Uzbek	22.81%	4.32%	14.10%	31.52%	23.21%	3.82%	15.53%	30.89%
	Pamiri dialect	2.31%	1.67%	-1.06%	5.67%	3.31%	2.22%	-1.16%	7.78%
	Other (specify):	0.00%	N.A.	N.A.	N.A.	2.00%	1.03%	-0.07%	4.07%
What is your highest level of education?	Secondary vocational education	5.25%	2.54%	0.13%	10.37%	12.74%	3.25%	6.20%	19.27%
	Incomplete high (university) education	1.70%	0.79%	0.11%	3.29%	3.93%	1.86%	0.19%	7.67%
	Complete high (university) education	93.05%	2.67%	87.67%	98.43%	83.34%	3.51%	76.27%	90.40%
	Other (do not specify)	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Attended in-service training or professional	Yes	57.48%	6.42%	44.55%	70.42%	67.50%	4.83%	57.79%	77.21%
	No	42.52%	6.42%	29.58%	55.45%	32.50%	4.83%	22.79%	42.21%

development sessions such as workshops in the last year.	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Have you attended any in-service training on how to teach reading?	Yes	74.69%	4.77%	65.07%	84.30%	64.44%	5.52%	53.34%	75.55%
	No	25.31%	4.77%	15.70%	34.93%	32.45%	5.48%	21.44%	43.47%
	Do not know	0.00%	N.A.	N.A.	N.A.	3.10%	2.09%	-1.11%	7.31%
If yes, did you receive training for this school year?	Yes	24.68%	8.89%	6.48%	42.88%	22.28%	7.20%	7.58%	36.97%
	No	75.32%	8.89%	57.12%	93.52%	77.72%	7.20%	63.03%	92.42%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Have you received any methodological support or assistance at school this past year on how to teach reading?	Yes	64.96%	6.14%	52.59%	77.34%	80.75%	4.92%	70.85%	90.65%
	No	35.04%	6.14%	22.66%	47.41%	19.25%	4.92%	9.35%	29.15%
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
If yes, what type of support was received?	External support visit from education officers	38.79%	7.58%	23.36%	54.22%	22.64%	4.19%	14.16%	31.11%
	Support visit from school methodologist	59.60%	8.05%	43.22%	75.97%	70.68%	4.68%	61.21%	80.15%
	Support visit from school level facilitator	38.21%	8.08%	21.77%	54.65%	19.35%	4.29%	10.66%	28.03%
	School of young teachers	27.17%	7.25%	12.42%	41.92%	20.72%	4.57%	11.48%	29.97%
	Mentorship	33.96%	4.65%	24.50%	43.41%	44.07%	5.97%	32.00%	56.14%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
What grade or grades do you teach in this school year?	Grade 1	19.96%	4.13%	11.64%	28.29%	20.91%	4.86%	11.13%	30.69%
	Grade 2	99.08%	0.54%	97.99%	100.17%	16.73%	4.52%	7.65%	25.82%
	Grade 3	33.49%	6.07%	21.27%	45.72%	19.71%	3.66%	12.35%	27.07%
	Grade 4	21.70%	4.60%	12.43%	30.96%	98.14%	1.59%	94.95%	101.33%
Do you have a written lesson plan for today?	Yes	96.97%	2.17%	92.59%	101.35%	98.26%	1.06%	96.13%	100.39%
	No	3.03%	2.17%	-1.35%	7.41%	1.74%	1.06%	-0.39%	3.87%

	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Last five school days: Students were assigned reading to do on their own in school time.	Never	0.52%	0.42%	-0.32%	1.36%	0.00%	N.A.	N.A.	N.A.
	Sometimes	10.81%	3.02%	4.72%	16.89%	12.71%	3.36%	5.97%	19.46%
	Frequently	42.94%	5.95%	30.96%	54.93%	39.16%	5.58%	27.93%	50.38%
	Every day	45.73%	5.69%	34.27%	57.19%	48.13%	5.39%	37.30%	58.96%
Last five school days: Students were assigned reading to do at home.	Never	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Sometimes	8.78%	3.32%	2.10%	15.47%	14.32%	3.71%	6.86%	21.77%
	Frequently	29.66%	6.06%	17.45%	41.87%	34.29%	5.20%	23.83%	44.76%
	Every day	61.56%	5.64%	50.19%	72.93%	51.39%	4.86%	41.61%	61.17%
Do you use the official reading curriculum in your classroom lessons?	Never	0.29%	0.19%	-0.09%	0.67%	0.00%	N.A.	N.A.	N.A.
	Sometimes	11.42%	3.34%	4.69%	18.15%	14.83%	3.18%	8.43%	21.23%
	Frequently	54.35%	5.84%	42.60%	66.11%	42.71%	5.39%	31.87%	53.55%
	Every day	33.94%	5.82%	22.22%	45.66%	42.46%	5.24%	31.93%	52.99%
Do you have teacher guides?	Yes	88.14%	3.37%	81.36%	94.92%	90.45%	1.87%	86.69%	94.22%
	No	11.86%	3.37%	5.08%	18.64%	9.55%	1.87%	5.78%	13.31%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
If yes, how useful do you find them?	Not very useful	3.34%	2.46%	-1.65%	8.32%	0.00%	N.A.	N.A.	N.A.
	Moderately useful	23.30%	7.30%	8.52%	38.08%	27.53%	5.15%	17.14%	37.91%
	Very Useful	73.36%	7.52%	58.15%	88.58%	72.47%	5.15%	62.09%	82.86%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Skipped	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
How do you measure your students' progress?	Written tests	48.56%	6.31%	35.85%	61.27%	51.85%	5.54%	40.70%	63.00%
	Oral evaluations	75.63%	5.15%	65.26%	85.99%	71.70%	5.14%	61.36%	82.05%

	Portfolios and other projects	21.44%	4.79%	11.79%	31.08%	28.28%	4.05%	20.13%	36.43%
	Homework	49.20%	6.03%	37.07%	61.34%	56.49%	4.88%	46.68%	66.30%
	End of term evaluations	41.78%	5.37%	30.98%	52.59%	40.13%	5.25%	29.58%	50.69%
	Other	13.99%	3.31%	7.33%	20.65%	17.15%	3.29%	10.53%	23.77%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
What reading skills should your students have at the end of the school year?	Read grade level stories	78.30%	3.20%	71.85%	84.75%	74.86%	5.37%	64.05%	85.66%
	Sound out words they don't know	27.55%	4.59%	18.30%	36.80%	28.14%	4.90%	18.29%	37.99%
	Understand stories that they read	88.30%	3.88%	80.48%	96.13%	100.00%	0.00%	100.00%	100.00%
	Know letter names	23.17%	6.07%	10.93%	35.40%	0.00%	N.A.	N.A.	N.A.
	Other (do not specify)	19.39%	3.74%	11.86%	26.92%	19.37%	3.12%	13.11%	25.63%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
At what grade level do you expect students to be reading the language of instruction fluently?	Grade 1	14.39%	3.54%	7.27%	21.52%	18.42%	3.08%	12.24%	24.61%
	Grade 2	59.71%	6.04%	47.54%	71.88%	38.10%	5.18%	27.69%	48.51%
	Grade 3	7.85%	3.48%	0.84%	14.85%	26.26%	5.06%	16.09%	36.43%
	Grade 4 or higher	18.05%	5.31%	7.36%	28.74%	17.22%	4.55%	8.06%	26.37%
	Don't know/refuse	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
At what grade level do you expect students to be writing the language of instruction?	Grade 1	81.99%	3.73%	74.47%	89.50%	71.88%	5.05%	61.73%	82.03%
	Grade 2	16.46%	3.44%	9.52%	23.39%	22.86%	4.62%	13.58%	32.15%
	Grade 3	1.56%	1.14%	-0.73%	3.85%	0.53%	0.38%	-0.24%	1.31%
	Grade 4 or higher	0.00%	N.A.	N.A.	N.A.	4.72%	1.93%	0.84%	8.61%
	Don't know/refuse	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Do students take books from school to read at home?	Never	1.67%	1.08%	-0.50%	3.84%	4.89%	2.39%	0.10%	9.69%
	Sometimes	34.67%	6.29%	21.99%	47.34%	24.24%	4.26%	15.67%	32.82%
	Frequently	53.09%	6.22%	40.55%	65.63%	70.10%	4.82%	60.41%	79.80%
	Every day	8.21%	2.87%	2.42%	13.99%	0.76%	0.42%	-0.09%	1.61%

	Refuse/no answer	2.37%	1.75%	-1.15%	5.88%	0.00%	N.A.	N.A.	N.A.
Do the teachers at this school work together as teams to solve problems related to teaching?	Yes	97.06%	2.15%	92.74%	101.38%	95.66%	2.00%	91.64%	99.67%
	No	2.94%	2.15%	-1.38%	7.26%	4.34%	2.00%	0.33%	8.36%
	Do not know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
In this classroom in this shift, do you teach students from more than one grade?	Yes	0.37%	0.23%	-0.08%	0.83%	10.62%	4.22%	2.14%	19.09%
	No	99.63%	0.23%	99.17%	100.08%	89.38%	4.22%	80.91%	97.86%
	Do not know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
How frequently does the deputy director observe your classes?	Never	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Once a year	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Once every 6 months	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Once every 2–3 months	0.29%	0.19%	-0.09%	0.67%	4.26%	1.56%	1.13%	7.38%
	Once every month	10.17%	3.73%	2.65%	17.70%	12.86%	3.15%	6.53%	19.19%
	Once every two weeks	32.65%	6.22%	20.13%	45.17%	20.15%	3.50%	13.11%	27.18%
	Once every week	53.86%	6.02%	41.74%	65.98%	61.45%	4.75%	51.91%	70.99%
	Daily	3.03%	1.77%	-0.54%	6.59%	1.29%	0.80%	-0.33%	2.91%
When you need some help with your teaching, who do you consult?	Never need help	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	No one to ask for help	0.00%	N.A.	N.A.	N.A.	0.45%	0.31%	-0.18%	1.08%
	Organize meetings with teachers	32.40%	6.43%	19.45%	45.34%	36.56%	5.66%	25.19%	47.94%
	Discuss casually with teachers	47.65%	6.10%	35.37%	59.93%	62.36%	4.58%	53.15%	71.56%
	Head teacher	4.92%	1.60%	1.70%	8.15%	4.62%	1.43%	1.74%	7.50%
	Assistant head teacher	54.68%	6.22%	42.15%	67.22%	38.04%	3.81%	30.39%	45.70%
	Seek advice from education supervisor or subject specialist	71.36%	5.50%	60.27%	82.44%	58.33%	5.45%	47.38%	69.29%
	Other	0.00%	N.A.	N.A.	N.A.	3.27%	1.20%	0.85%	5.69%
	Don't know/refuse	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.

How do you use the results of students' oral and written assessments in your teaching?	Grade students	52.87%	6.13%	40.52%	65.22%	38.24%	5.08%	28.03%	48.45%
	Evaluate students' understanding of subject matter	59.46%	4.73%	49.92%	68.99%	61.37%	4.29%	52.75%	69.99%
	Plan teaching activities	35.10%	5.46%	24.10%	46.09%	41.87%	5.26%	31.30%	52.44%
	Adapt teaching to better suit students' needs	19.08%	5.76%	7.48%	30.68%	33.51%	5.12%	23.23%	43.80%
	Other	19.06%	3.35%	12.32%	25.80%	19.32%	3.15%	12.99%	25.66%
	Do not know / refuse to respond	0.00%	N.A.	N.A.	N.A.	0.97%	0.50%	-0.03%	1.96%
In your class, how many parents or guardians review students' homework?	None	0.00%	N.A.	N.A.	N.A.	0.34%	0.31%	-0.28%	0.97%
	Some	18.35%	5.25%	7.77%	28.92%	21.56%	3.82%	13.88%	29.24%
	Most	70.30%	5.69%	58.83%	81.77%	62.13%	5.11%	51.84%	72.41%
	All	11.35%	2.84%	5.64%	17.06%	15.97%	3.75%	8.43%	23.52%
	Don't know/refuse	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Are you generally satisfied with parents' involvement in their children's schoolwork?	Yes	71.24%	6.49%	58.17%	84.31%	72.36%	5.21%	61.88%	82.84%
	No	28.76%	6.49%	15.69%	41.83%	27.64%	5.21%	17.16%	38.12%
	Do not know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.

Table F4: Endline (2021) Teacher Survey: Means of Responses by Question, Russian Schools

Question	Russian Grade 2				Russian Grade 4			
	Mean	SE	95% Conf. Interval		Mean	SE	95% Conf. Interval	
			Lower	Upper			Lower	Upper
How many years of teaching experience do you have?	24.06	1.28	21.49	26.64	19.07	0.96	17.14	21.01
How many hours of in-service training on how to teach reading have you received?	63.49	1.70	58.77	68.21	53.22	14.76	6.25	100.19
In this class, how many boys are enrolled?	21.26	0.63	19.99	22.52	22.89	0.78	21.31	24.46

In this class, how many girls are enrolled?	17.55	0.64	16.27	18.84	15.34	0.55	14.24	16.45
---	-------	------	-------	-------	-------	------	-------	-------

ANNEX G: DETAILED DIRECTOR SSME TABLES

Table G1: Endline (2021) School Director Tool: Frequency of Responses by Question, All Schools

Question	Responses	Tajik				Russian			
		Proportion	SE	95% Conf. Interval		Proportion/Mean	SE	95% Conf. Interval	
				Lower	Upper			Lower	Upper
What is your position at this school?	Principal	41.83%	4.86%	32.20%	51.45%	19.28%	4.39%	10.43%	28.12%
	Deputy director	58.17%	4.86%	48.55%	67.80%	80.72%	4.39%	71.88%	89.57%
	Neither is available	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Participant sex	Male	42.30%	4.96%	32.48%	52.12%	9.64%	4.19%	1.19%	18.10%
	Female	57.70%	4.96%	47.88%	67.52%	90.36%	4.19%	81.90%	98.81%
Does your school have any programming support to help you be more effective in teaching students with disabilities?	Yes	22.23%	4.87%	12.59%	31.87%	20.72%	3.06%	14.55%	26.88%
	No	77.77%	4.87%	68.13%	87.41%	79.28%	3.06%	73.12%	85.45%
Who is providing program support to help you be more effective in teaching students with disabilities?	The state	22.32%	14.08%	-9.05%	53.69%	93.86%	1.12%	91.39%	96.32%
	International organizations	51.90%	18.23%	11.28%	92.52%	26.27%	14.27%	-5.14%	57.69%
	Public organizations	45.05%	19.30%	2.05%	88.06%	15.89%	7.78%	-1.24%	33.02%
	Other	2.20%	2.55%	-3.49%	7.89%	0.00%	N.A.	N.A.	N.A.
Outside of in-service teacher training, have any of your primary school teachers received training on how to teach reading?	Yes	97.15%	1.26%	94.66%	99.64%	95.05%	2.04%	90.94%	99.16%
	No	2.56%	1.24%	0.10%	5.01%	4.95%	2.04%	0.84%	9.06%
	Don't know	0.29%	0.21%	-0.12%	0.71%	0.00%	N.A.	N.A.	N.A.

If any of your primary school teachers have received training on how to teach reading, who provided those courses?	Advanced training courses in the institutes of teacher training	34.53%	3.83%	26.95%	42.11%	48.30%	6.62%	34.92%	61.69%
	Courses, provided by international agencies or projects	98.81%	0.49%	97.85%	99.78%	96.77%	2.52%	91.68%	101.86%
	Other	10.43%	3.32%	3.86%	17.00%	0.00%	N.A.	N.A.	N.A.
	Don't know	0.38%	0.36%	-0.33%	1.10%	3.23%	2.52%	-1.86%	8.32%
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
If you do go into classrooms to observe your teachers teaching, how often?	Once every 2-3 months	2.21%	1.12%	0.00%	4.42%	1.58%	0.96%	-0.36%	3.52%
	Once every month	9.77%	2.49%	4.86%	14.69%	2.13%	0.88%	0.36%	3.90%
	Once every two weeks	24.19%	4.60%	15.10%	33.29%	5.81%	1.52%	2.74%	8.88%
	Once every week	56.42%	5.03%	46.46%	66.38%	66.38%	5.67%	54.95%	77.82%
	Daily	7.41%	2.38%	2.69%	12.12%	24.10%	5.71%	12.59%	35.61%
	Do not know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Do you have a feeding program at school?	Yes	46.45%	4.72%	37.11%	55.79%	18.17%	2.84%	12.45%	23.89%
	No	53.55%	4.72%	44.21%	62.89%	81.83%	2.84%	76.11%	87.55%
	Don't Know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Do you have sufficient resource materials or textbooks?	Yes	66.65%	4.93%	56.91%	76.40%	21.02%	3.67%	13.62%	28.41%
	No	33.12%	4.92%	23.38%	42.86%	78.98%	3.67%	71.59%	86.38%
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Do you have a library or reading room?	Yes	93.95%	2.74%	88.53%	99.37%	99.37%	0.21%	98.94%	99.80%
	No	6.05%	2.74%	0.63%	11.47%	0.63%	0.21%	0.20%	1.06%
Did the librarian have any trainings or courses for the last five years?	Yes	75.24%	4.63%	66.08%	84.40%	86.69%	2.74%	81.17%	92.20%
	No	24.76%	4.63%	15.60%	33.92%	13.31%	2.74%	7.80%	18.83%

Do you hold regular parents-teachers association meetings?	Yes	99.85%	0.14%	99.56%	100.13%	100.00%	0.00%	100.00%	100.00%
	No	0.15%	0.14%	-0.13%	0.44%	0.00%	N.A.	N.A.	N.A.
How many parents come to the PTA meetings?	Few	7.81%	2.12%	3.62%	12.00%	4.89%	2.29%	0.28%	9.50%
	Some	22.58%	4.26%	14.14%	31.01%	9.76%	4.12%	1.45%	18.07%
	Most	69.61%	4.60%	60.51%	78.70%	85.35%	4.66%	75.97%	94.74%
Are you generally satisfied with the level of support the parent-teacher association provides to the school?	Yes	91.03%	2.04%	87.00%	95.06%	75.26%	5.62%	63.93%	86.58%
	No	8.45%	2.03%	4.43%	12.46%	24.74%	5.62%	13.42%	36.07%
	Don't know/refuse	0.53%	0.24%	0.05%	1.00%	0.00%	N.A.	N.A.	N.A.
Percentage of students with disabilities by grade	Grade 1	6.41%	3.14%	0.19%	12.64%	0.50%	0.10%	0.30%	0.69%
	Grade 2	6.44%	3.12%	0.25%	12.64%	0.40%	0.09%	0.22%	0.57%
	Grade 3	6.42%	3.22%	0.03%	12.80%	0.38%	0.11%	0.16%	0.60%
	Grade 4	6.40%	3.05%	0.34%	12.47%	0.52%	0.10%	0.32%	0.72%

Table G2: Endline (2021) School Director Tool: Means of Responses by Question, All Schools

Question	Responses	Tajik				Russian			
		Mean	SE	95% Conf. Interval		Mean	SE	95% Conf. Interval	
				Lower	Upper			Lower	Upper
How many years have you been a principal?	All	8.56	0.78	7.02	10.09	7.48	0.81	5.85	9.12
Enrollment by grade	Grade 1	131.20	11.16	109.12	153.29	213.32	19.98	173.06	253.58
	Grade 2	125.38	10.49	104.64	146.13	209.96	17.69	174.31	245.61
	Grade 3	123.05	9.26	104.74	141.36	208.00	20.61	166.47	249.53
	Grade 4	129.22	10.90	107.65	150.78	183.12	14.37	154.15	212.09
Girl-to-boy ratio by grade	Grade 1	1.04	0.03	0.98	1.09	0.73	0.02	0.69	0.77
	Grade 2	1.00	0.02	0.95	1.05	0.74	0.01	0.72	0.77

	Grade 3	1.03	0.02	0.99	1.07	0.98	0.18	0.62	1.35
	Grade 4	1.01	0.02	0.97	1.05	0.68	0.02	0.65	0.71
Student-to-class ratio by grade	Grade 1	26.56	0.73	25.12	28.00	36.54	0.69	35.15	37.93
	Grade 2	25.85	0.50	24.87	26.84	37.71	0.51	36.68	38.75
	Grade 3	25.26	0.53	24.21	26.30	36.23	0.85	34.51	37.95
	Grade 4	25.39	0.55	24.31	26.48	36.42	0.58	35.25	37.59

ANNEX H: DETAILED SCHOOL INVENTORY SSME TABLES

Table H1: Endline (2021) School Inventory Tool: Frequency of Responses by Question, All Schools

Questions	Response options	Tajik				Russian			
		Proportion	SE	95% Conf. Interval		Proportion	SE	95% Conf. Interval	
				Lower	Upper			Lower	Upper
Type of school	Primary (Grades 1–4)	3.14%	1.58%	0.01%	6.26%	0.00%	N.A.	N.A.	N.A.
	Compulsory (Grades 1–9)	1.78%	0.75%	0.30%	3.26%	2.09%	1.45%	-0.84%	5.02%
	Secondary complete (Grades 1–11)	95.08%	1.72%	91.68%	98.48%	97.91%	1.45%	94.98%	100.84%
	Boarding school	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
If the school is rural, how far is it from the town or district center?	5–20 km	70.59%	4.98%	60.70%	80.49%	100.00%	0.00%	100.00%	100.00%
	21–40 km	19.26%	4.36%	10.61%	27.91%	0.00%	N.A.	N.A.	N.A.
	41–60 km	5.10%	2.01%	1.10%	9.10%	0.00%	N.A.	N.A.	N.A.
	60–100 km	1.32%	0.78%	-0.23%	2.87%	0.00%	N.A.	N.A.	N.A.
	More than 100 km	3.73%	1.79%	0.18%	7.28%	0.00%	N.A.	N.A.	N.A.
	Skipped	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Language of instruction	Tajik	100.00%	0.00%	100.00%	100.00%	96.66%	1.48%	93.69%	99.64%
	Russian	6.03%	2.68%	0.73%	11.34%	92.79%	3.28%	86.19%	99.39%

	Uzbek	17.39%	4.11%	9.26%	25.52%	3.99%	1.28%	1.42%	6.56%
	Kyrgyz	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Other	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
How many shifts are there in the school?	1	5.99%	1.70%	2.63%	9.35%	0.00%	N.A.	N.A.	N.A.
	2	87.93%	3.48%	81.05%	94.80%	99.50%	0.45%	98.59%	100.41%
	3	6.09%	3.20%	-0.25%	12.42%	0.50%	0.45%	-0.41%	1.41%
Is there a library in the school?	Yes	97.67%	0.80%	96.09%	99.26%	99.38%	0.21%	98.95%	99.80%
	No	2.33%	0.80%	0.74%	3.91%	0.62%	0.21%	0.20%	1.05%
If yes, are there any students at the time of the visit?	Yes	37.01%	5.08%	26.95%	47.08%	28.55%	4.45%	19.58%	37.52%
	No	62.99%	5.08%	52.92%	73.05%	71.45%	4.45%	62.48%	80.42%
	Skipped	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
If no, indicate for what reason.	The library is closed or locked	4.03%	1.98%	0.10%	7.97%	20.11%	6.36%	7.07%	33.15%
	Students are all in class	91.99%	2.97%	86.09%	97.90%	79.89%	6.36%	66.85%	92.93%
	Other	3.37%	2.20%	-1.01%	7.75%	0.00%	N.A.	N.A.	N.A.
	Don't know	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Refuse/no answer	0.61%	0.57%	-0.52%	1.73%	0.00%	N.A.	N.A.	N.A.
Approximately how many books for primary students are there in the library?	No books	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	1-50	0.00%	N.A.	N.A.	N.A.	5.89%	1.63%	2.61%	9.17%
	51-100	3.81%	1.13%	1.56%	6.06%	15.93%	3.17%	9.53%	22.32%
	More than 100	92.60%	1.77%	89.09%	96.11%	63.82%	5.43%	52.87%	74.77%
	Do not know (unable to access the library)	3.59%	1.31%	1.00%	6.18%	14.37%	3.84%	6.63%	22.11%
	Refuse/no answer	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
The books provided by the project are available.	Yes	94.47%	2.09%	90.34%	98.59%	83.33%	4.25%	74.77%	91.88%
	No	2.61%	1.75%	-0.85%	6.07%	2.39%	1.30%	-0.23%	5.01%
	Don't know	2.93%	1.24%	0.48%	5.37%	14.28%	4.02%	6.18%	22.38%

All four parts of the project books logbook are completed correctly and timely.	Yes	51.39%	4.91%	41.67%	61.10%	61.51%	5.80%	49.82%	73.20%
	No	43.24%	4.72%	33.90%	52.59%	24.21%	4.92%	14.29%	34.13%
	Don't know	5.37%	2.11%	1.19%	9.55%	14.28%	4.02%	6.18%	22.38%
The last quarterly report to the District Education Department (DED) appears complete in the logbook.	Yes	71.71%	4.24%	63.33%	80.10%	59.72%	5.75%	48.15%	71.30%
	No	22.34%	4.00%	14.43%	30.25%	26.00%	4.83%	16.27%	35.72%
	Don't know	5.95%	2.19%	1.61%	10.29%	14.28%	4.02%	6.18%	22.38%
A reading corner has been organized in the library.	Yes	73.28%	4.13%	65.10%	81.46%	83.04%	4.21%	74.55%	91.52%
	No	23.55%	3.93%	15.78%	31.32%	2.68%	1.07%	0.52%	4.85%
	Don't know	3.17%	1.25%	0.70%	5.64%	14.28%	4.02%	6.18%	22.38%
The librarian logbook records that an extracurricular reading event for primary school students was held within the last 30 days.	Yes	53.37%	5.12%	43.24%	63.49%	63.19%	5.72%	51.67%	74.70%
	No	41.26%	4.93%	31.50%	51.02%	22.53%	4.79%	12.89%	32.18%
	Don't know	5.37%	2.11%	1.19%	9.55%	14.28%	4.02%	6.18%	22.38%
The school building and the school grounds are clean and tidy?	Yes	95.82%	2.09%	91.69%	99.95%	97.68%	1.58%	94.49%	100.87%
	No	4.18%	2.09%	0.05%	8.31%	2.32%	1.58%	-0.87%	5.51%

Table H2: Endline (2021) School Inventory Tool: Means of Responses by Question, Tajik Schools

Question	Tajik Grade 2				Tajik Grade 4			
	Mean	SE	95% Conf. Interval		Mean	SE	95% Conf. Interval	
			Lower	Upper			Lower	Upper
School environment index (out of 4)=1	0.02	0.01	-0.01	0.05	0.02	0.02	-0.02	0.05
School environment index (out of 4)=2	0.25	0.04	0.17	0.33	0.22	0.04	0.14	0.31
School environment index (out of 4)=3	0.73	0.04	0.64	0.81	0.75	0.05	0.66	0.84
School environment index (out of 4)=4	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.02

Table H3: Endline (2021) School Inventory Tool: Frequency of Responses by Question, Russian Schools

Question	Russian Grade 2				Russian Grade 4			
	Mean	SE	95% Conf. Interval		Mean	SE	95% Conf. Interval	
			Lower	Upper			Lower	Upper
School environment index (out of 4)=1	0.00	N.A.	N.A.	N.A.	0.00	N.A.	N.A.	N.A.
School environment index (out of 4)=2	0.05	0.02	0.01	0.09	0.07	0.02	0.02	0.12
School environment index (out of 4)=3	0.95	0.02	0.91	0.99	0.93	0.02	0.88	0.98
School environment index (out of 4)=4	0.00	N.A.	N.A.	N.A.	0.00	N.A.	N.A.	N.A.

ANNEX I: DETAILED CLASSROOM OBSERVATION SSME TABLES

Table 11: Endline (2021) Classroom Observation Tool: Frequency of Responses by Question, Tajik Schools

Questions	Response options	Tajik Grade 2				Tajik Grade 4			
		Proportion	SE	95% Conf. Interval		Proportion	SE	95% Conf. Interval	
				Lower	Upper			Lower	Upper
The teacher explicitly articulates the objectives of the lesson and relates classroom activities to the objectives.	Yes	81.17%	3.37%	74.50%	87.83%	70.91%	4.96%	61.09%	80.73%
	No	18.83%	3.37%	12.17%	25.50%	29.09%	4.96%	19.27%	38.91%
The teacher's explanation of content is clear.	Yes	57.26%	4.79%	47.78%	66.73%	54.23%	5.54%	43.27%	65.20%
	No	42.74%	4.79%	33.27%	52.22%	45.77%	5.54%	34.80%	56.73%
The teacher makes connections in the lesson that relate to other content knowledge or students' daily lives.	Yes	32.62%	5.59%	21.56%	43.69%	34.46%	6.22%	22.15%	46.78%
	No	67.38%	5.59%	56.31%	78.44%	65.54%	6.22%	53.22%	77.85%
During the lesson, which of the following reading skills were developed or formed?	Looking for text conventions	18.68%	3.61%	11.54%	25.82%	19.11%	4.03%	11.13%	27.08%
	Phonemic consciousness (working with sounds)	52.69%	5.26%	42.27%	63.10%	31.04%	6.16%	18.85%	43.23%

	Fluent reading	83.57%	3.41%	76.82%	90.32%	84.91%	3.76%	77.48%	92.35%
	Vocabulary (passive or active vocabulary)	77.09%	4.04%	69.09%	85.09%	80.85%	4.52%	71.91%	89.79%
	Reading comprehension	67.48%	4.69%	58.20%	76.75%	84.57%	3.42%	77.81%	91.33%
	None of the above	0.74%	0.53%	-0.31%	1.79%	1.62%	1.52%	-1.39%	4.63%
Teacher did the following pre-reading activities: Work with students to predict the content or themes of the text from an illustration or picture related to the text	Yes	40.23%	5.18%	29.99%	50.47%	36.78%	5.81%	25.28%	48.27%
	No	59.77%	5.18%	49.53%	70.01%	63.22%	5.81%	51.73%	74.72%
Teacher did the following while reading the text: Work with students to predict the meaning of the text based on the text title	Yes	52.57%	5.12%	42.44%	62.70%	55.73%	5.66%	44.52%	66.93%
	No	47.43%	5.12%	37.30%	57.56%	44.27%	5.66%	33.07%	55.48%
Teacher did the following pre-reading activities: Ask questions or lead a discussion related to the theme of the text	Yes	87.84%	2.64%	82.61%	93.07%	86.28%	3.18%	79.98%	92.58%
	No	12.16%	2.64%	6.93%	17.39%	13.72%	3.18%	7.42%	20.02%
Teacher did the following while reading the text: Define key words in the text and teach them	Yes	17.72%	5.02%	7.80%	27.64%	17.35%	4.73%	7.98%	26.71%
	No	82.28%	5.02%	72.36%	92.20%	82.65%	4.73%	73.29%	92.02%
Teacher did the following pre-reading activities: Introduce new vocabulary words to help with reading	Yes	41.58%	5.24%	31.21%	51.94%	33.12%	6.22%	20.82%	45.43%
	No	58.42%	5.24%	48.06%	68.79%	66.88%	6.22%	54.57%	79.18%
Teacher did the following while reading the text: Model reading the text aloud, when the students listened attentively.	Yes	54.49%	4.76%	45.08%	63.90%	53.23%	5.27%	42.81%	63.65%
	No	45.51%	4.76%	36.10%	54.92%	46.77%	5.27%	36.35%	57.19%
Teacher did the following while reading the text: Teacher and students develop success criteria for the reading activity together	Yes	2.57%	1.13%	0.34%	4.80%	1.69%	1.34%	-0.95%	4.33%
	No	97.43%	1.13%	95.20%	99.66%	98.31%	1.34%	95.67%	100.95%

Teacher did the following pre-reading activities: None of the above	Yes	0.35%	0.17%	0.00%	0.69%	0.88%	0.56%	-0.24%	1.99%
	No	99.65%	0.17%	99.31%	100.00%	99.12%	0.56%	98.01%	100.24%
During the while-reading activity, did the teacher: Model reading or play audio recordings of the text being read, when the students follow the text with a finger.	Yes	72.05%	4.57%	63.01%	81.08%	73.44%	5.00%	63.55%	83.32%
	No	27.95%	4.57%	18.92%	36.99%	26.56%	5.00%	16.68%	36.45%
During the while-reading activity, did the teacher: Practice reading through choral reading, pair and group reading	Yes	38.26%	5.19%	27.99%	48.52%	28.23%	5.08%	18.17%	38.29%
	No	61.74%	5.19%	51.48%	72.01%	71.77%	5.08%	61.71%	81.83%
During the while-reading activity, did the teacher: Give (comprehension) tasks to students to do while reading	Yes	75.10%	4.19%	66.82%	83.39%	85.11%	3.73%	77.74%	92.48%
	No	24.90%	4.19%	16.61%	33.18%	14.89%	3.73%	7.52%	22.26%
During the while-reading activity, did the teacher: Use any supplemental materials, such as: Project books, Other books, Texts (handouts and digital), Picture dictionaries and cards, Reading books	Yes	44.59%	5.33%	34.04%	55.14%	51.32%	6.07%	39.30%	63.34%
	No	55.41%	5.33%	44.86%	65.96%	48.68%	6.07%	36.66%	60.70%
During the while-reading activity, did the teacher: Ask students to practice chain reading or individual, silent reading	Yes	87.58%	2.87%	81.90%	93.27%	93.21%	2.92%	87.42%	99.00%
	No	12.42%	2.87%	6.73%	18.10%	6.79%	2.92%	1.00%	12.58%
During the while-reading activity, did the teacher: did none of the above	Yes	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	No	100.00%	0.00%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%
During the post-reading activity, did the teacher: related the text to art	Yes	16.10%	4.71%	6.77%	25.42%	28.08%	5.54%	17.11%	39.05%
	No	83.90%	4.71%	74.58%	93.23%	71.92%	5.54%	60.95%	82.89%
During the post-reading activity, did the teacher write conclusions related to the text	Yes	20.88%	4.48%	12.00%	29.75%	39.78%	6.28%	27.36%	52.21%
	No	79.12%	4.48%	70.25%	88.00%	60.22%	6.28%	47.79%	72.64%
	Yes	70.05%	4.35%	61.43%	78.66%	77.52%	4.61%	68.41%	86.64%

During the post-reading activity, did the teacher: conduct student-centered activities	No	29.95%	4.35%	21.34%	38.57%	22.48%	4.61%	13.36%	31.59%
During the post-reading activity, did the teacher: role-play	Yes	13.66%	3.62%	6.49%	20.83%	22.32%	5.97%	10.51%	34.14%
	No	86.34%	3.62%	79.17%	93.51%	77.68%	5.97%	65.86%	89.49%
During the post-reading activity, did the teacher: modelled an action	Yes	44.04%	5.08%	33.99%	54.09%	51.69%	5.98%	39.86%	63.52%
	No	55.96%	5.08%	45.91%	66.01%	48.31%	5.98%	36.48%	60.14%
During the post-reading activity, did the teacher: revisited success criteria	Yes	1.49%	0.93%	-0.35%	3.34%	1.72%	1.34%	-0.94%	4.37%
	No	98.51%	0.93%	96.66%	100.35%	98.28%	1.34%	95.63%	100.94%
During the post-reading activity, did the teacher: use questions, prompts or other strategies to determine students' level of understanding	Yes	74.43%	4.02%	66.47%	82.38%	85.66%	3.06%	79.61%	91.71%
	No	25.57%	4.02%	17.62%	33.53%	14.34%	3.06%	8.29%	20.39%
During the post-reading activity, did the teacher: use any supplemental materials	Yes	56.16%	5.21%	45.84%	66.47%	55.02%	5.76%	43.62%	66.41%
	No	43.84%	5.21%	33.53%	54.16%	44.98%	5.76%	33.59%	56.38%
During the post-reading activity, did the teacher: none of the above	Yes	4.85%	2.02%	0.86%	8.85%	1.16%	0.79%	-0.40%	2.73%
	No	95.15%	2.02%	91.15%	99.14%	98.84%	0.79%	97.27%	100.40%
During the entirety of the lesson, did the teacher use the Gradual Release of Responsibility Approach?	Yes	80.08%	3.71%	72.74%	87.43%	78.38%	3.94%	70.58%	86.19%
	No	19.92%	3.71%	12.57%	27.26%	21.62%	3.94%	13.81%	29.42%
During the entirety of the lesson, did the teacher use differentiated instruction?	Yes	57.44%	4.83%	47.87%	67.00%	64.06%	5.44%	53.30%	74.82%
	No	42.56%	4.83%	33.00%	52.13%	35.94%	5.44%	25.18%	46.70%
During the entirety of the lesson, did the teacher use a guided group strategy?	Yes	23.29%	4.92%	13.55%	33.03%	24.31%	6.04%	12.35%	36.26%
	No	76.71%	4.92%	66.97%	86.45%	75.69%	6.04%	63.74%	87.65%
During the entirety of the lesson, did the teacher use wait time as a strategy?	Yes	55.28%	5.14%	45.10%	65.45%	67.70%	5.32%	57.18%	78.23%
	No	44.72%	5.14%	34.55%	54.90%	32.30%	5.32%	21.77%	42.82%
	Board (white or plain)	99.63%	0.34%	98.96%	100.30%	99.11%	0.89%	97.35%	100.87%

Does the teacher have the following materials?	Chalk or markers for white board	98.92%	0.96%	97.01%	100.83%	99.72%	0.28%	99.17%	100.27%
	Interactive board	2.13%	0.58%	0.99%	3.27%	3.66%	1.35%	1.00%	6.33%
	Visual aids	67.21%	4.47%	58.36%	76.06%	69.04%	4.97%	59.20%	78.88%
	None of the above	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Teacher has lesson plans developed by the teacher himself.	Yes	97.18%	1.39%	94.43%	99.93%	99.62%	0.29%	99.05%	100.19%
	No	2.82%	1.39%	0.07%	5.57%	0.38%	0.29%	-0.19%	0.95%
Teacher has a reading guide or methodical guide.	Yes	83.68%	3.73%	76.30%	91.06%	83.21%	4.08%	75.14%	91.27%
	No	16.32%	3.73%	8.94%	23.70%	16.79%	4.08%	8.73%	24.86%
Number of books other than textbooks that are available and accessible to students.	None	20.90%	3.82%	13.34%	28.45%	19.17%	3.96%	11.34%	26.99%
	1–4	10.85%	3.22%	4.47%	17.22%	2.45%	1.19%	0.10%	4.80%
	5–9	17.64%	4.05%	9.63%	25.64%	19.04%	4.41%	10.32%	27.76%
	10–19	27.63%	4.23%	19.26%	36.00%	45.86%	6.07%	33.85%	57.87%
	20–39	19.85%	4.97%	10.01%	29.69%	10.76%	3.56%	3.73%	17.80%
	40+	3.14%	1.42%	0.34%	5.95%	2.72%	1.17%	0.40%	5.04%
Class has newspapers and magazines that students can read.	Yes	53.51%	4.62%	44.37%	62.65%	61.25%	5.28%	50.81%	71.69%
	No	46.49%	4.62%	37.35%	55.63%	38.75%	5.28%	28.31%	49.19%
At least 90 percent of students in the class have reading books.	Yes	86.35%	3.26%	79.89%	92.80%	84.66%	3.65%	77.45%	91.88%
	No	13.65%	3.26%	7.20%	20.11%	15.34%	3.65%	8.12%	22.55%
At least 90 percent of students in the class have language books.	Yes	97.00%	1.43%	94.17%	99.83%	97.11%	1.78%	93.58%	100.63%
	No	3.00%	1.43%	0.17%	5.83%	2.89%	1.78%	-0.63%	6.42%
At least 90 percent of students in the class have vocabulary books.	Yes	47.92%	5.26%	37.52%	58.32%	60.03%	5.70%	48.76%	71.30%
	No	52.08%	5.26%	41.68%	62.48%	39.97%	5.70%	28.70%	51.24%
Are RWM reading cards being used	Not present	88.28%	8.48%	69.38%	107.19%	86.81%	11.02%	62.25%	111.38%
	Present (on the walls)	2.12%	2.24%	-2.86%	7.10%	0.00%	N.A.	N.A.	N.A.

	Used in the lesson	9.59%	7.89%	-7.98%	27.17%	13.19%	11.02%	-11.38%	37.75%
--	--------------------	-------	-------	--------	--------	--------	--------	---------	--------

Table 12: Endline (2021) Classroom Observation Tool: Means of Responses by Question, Tajik Schools

Questions	Tajik Grade 2				Tajik Grade 4			
	Mean	SE	95% Conf. Interval		Mean	SE	95% Conf. Interval	
			Lower	Upper			Lower	Upper
How many boys are present in the class at the time of your visit?	11.85	0.36	11.13	12.57	11.85	0.53	10.81	12.89
How many girls are present in the class at the time of your visit?	12.38	0.48	11.43	13.34	12.96	0.44	12.09	13.83
How much time did the teacher spend on pre-reading activities?	8.15	0.53	7.10	9.20	8.33	0.78	6.79	9.87
How much time did the teacher spend on reading activities?	21.92	0.65	20.64	23.21	21.82	0.81	20.21	23.43
How much time did the teacher spend on post-reading activities?	17.92	1.01	15.93	19.91	19.53	0.92	17.72	21.35

Table 13: Endline (2021) Classroom Observation Tool: Frequency of Responses by Question, Russian Schools

Questions	Response options	Russian Grade 2				Russian Grade 4			
		Proportion	SE	95% Conf. Interval		Proportion	SE	95% Conf. Interval	
				Lower	Upper			Lower	Upper
The teacher explicitly articulates the objectives of the lesson and relates classroom activities to the objectives.	Yes	60.13%	5.51%	49.03%	71.23%	65.22%	5.68%	53.80%	76.64%
	No	39.87%	5.51%	28.77%	50.97%	34.78%	5.68%	23.36%	46.20%
The teacher's explanation of content is clear.	Yes	64.07%	5.12%	53.75%	74.39%	63.00%	5.45%	52.03%	73.96%
	No	35.93%	5.12%	25.61%	46.25%	37.00%	5.45%	26.04%	47.97%
	Yes	31.42%	6.03%	19.27%	43.57%	38.11%	4.69%	28.68%	47.54%

The teacher makes connections in the lesson that relate to other content knowledge or students' daily lives.	No	68.58%	6.03%	56.43%	80.73%	61.89%	4.69%	52.46%	71.32%
During the lesson, which of the following reading skills were developed or formed?	Looking for text conventions	37.31%	6.40%	24.43%	50.19%	28.00%	4.92%	18.10%	37.90%
	Phonemic consciousness (working with sounds)	33.45%	5.27%	22.83%	44.08%	9.25%	2.79%	3.63%	14.86%
	Fluent reading	66.81%	4.50%	57.74%	75.89%	69.85%	4.03%	61.74%	77.96%
	Vocabulary (passive or active vocabulary)	87.86%	2.70%	82.42%	93.31%	77.19%	3.44%	70.28%	84.10%
	Reading comprehension	71.03%	4.21%	62.54%	79.51%	74.65%	5.09%	64.42%	84.88%
	None of the above	4.56%	1.96%	0.60%	8.52%	2.30%	0.99%	0.31%	4.29%
Teacher did the following pre-reading activities: Work with students to predict the content or themes of the text from an illustration or picture related to the text	Yes	46.54%	6.06%	34.34%	58.75%	45.42%	5.80%	33.75%	57.09%
	No	53.46%	6.06%	41.25%	65.66%	54.58%	5.80%	42.91%	66.25%
Teacher did the following while reading the text: Work with students to predict the meaning of the text based on the text title	Yes	52.88%	5.59%	41.61%	64.14%	59.17%	4.85%	49.42%	68.92%
	No	47.12%	5.59%	35.86%	58.39%	40.83%	4.85%	31.08%	50.58%
Teacher did the following pre-reading activities: Ask questions or lead a discussion related to the theme of the text	Yes	82.40%	3.27%	75.81%	89.00%	69.54%	3.62%	62.26%	76.81%
	No	17.60%	3.27%	11.00%	24.19%	30.46%	3.62%	23.19%	37.74%
Teacher did the following while reading the text: Define key words in the text and teach them	Yes	55.37%	4.65%	46.00%	64.74%	54.13%	4.57%	44.94%	63.33%
	No	44.63%	4.65%	35.26%	54.00%	45.87%	4.57%	36.67%	55.06%

Teacher did the following pre-reading activities: Introduce new vocabulary words to help with reading	Yes	73.90%	4.28%	65.28%	82.53%	64.68%	4.69%	55.25%	74.10%
	No	26.10%	4.28%	17.47%	34.72%	35.32%	4.69%	25.90%	44.75%
Teacher did the following while reading the text: Model reading the text aloud, when the students listened attentively.	Yes	69.18%	5.24%	58.62%	79.75%	66.05%	3.86%	58.29%	73.80%
	No	30.82%	5.24%	20.25%	41.38%	33.95%	3.86%	26.20%	41.71%
Teacher did the following while reading the text: Teacher and students develop success criteria for the reading activity together	Yes	30.20%	6.29%	17.54%	42.87%	33.48%	5.84%	21.74%	45.22%
	No	69.80%	6.29%	57.13%	82.46%	66.52%	5.84%	54.78%	78.26%
Teacher did the following pre-reading activities: None of the above	Yes	5.30%	2.12%	1.04%	9.57%	9.04%	2.20%	4.61%	13.46%
	No	94.70%	2.12%	90.43%	98.96%	90.96%	2.20%	86.54%	95.39%
During the while-reading activity, did the teacher: Model reading or play audio recordings of the text being read, when the students follow the text with a finger.	Yes	71.01%	5.40%	60.14%	81.89%	68.05%	5.24%	57.51%	78.58%
	No	28.99%	5.40%	18.11%	39.86%	31.95%	5.24%	21.42%	42.49%
During the while-reading activity, did the teacher: Practice reading through choral reading, pair and group reading	Yes	62.89%	4.48%	53.86%	71.91%	53.79%	4.24%	45.27%	62.31%
	No	37.11%	4.48%	28.09%	46.14%	46.21%	4.24%	37.69%	54.73%
During the while-reading activity, did the teacher: Give (comprehension) tasks to students to do while reading	Yes	77.31%	4.83%	67.57%	87.04%	79.54%	3.66%	72.18%	86.91%
	No	22.69%	4.83%	12.96%	32.43%	20.46%	3.66%	13.09%	27.82%
During the while-reading activity, did the teacher: Use any supplemental materials, such as: Project books, Other books, Texts (handouts and digital), Picture dictionaries and cards, Reading books	Yes	44.42%	6.23%	31.86%	56.97%	51.79%	5.45%	40.82%	62.75%
	No	55.58%	6.23%	43.03%	68.14%	48.21%	5.45%	37.25%	59.18%
During the while-reading activity, did the teacher: Ask students to	Yes	77.54%	4.55%	68.36%	86.71%	77.56%	4.15%	69.21%	85.91%
	No	22.46%	4.55%	13.29%	31.64%	22.44%	4.15%	14.09%	30.79%

practice chain reading or individual, silent reading									
During the while-reading activity, did the teacher: did none of the above	Yes	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	No	100.00%	0.00%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%
During the post-reading activity, did the teacher: related the text to art	Yes	43.10%	6.23%	30.54%	55.65%	28.11%	5.08%	17.89%	38.32%
	No	56.90%	6.23%	44.35%	69.46%	71.89%	5.08%	61.68%	82.11%
During the post-reading activity, did the teacher: wrote conclusions related to the text	Yes	13.11%	5.87%	1.29%	24.94%	12.77%	3.58%	5.58%	19.97%
	No	86.89%	5.87%	75.06%	98.71%	87.23%	3.58%	80.03%	94.42%
During the post-reading activity, did the teacher: conduct student-centered activities	Yes	57.65%	5.74%	46.08%	69.22%	52.00%	4.26%	43.43%	60.58%
	No	42.35%	5.74%	30.78%	53.92%	48.00%	4.26%	39.42%	56.57%
During the post-reading activity, did the teacher: role-play	Yes	30.65%	6.34%	17.88%	43.42%	21.52%	5.28%	10.91%	32.13%
	No	69.35%	6.34%	56.58%	82.12%	78.48%	5.28%	67.87%	89.09%
During the post-reading activity, did the teacher: modelled an action	Yes	42.54%	5.26%	31.94%	53.14%	56.16%	5.50%	45.10%	67.23%
	No	57.46%	5.26%	46.86%	68.06%	43.84%	5.50%	32.77%	54.90%
During the post-reading activity, did the teacher: revisited success criteria	Yes	6.91%	2.96%	0.96%	12.86%	12.58%	4.46%	3.62%	21.55%
	No	93.09%	2.96%	87.14%	99.04%	87.42%	4.46%	78.45%	96.38%
During the post-reading activity, did the teacher: use questions, prompts or other strategies to determine students' level of understanding	Yes	81.89%	3.22%	75.40%	88.38%	82.08%	4.22%	73.59%	90.58%
	No	18.11%	3.22%	11.62%	24.60%	17.92%	4.22%	9.42%	26.41%
During the post-reading activity, did the teacher: use any supplemental materials	Yes	58.13%	5.40%	47.25%	69.02%	52.31%	5.32%	41.62%	63.00%
	No	41.87%	5.40%	30.98%	52.75%	47.69%	5.32%	37.00%	58.38%
During the post-reading activity, did the teacher: none of the above	Yes	9.09%	2.65%	3.75%	14.43%	5.25%	1.74%	1.75%	8.74%
	No	90.91%	2.65%	85.57%	96.25%	94.75%	1.74%	91.26%	98.25%
During the entirety of the lesson, did the teacher use the Gradual Release of Responsibility Approach?	Yes	70.76%	4.23%	62.24%	79.28%	68.26%	4.53%	59.15%	77.36%
	No	29.24%	4.23%	20.72%	37.76%	31.74%	4.53%	22.64%	40.85%

During the entirety of the lesson, did the teacher use differentiated instruction?	Yes	54.16%	5.48%	43.13%	65.19%	48.39%	5.56%	37.21%	59.58%
	No	45.84%	5.48%	34.81%	56.87%	51.61%	5.56%	40.42%	62.79%
During the entirety of the lesson, did the teacher use a guided group strategy?	Yes	46.53%	5.67%	35.12%	57.94%	50.77%	4.20%	42.32%	59.21%
	No	53.47%	5.67%	42.06%	64.88%	49.23%	4.20%	40.79%	57.68%
During the entirety of the lesson, did the teacher use wait time as a strategy?	Yes	66.65%	4.47%	57.64%	75.66%	66.11%	4.35%	57.35%	74.86%
	No	33.35%	4.47%	24.34%	42.36%	33.89%	4.35%	25.14%	42.65%
Does the teacher have the following materials?	Board (white or plain)	100.00%	0.00%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%
	Chalk or markers for white board	86.61%	3.13%	80.30%	92.92%	86.21%	2.70%	80.77%	91.64%
	Interactive board	26.48%	5.88%	14.64%	38.32%	14.37%	4.25%	5.82%	22.92%
	Visual aids	80.26%	3.14%	73.93%	86.60%	75.30%	2.80%	69.67%	80.93%
	None of the above	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
Teacher has lesson plans developed by the teacher himself.	Yes	89.04%	4.34%	80.31%	97.77%	94.26%	1.92%	90.41%	98.12%
	No	10.96%	4.34%	2.23%	19.69%	5.74%	1.92%	1.88%	9.59%
Teacher has a reading guide or methodical guide.	Yes	87.45%	2.76%	81.88%	93.02%	82.14%	2.82%	76.47%	87.80%
	No	12.55%	2.76%	6.98%	18.12%	17.86%	2.82%	12.20%	23.53%
Number of books other than textbooks that are available and accessible to students.	None	7.02%	2.23%	2.52%	11.51%	15.35%	4.08%	7.15%	23.56%
	1–4	3.10%	2.26%	-1.46%	7.66%	2.46%	1.41%	-0.38%	5.29%
	5–9	22.00%	5.23%	11.46%	32.54%	5.91%	1.59%	2.73%	9.10%
	10–19	29.74%	5.67%	18.33%	41.15%	49.93%	5.80%	38.27%	61.58%
	20–39	28.11%	5.81%	16.40%	39.81%	15.65%	3.20%	9.22%	22.07%
	40+	10.04%	4.07%	1.84%	18.23%	10.70%	4.38%	1.90%	19.51%
Class has newspapers and magazines that students can read.	Yes	21.66%	5.62%	10.34%	32.97%	21.08%	5.43%	10.17%	31.99%
	No	78.34%	5.62%	67.03%	89.66%	78.92%	5.43%	68.01%	89.83%

At least 90 percent of students in the class have reading books.	Yes	94.65%	1.95%	90.74%	98.57%	90.66%	1.85%	86.94%	94.38%
	No	5.35%	1.95%	1.43%	9.26%	9.34%	1.85%	5.62%	13.06%
At least 90 percent of students in the class have language books.	Yes	67.56%	3.76%	60.00%	75.12%	82.66%	2.43%	77.77%	87.54%
	No	32.44%	3.76%	24.88%	40.00%	17.34%	2.43%	12.46%	22.23%
At least 90 percent of students in the class have vocabulary books.	Yes	36.19%	5.55%	25.02%	47.36%	50.69%	5.36%	39.93%	61.45%
	No	63.81%	5.55%	52.64%	74.98%	49.31%	5.36%	38.55%	60.07%
Are RWM reading cards being used	Not present	100.00%	0.00%	100.00%	100.00%	100.00%	0.00%	100.00%	100.00%
	Present (on the walls)	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.
	Used in the lesson	0.00%	N.A.	N.A.	N.A.	0.00%	N.A.	N.A.	N.A.

Table 14: Endline (2021) Classroom Observation Tool: Means of Responses by Question, Russian Schools

Questions	Russian Grade 2				Russian Grade 4			
	Mean	SE	95% Conf. Interval		Mean	SE	95% Conf. Interval	
			Lower	Upper			Lower	Upper
How many boys are present in the class at the time of your visit?	18.84	0.44	17.95	19.73	19.75	0.47	18.80	20.70
How many girls are present in the class at the time of your visit?	14.90	0.60	13.69	16.11	12.68	0.50	11.68	13.69
How much time did the teacher spend on pre-reading activities?	7.01	0.26	6.48	7.54	7.07	0.32	6.42	7.72
How much time did the teacher spend on reading activities?	16.71	0.55	15.61	17.82	16.90	0.51	15.88	17.92
How much time did the teacher spend on post-reading activities?	9.67	0.47	8.73	10.61	9.66	0.40	8.84	10.47

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