



USAID
FROM THE AMERICAN PEOPLE



EARLY GRADE READING ASSESSMENT BASELINE REPORT

SINDH

MAY 2014

This publication was produced for review by the United States Agency for International Development. It was prepared by Management Systems International (MSI) with School-to-School International (STS) under the Monitoring and Evaluation Program (MEP).

EARLY GRADE READING ASSESSMENT BASELINE REPORT SINDH

Contracted under Order No. AID-391-C-13-00005

Monitoring and Evaluation Program (MEP)

DISCLAIMER

This study/report is made possible by the support of the American people through the United States Agency for International Development (USAID). The contents are the sole responsibility of Management Systems International and do not necessarily reflect the views of USAID or the United States Government.

ACKNOWLEDGEMENTS

We would like to thank the Education team of USAID/Pakistan for their forward planning to be able to collect baseline data before the roll out of the two important reading programs. Their support and responsiveness under a demanding timeline made this study possible. We would also like to thank the Government of Pakistan, the Ministry of Sindh, and the Education and Literacy Department for their support of this activity. Finally, this effort would not have been possible without the dedication of our field teams of quality control officers and our local data collection partner, Voice Tel Tech (VTT).

CONTENTS

Executive Summary	1
Chapter 1: Introduction.....	8
Chapter 2: Design and Methodology	10
Chapter 3: Urdu Findings and Results.....	21
Chapter 4: Sindhi Findings and Results.....	39
Chapter 5: Conclusions and Recommendations	57
Annexes.....	62
Annex 1: Complete Item Statistics by Grade.....	63
Annex 2: Box Plots for Phonics and Reading-Rate Fluency Tasks.....	65
Annex 3: Example of Fluency Score Threshold Calculations.....	70
Annex 4: Distribution of Reading Fluency and Comprehension Scores using Fixed Intervals	73

Tables

Table 1: Timeline (January 2013 to May 2014)	12
Table 2: Urdu and Sindhi - Medium Sample Schools by District, Gender, and Location for Full Treatment	14
Table 3: Urdu and Sindhi - Medium Sample Schools by District, Gender, and Location for Light Treatment.....	15
Table 4: Urdu and Sindhi Reliability Estimates.....	18
Table 5: EGRA Score Ranges and Calculations	19
Table 6: Example of EGRA Percent Correct and Summary Scores	20
Table 7: Example of EGRA Timed Task Scores.....	20

Urdu Tables

Table 8u: Urdu Actual Student Sample by Grade and Gender	21
Table 9u: Urdu Tasks Statistics (Full and Light Treatment Groups)	22
Table 10u: Urdu Percent Correct Scores by Grade and Task (Full and Light Treatment Groups)	24
Table 11u: Urdu Scores by Grade, Task, and Group.....	24
Table 12u: Urdu Scores by Grade, Task, and Gender (Full and Light Treatment Groups).....	26
Table 13u: Urdu Percent Correct Scores by Group, Grade, Gender and Task.....	27
Table 14u: Urdu Baseline Maximum Scores on Fluency (timed) Tasks (Full and Light Treatment Groups)	28
Table 15u: Urdu Phonics and Reading-Rate Fluency Task Means by Grade (Full and Light Treatment Groups)	28
Table 16u: Urdu Phonics and Reading-Rate Fluency Task Means by Grade and Group (Full and Light Treatment Groups).....	29
Table 17u: Urdu Phonics and Reading-Rate Fluency Task Means by Grade and Gender (Full and Light Treatment Groups).....	29
Table 18u: Urdu Phonics and Reading-Rate Fluency Task Means by Group, Grade, and Gender.....	30
Table 19u: Urdu Summary Scores by Student Age	31
Table 20u: Urdu Summary Scores by Reading the Quran at Home.....	31
Table 21u: Urdu Summary Scores by the Presence of a Library at the School	31
Table 22u: Urdu Summary Scores by the Presence of Newspapers at Home	32
Table 23u: Urdu Summary Scores by the Presence of Magazines at Home	32
Table 24u: Urdu Summary Scores by the Presence of Books at Home.....	32
Table 25u: Urdu Summary Scores By Children Having Someone Read to Them at Home	33
Table 26u: Urdu Summary Scores by Children Reading to Someone Else at Home	33

Table 27u: Urdu Summary Scores by Children Reading Silently at Home.....	33
Table 28u: Urdu Summary Scores by Teacher Academic Qualification.....	34
Table 29u: Urdu Summary Scores by Teacher Professional Qualification.....	34
Table 30u: Urdu Summary Scores by Teacher Age.....	34
Table 31u: Urdu Summary Scores by Teacher Experience.....	35
Table 32u: Urdu Summary Scores by Teacher In-Service Training.....	35
Table 33u: Urdu Summary Scores by Head Teacher Academic Qualification.....	35
Table 34u: Urdu Summary Scores by Head Teacher Professional Qualification.....	36
Table 35u: Urdu Summary Scores by Head Teacher Experience.....	36
Table 36u: Urdu Summary Scores by Head Teacher In-Service Training.....	36
Table 37u: Urdu Summary Scores by Head Teacher Support to Teachers in Reading.....	37
Table 38u: Urdu Summary Scores by Head Teacher Training in Teaching Reading.....	37
Table 39u: Urdu Summary Scores by School Gender.....	37
Table 40u: Urdu Summary Scores by PTA/SMC/PTSMC/PTC.....	38
Table 41u: Urdu Summary Scores by Presence of a School Library.....	38
Table 42u: Urdu Summary Scores by Infrastructure (Drinking Water, Electricity, Toilets).....	38
Table A1u: Urdu Complete Item Statistics by Grade.....	63
Table A2u: Urdu Thresholds for CWPM with 80 Percent Comprehension.....	70
Table A3u: Urdu Thresholds for CWPM with Fixed Intervals.....	71
Table A4u: Grade 3 Reading Fluency and Comprehension, Urdu.....	73
Table A5u: Grade 5 Reading Fluency and Comprehension, Urdu.....	74

Sindhi Tables

Table 8s: Sindhi Actual Student Sample by Grade and Gender.....	39
Table 9s: Sindhi Task Statistics (Full and Light Treatment Groups).....	40
Table 10s: Sindhi Percent Correct Scores by Grade and Task (Full and Light Treatment Groups).....	42
Table 11s: Sindhi Scores by Grade, Task, and Group.....	42
Table 12s: Sindhi Scores by Grade, Task, and Gender (Full and Light Treatment Groups).....	44
Table 13s: Sindhi Percent Correct Scores by Group, Grade, Gender and Task.....	45
Table 14s: Baseline Maximum Scores on Fluency (Timed) Tasks (Full and Light Treatment Groups).....	46
Table 15s: Sindhi Phonics and Reading-Rate Fluency Task Means by Grade (Full and Light Treatment Groups).....	46
Table 16s: Sindhi Phonics and Reading-Rate Fluency Task Means by Grade and Group.....	47
Table 17s: Sindhi Phonics and Reading-Rate Fluency Task Means by Grade and Gender (Full and Light Treatment Groups).....	47
Table 18s: Sindhi Phonics and Reading-Rate Fluency Task Means by Group, Grade, and Gender.....	48
Table 19s: Sindhi Summary Scores by Student Age.....	49
Table 20s: Sindhi Summary Scores by Reading the Quran at Home.....	49
Table 21s: Sindhi Summary Scores by the Presence of a Library at the School.....	49
Table 22s: Sindhi Summary Scores by the Presence of Newspapers at Home.....	50
Table 23s: Sindhi Summary Scores by the Presence of Magazines at Home.....	50
Table 24s: Sindhi Summary Scores by the Presence of Books at Home.....	50
Table 25s: Sindhi Summary Scores by Children Having Someone Read to Them at Home.....	51
Table 26s: Sindhi Summary Scores by Children Reading to Someone Else at Home.....	51
Table 27s: Sindhi Summary Scores by Children Reading Silently at Home.....	51
Table 28s: Sindhi Summary Scores by Teacher Academic Qualification.....	52
Table 29s: Sindhi Summary Scores by Teacher Professional Qualification.....	52
Table 30s: Sindhi Summary Scores by Teacher Age.....	52
Table 31s: Sindhi Summary Scores by Teacher Experience.....	53
Table 32s: Sindhi Summary Scores by Teacher In-Service Training.....	53
Table 33s: Sindhi Summary Scores by Head Teacher Academic Qualification.....	54

Table 34s: Sindhi Summary Scores by Head Teacher Professional Qualification.....	54
Table 35s: Sindhi Summary Scores by Head Teacher Experience.....	54
Table 36s: Sindhi Summary Scores by Head Teacher In-Service Training	55
Table 37s: Sindhi Summary Scores by Head Teacher Support to Teachers in Reading	55
Table 38s: Sindhi Summary Scores by Head Teacher Training in Teaching Reading	55
Table 39s: Sindhi Summary Scores by School Gender.....	56
Table 40s: Sindhi Summary Scores by PTA/SMC/PTSMC/PTC.....	56
Table 41s: Sindhi Summary Scores by Presence of School Library.....	56
Table 42s: Sindhi Summary Scores by Infrastructure (Drinking Water, Electricity, Toilets)	56
Table A1s: Sindhi Complete Item Statistics by Grade.....	64
Table A2s: Sindhi Thresholds for CWPM with 80 Percent Comprehension	71
Table A3s: Sindhi Thresholds for CWPM with Fixed Intervals	72
Table A4s: Grade 3 Reading Fluency and Comprehension, Sindhi.....	75
Table A5s: Grade 5 Reading Fluency and Comprehension, Sindhi.....	76

Figures

Figure 1: Evaluation Design.....	10
Figure A1: Understanding Boxplots	65

Urdu Figures

Figure 2u: Urdu Grade 3 Summary Scores	23
Figure 3u: Urdu Grade 5 Summary Scores	23
Figure 4u: Urdu Full Treatment scores by Grade and Task (Full and Light Treatment Groups).....	25
Figure 5u: Urdu Light Treatment Scores by Grade and Task (Full and Light Treatment Groups).....	25
Figure 6u: Urdu Grade 3 Scores by Task and Gender (Full and Light Treatment Groups)	26
Figure 7u: Urdu Grade 5 Scores by Task and Gender (Full and Light Treatment Groups)	27
Figure A2u: Phonics and Reading-Rate Fluency Box Plots for Grade 3, Urdu	66
Figure A3u: Phonics and Reading-Rate Fluency Box Plots for Grade 5, Urdu	67
Figure A4u: Grade 3 Reading Fluency and Comprehension, Urdu.....	74
Figure A5u: Grade 5 Reading Fluency and Comprehension, Urdu.....	75

Sindhi Figures

Figure 2s: Sindhi Grade 3 Summary Scores.....	41
Figure 3s: Sindhi Grade 5 Summary Scores.....	41
Figure 4s: Sindhi Full Treatment Scores by grade and Task.....	43
Figure 5s: Sindhi Light Treatment Summary Scores by grade and Task	43
Figure 6s: Sindhi Grade 3 Scores by Task and Gender (Full and Light Treatment Groups).....	44
Figure 7s: Sindhi Grade 5 Scores by Task and Gender (Full and Light Treatment Groups).....	45
Figure A2s: Phonics and Reading-Rate Fluency Box Plots for Grade 3, Sindhi.....	68
Figure A3s: Phonics and Reading-Rate Fluency Box Plots for Grade 5, Sindhi.....	69
Figure A4s: Grade 3 Reading Fluency and Comprehension, Sindhi	76
Figure A5s: Grade 5 Reading Fluency and Comprehension, Sindhi	77

ACRONYMS

AJK	Azad Jammu and Kashmir
B.A.	Bachelor of Arts
BEFARe	Basic Education for Awareness, Reforms, and Empowerment
B.Sc.	Bachelor of Science
C.T.	Certificate of Teaching (Grade 12 plus FA/FSC Certificate)
DOE	Education and Literacy Department
EGRA	Early Grade Reading Assessment
F.A.	Fellow in Arts
FATA	Federally Administered Tribal Areas
F.Sc.	Fellow in Sciences
GB	Gilgit-Baltistan
ICT	Islamabad Capital Territory
KP	Khyber Pakhtunkhwa
M.A.	Master of Arts
Matric	Secondary School (Grade 10) Certificate (Matriculation)
M.Ed.	Master of Education
M.Sc.	Master of Science
MSI	Management Systems International
MT	Master Trainers
NEAS	National Education Assessment System
NEMIS	National Education Management Information System
PRP	Pakistan Reading Project
P.T.C.	Primary Teaching (Grade 12) Certificate
QCO	Quality Control Officer
SEMIS	Sindh Education Management Information System
SPSS	Statistical Package for the Social Sciences
SRP	Sindh Reading Project
STS	School-to-School International
USAID	United States Agency for International Development
VTT	Voice Tel Tech

EXECUTIVE SUMMARY

Overview

In 2013, Management Systems International (MSI) and School-to-School International (STS) conducted a baseline reading assessment for primary school children prior to the launching of two USAID-funded projects: the Pakistan Reading Project (PRP) and the Sindh Reading Program (SRP). PRP is targeting improved reading for 910,000 children in Azad Jammu and Kashmir (AJK), Balochistan, the Federally Administered Tribal Areas (FATA), Gilgit-Baltistan (GB), the Islamabad Capital Territory (ICT), Khyber Pakhtunkhwa (KP), and Sindh, while the SRP is targeting improved reading and mathematics for 750,000 children in Sindh. Targets will be achieved through support for 1) improved policies, laws, and guidelines for teachers and administrators, and 2) improved reading instruction for children in the primary grades.

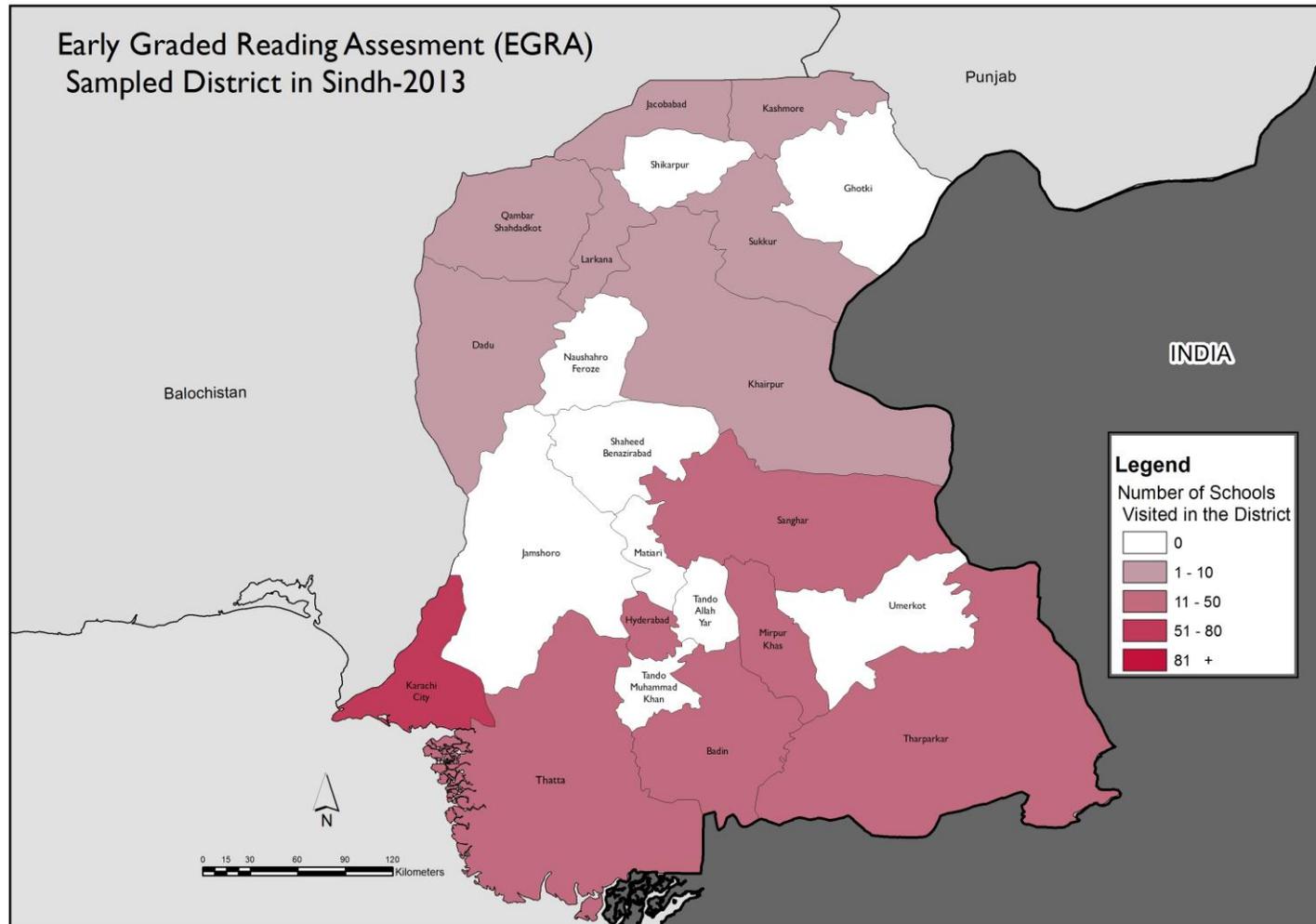
To measure results from PRP and SRP, a rigorous external evaluation is being conducted. This report covers the baseline assessment in Sindh which took place in September 2013. In May 2013, GB, AJK, and ICT were part of Round 1 of the baseline data collection, while data collection in KP, along with Sindh, was completed in September 2013 and data collection in Punjab, Balochistan, and FATA was completed in October 2013. The following activities were carried out for all of the provinces, including Sindh: 1) design, 2) sampling, 3) instrumentation, 4) planning, 5) training, 6) implementation, 7) analysis, and 8) reporting.

The external evaluation design, which was developed prior to the baseline assessment, was tailored to the implementation of the PRP and SRP in each province. In most of the provinces, a quasi-experimental design will be used, with two treatment groups: full treatment and light treatment. The full treatment group will receive both the first and second kinds of support, i.e., 1) policy, laws, and guidelines, and 2) improved instruction. The light treatment group will only receive the first kind of support.

In accordance with the USAID evaluation guidelines, students at two selected grade levels – grades 3 and 5 – will be assessed at three time points: baseline, midline, and endline. An internationally accepted assessment tool, the Early Grade Reading Assessment (EGRA), will be individually administered to over 30,000 children in over 1,000 schools throughout the country. Over the course of the two projects, the evaluators will compare the baseline results with those at the midline and endline to examine success in improving children's reading levels in Pakistan. The sampling was designed so that each province could be evaluated independently.

For the baseline in Sindh, all activities were completed by the end of November 2013, with the draft report submitted in early December. The EGRA baseline results were presented and discussed at a consultative meeting in Karachi in April 2014. Representatives from the provincial Education and Literacy Department (DOE), USAID, and the contractors (MSI and STS) attended the consultation. Revisions were then be made to this report based on the discussions between the stakeholders.

Map of Sampled Districts



Key Points

Several key points from the EGRA baseline assessment in Sindh are highlighted below:

Implementation

1. Sindh schools teaching in Urdu and Sindhi are being evaluated separately. Both evaluations involve two kinds of comparisons: 1) a comparison of full and light treatment groups to determine the effects of full treatment above and beyond that of the light treatment, and 2) a comparison of each group to itself at the baseline, midline, and endline. (Please see Figure 1 and the accompanying text for a fuller description of the evaluation design.)
2. District selection was finalized following consultative meetings between the DOE and USAID in August 2013. For the schools teaching in the Urdu language, USAID and the DOE under the SRP chose selected Karachi City towns for full treatment. Hyderabad, Mirpur Khas, and Sanghar were randomly selected for light treatment for the EGRA evaluation. For Sindhi-medium schools, the full treatment schools were in the following districts: Dadu, Jacobabad, Kambar-Shahdadt, Karachi City, Kashmore, Khairpur Mirs, Larkana, and Sukkur. These seven districts and selected Karachi City towns were chosen by USAID and the DOE for SRP interventions. Three districts, Badin, Tharparkar, and Thatta, were randomly selected for light treatment for the EGRA evaluation.
3. EGRA in Urdu and Sindhi was used in the Sindh province. The EGRA tools, which have been administered in various forms in over 40 countries, were successfully adapted for use in Pakistan. These included individually administered reading tests for students, along with questionnaires for students, teachers, and head teachers. The Urdu version of the tools was piloted in AJK, ICT, and KP. The Sindhi version of the tools was piloted in Sindh.
4. For each language, a total of 140 schools, with 70 schools from each group (full and light treatment), were selected for the baseline.
5. The Urdu baseline data were collected in schools in three districts and in three towns of Karachi City. The Sindhi baseline data were collected in 10 districts and in three towns of Karachi City.
6. For each language, a random sample of male and female schools was selected within the districts and towns, followed by a random sample of grades 3 and 5 students within those schools. The results from this sample are presented in this report as a generalized view of the reading levels for students in Urdu- and Sindhi-medium schools.
7. The results from this sample are presented in this report as a generalized view of the reading levels for students in the Sindh schools. Please note that district comparisons are not possible because the districts were not evenly sampled; the number of sampled schools varied by district, and the sample sizes are limited for each district.
8. The EGRA testing window was September 2013 for both languages. All schools were tested within this period of time.
9. The Urdu-medium assessment tools were successfully administered in (with a percentage of the target reached in parentheses) 140 schools (100.0 percent) to 3,881 students (92.4 percent), 262 teachers (93.6 percent – with some teachers covering both grades 3 and 5, thus reducing the total number of teachers in the survey) and 140 head teachers (100.0 percent). The Sindhi-medium assessment tools were successfully administered in 140 schools (100.0 percent) to 4,076 students (97.0 percent), 236 teachers (84.3 percent), and 140 head teachers (100.0 percent).
10. The validity and reliability of the tools for both languages was very good. Validity was assured through the adaptation process, which involved 17 educationists from throughout the country who participated in a workshop in Islamabad. Reliability was assured through the high quality of

the assessment tools and the standardized administration of the tools in Sindh. Reliability estimates (of internal consistency) were calculated using the coefficient alpha.

11. The data entry and data cleaning process followed international standards. All student data were entered twice into two separate databases. All data were reconciled across the two databases and with the assessment booklets. A clean data file was produced for analysis.

Results

Urdu- and Sindhi-medium schools were analyzed separately. In the analysis phase, scores were calculated in three ways: 1) percentage correct scores for the reading tasks, 2) average percentage correct (grand means) for reading summary scores, and 3) adjusted raw scores for the timed reading tasks. These scores provide a comprehensive picture of student performance. Analysis of student, teacher, head teacher, and school characteristics was carried out using the summary scores.

Urdu-Medium Schools

Several key findings emerged from the Urdu baseline assessment in Sindh. These are as follows:

1. EGRA was administered to 1,966 grade 3 students and 1,915 grade 5 students. The reliability estimates were high for both grades (alpha = 0.88 for grade 3 and 0.87 for grade 5), indicating that the items worked well in measuring reading constructs at each grade level.
2. The task and item statistics showed that the EGRA discriminates well between low- and high-achieving students in both grades. The task p-values for grade 3 provided a spread on the lower to middle section of the difficulty range, while p-values for grade 5 were higher and covered the middle parts of the spectrum. All task scores at grades 3 and 5 had item-total correlations equal to or greater than 0.25, indicating good discrimination quality for these tasks. (Complete item statistics are listed in Annex 1.)
3. Grade 3 students did relatively better on the orientation to print, letter name recognition, and passage reading. The lowest scores were in phonics (non-word reading and letter sound knowledge) and comprehension (passage and listening). The best scores for grade 5 students were in the two reading tasks (familiar word and passage). They still had relatively low scores in phonics (non-word reading, letter sound knowledge) and comprehension (passage and listening).
4. There was also substantial progression from grade 3 to grade 5 on the summary score (17 points) and on some of the tasks – especially in familiar word, and passage reading and comprehension.
5. There were differences between boys and girls on the task and summary scores, but most of these differences were small. For both grades there was about an 11-point difference on the summary score favoring the girls. Girls had higher scores on familiar word reading, non-word reading, and passage reading and comprehension. In addition, girls also had higher fluency scores on the timed reading tasks.
6. Summary scores for students from full treatment schools were about six points higher in grade 3 and nine points higher in grade 5 than in light treatment schools. The full and light treatment groups had similar scores for most of the tasks, though the full treatment scores were higher on some tasks. The largest differences were in the phonics areas of letter sound knowledge and listening comprehension. These differences will be corrected statistically during the midline and endline assessments.
7. Students were timed on five tasks as they read words or passages. These tasks were categorized into phonics (letter name recognition, letter sound knowledge, and non-word reading) and reading-rate fluency (familiar word and passage reading). Passage reading was approximately 30 points higher in grade 5 (69 points) than in grade 3 (39 points). Although the passage was designed for grade 3, this difference shows that the reading levels in grade 3 are low but that students can make substantial progress in the early grades if expectations are high enough and if

they are provided with the opportunity to learn. Specifically, mastery of phonics, such as letter sound knowledge, phonemic awareness, and non-word reading, should help the students become better overall readers. It is clear that these types of knowledge and skills are not receiving an appropriate emphasis in Urdu-medium schools in Sindh.

8. Student questionnaire findings revealed two interesting findings. The first positive finding was that having reading materials and opportunities to read in the home seemed to have a positive effect on reading outcomes for both grade 3 and grade 5 students. Second, grade 3 summary scores increased with relative age (younger-than-normal, normal, older-than-normal age); older students in the grade had higher reading scores. However, by grade 5 that advantage was no longer significant.
9. School, teacher, and head teacher questionnaire findings were mostly inconclusive, due to small sample sizes and the lack of variation in the scores that were related to their characteristics. For teachers, those who attended one or more in-service trainings had higher scores than those who never attended such trainings. For head teachers, attending one or two in-service trainings and in-service training in teacher reading tended to relate to higher reading scores for students. For the schools, the presence of a library, and better infrastructure were associated with better student reading scores.

Sindhi-Medium Schools

Several key findings emerged from the Sindhi baseline assessment in Sindh. These are as follows:

1. EGRA was administered to 2,038 grade 3 students and 2,038 grade 5 students. The reliability estimates were high for both grades ($\alpha = 0.88$ for grade 3 and 0.88 for grade 5), indicating that the items worked well in measuring reading constructs at each grade level.
2. The task and item statistics showed that the EGRA discriminates well between low- and high-achieving students in both grades. The task p-values for grade 3 provided a spread on the lower to middle section of the difficulty range, while p-values for grade 5 were higher and covered the middle parts of the spectrum. All task scores at grades 3 and 5 had item-total correlations equal to or greater than 0.30, indicating good discrimination quality for these tasks. (Complete item statistics are listed in Annex 1.)
3. Grades 3 and 5 showed the same patterns on the reading tasks. Students did relatively better on the orientation to print, listening comprehension, familiar word and passage reading. The lowest scores were in phonics (non-word reading and letter sound knowledge) and reading comprehension.
4. There was also substantial progression from grade 3 to grade 5 on the summary score (15 points) and on some of the task scores – especially in fluency (familiar word reading, passage reading, non-word reading) and passage comprehension.
5. There were differences between boys and girls on the task and summary scores. For all of the task scores and the summary score, boys had statistically higher ($p < 0.01$) scores than girls. The summary score difference was 7 and 12 points, respectively for grades 3 and 5. The tasks with the largest differences were in familiar word reading and passage reading.
6. Summary scores for students from full treatment schools were about two points higher in grade 3 and five points higher in grade 5. There were no consistent differences among the reading tasks between the full and light treatment groups. The light treatment group had higher scores on some tasks while the full treatment group had higher scores on others. Most of these differences were slight, with the exception of letter name recognition and listening comprehension favoring the full treatment group. These differences will be corrected statistically at the midline and endline.
7. Passage reading (fluency) was approximately 18 points higher in grade 5 than in grade 3. However, no large discrepancies were found between the full and light treatment groups.

Although the passage was designed for grade 3, this difference shows that the reading levels in grade 3 are low, but that children can make substantial progress in the early grades if expectations are high enough and if they are provided with the opportunity to learn. Specifically, mastery of phonics, such as letter sound knowledge, phonemic awareness, and non-word reading, should help the students become better overall readers. It is clear that these types of knowledge and skills are not receiving an appropriate emphasis in Sindhi-medium schools.

8. Student questionnaire findings revealed two interesting findings. The first positive finding was that having reading materials and opportunities to read in the home seemed to have a positive effect on reading outcomes for both grade 3 and grade 5. Second, grade 3 summary scores increased with relative age (younger than normal, normal, older than normal age); older children in the grade had higher reading scores. However, by grade 5 that advantage was no longer significant.
9. School, teacher, and head teacher questionnaire findings were mostly inconclusive, due to small sample sizes and the lack of variation in the scores that were related to their characteristics. For teachers and head teachers, reading scores were not significantly higher for those who attended in-service trainings. For the schools, better reading scores were associated with the presence of a library (grade 3 only) and better infrastructure (grade 5 only).

Evaluation Recommendations

Given the success of the baseline assessment in Sindh (and in the other provinces), the methods used in 2013 should be repeated as much as possible for the midline and endline assessments in future years. This should be conducted as follows:

1. The EGRA instruments proved to be of high quality, and equivalent versions of those tools should be developed – through trans-adaptation, piloting, and revision – for the midline and endline assessments so that progress can be accurately measured over time.
2. The EGRA items and tasks had good discrimination (quality) values and covered the low-to-middle part of the difficulty range. At baseline, the reading scores were relatively low for both grades and show room for growth. In addition, histograms and box plots provided evidence that the tool is expected to measure higher levels of reading-rate fluency that are anticipated following project-led interventions. Therefore, the baseline data indicates that the EGRA is appropriate for measuring increases in reading ability at midline and endline.
3. The sampling was reasonable in terms of finding a balance between the resources available, the required sample size, and the geographic coverage. It should be maintained in the midline and endline, i.e., keep the same districts and schools, along with the sampling methods at the school level.
4. The systems for field data collection should be replicated, with the same systems for recruitment and training for the master trainers (MTs), field supervisors, quality control officers (QCOs), and enumerators as used in the baseline.
5. The data entry system should continue to be used, with the same systems for recruitment and training of data entry supervisors and operators, along with implementation through networked computers, double data entry, and reconciliation of errors.
6. The analysis should follow the same procedures, with calculations of task scores, summary scores, and timed task scores. The baseline, midline, and endline scores should be comparable so that improvements in students' reading can be accurately examined.
7. Reading proficiency levels should be created to provide educators and other stakeholders with meaningful results. Most parents and educators better understand reading achievement in useful terms or levels, such as emerging, proficient, or advanced, rather than interpreting a percent-correct test score that may differ by test or reading passage difficulty. Education officials are

encouraged to select specific EGRA scores to serve as levels of reading proficiency for both grades. Percent correct for each task, summary score, as well as fluency rates are recommended for this purpose. The baseline EGRA data can be used for establishing these reading proficiency levels.

8. Finally, it may be advisable to add items to the student, teacher, and head teacher questionnaires for collecting data on SRP- and PRP-supported interventions so that student scores can be correlated with these indicators.

CHAPTER I: INTRODUCTION

The Pakistan Reading Project (PRP) and the Sindh Reading Program (SRP) are two five-year initiatives funded by USAID. The projects/programs will cover over 40,000 government schools in Pakistan's eight provinces/areas/territories (hereafter referred to as provinces). PRP is targeting improved reading for 910,000 children in AJK, Balochistan, FATA, GB, ICT, KP, and Sindh, while the SRP is targeting improved reading and mathematics for 750,000 children in Sindh. Targets will be achieved through support for 1) improved policies, laws, and guidelines for teachers and educational administrators, and 2) improved reading instruction for children in primary grades. Some districts in Pakistan will receive both kinds of support, i.e., "full treatment," while others will receive only the policy support, i.e., "light treatment." All schools within districts will receive the same type of treatment.

To measure results from PRP and SRP, a rigorous external evaluation is being conducted. The evaluation baseline is taking place in 2013, prior to the launch of the reading interventions. In accordance with USAID program evaluation guidelines, samples of students in two selected grade levels – grade 3 and grade 5 – are being assessed throughout Pakistan so that independent baselines can be established in each province. Students at the same grade levels will be assessed at the midline and endline time points to evaluate the success of the interventions, taking into account the two treatment groups.

This report covers Sindh Province. Along with KP, Sindh was part the baseline data collection in September 2013; data from Pakistan's other six provinces were collected in May 2013 (ICT, AJK, GB) and October 2013 (Punjab, Balochistan, and FATA). The following activities were planned for all of the provinces, including Sindh:

1. Design – USAID required a cross-sectional design, i.e., assessing students at the same grade levels (grades 3 and 5) over the course of PRP and SRP. In most provinces, including Sindh, this was complemented by a quasi-experimental design with the two treatment groups (full and light). In Sindh the design also included assessments in two languages, Urdu and Sindhi, each of which had its own sample.
2. Sampling – Schools were selected from the full and light treatment districts. The sample enabled the collection of student reading assessment data that were representative of the treatment groups, grade levels, and gender. There was also some stratification by urban/rural zones, though at times, achieving a balance for this variable was difficult due to towns and districts within a group being primarily urban or primarily rural.
3. Instrumentation – EGRA tools were developed, with tests at the grade 3 level in English, Sindhi, and Urdu, and questionnaires for teachers, head teachers, and students in Urdu and Sindhi. Model EGRA instruments were trans-adapted, piloted, revised, and finalized for use in Pakistan. The Urdu and Sindhi tools were used in Sindh.
4. Planning – A field administration plan was developed for the baseline administration that would ensure the reliability of the data collected. The plan specified the timeline, training, logistics, field activities, supervision, data entry, analysis, reporting, and quality control.
5. Training – Workshops were conducted to train all MTs, supervisors, enumerators, and QCOs. Enumerators and supervisors were observed to ensure clear comprehension and skills adequate to implement the EGRA tools.
6. Implementation – The baseline survey was implemented according to the plan. It ensured that all of the field activities took place in a standardized manner, as verified by the QCOs. The fieldwork was followed by data entry and preparation of a clean data file.
7. Analysis – Data were analyzed using spreadsheet (Excel) and statistical (Statistical Package for the Social Sciences [SPSS]) software. Experienced statisticians/psychometricians conducted the analysis, produced data tables and graphs, and ensured quality control.

8. Reporting – Provincial-level reports were produced and will be disseminated to the provincial education authorities. A template was developed according to guidelines from the USAID contract.

This report is organized into five chapters: 1) introduction, 2) methodology, 3) Urdu findings and results, 4) Sindhi findings and results, and 5) conclusions and recommendations. Annexes with item statistics, box plots for the timed tasks, and a possible process for establishing a reading proficiency threshold follow the chapters.

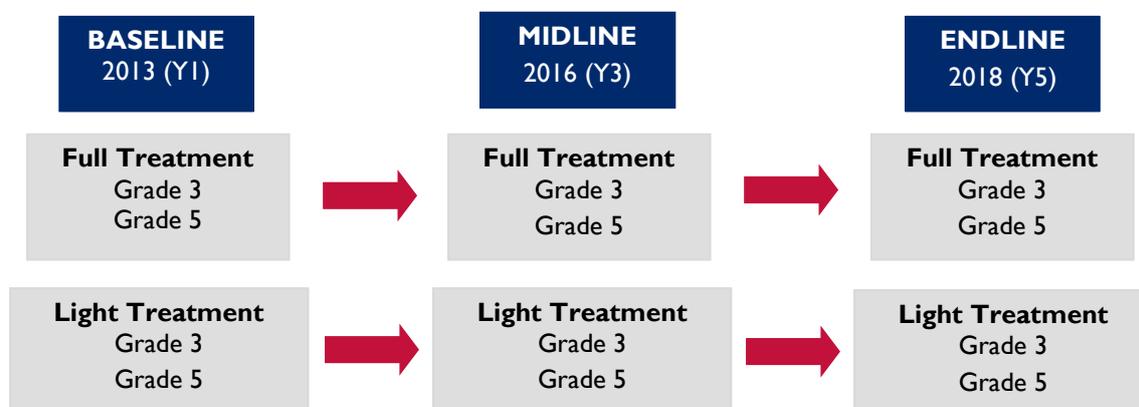
CHAPTER 2: DESIGN AND METHODOLOGY

This chapter presents the evaluation design and methodology, including the systems used for collecting the EGRA baseline data for both Urdu and Sindhi schools in Sindh. There are sections on the evaluation design, timeline, sampling, instrument development, data collection, data entry, and data analysis.

Evaluation Design

Following USAID policy, a cross-sectional evaluation design was developed prior to the baseline data collection. As shown in Figure 1, the design features two grade levels (3 and 5) and three time points (baseline, midline, and endline). Different groups of grade 3 and grade 5 students will be compared against each other across the three time points. In Figure 1, the years for the midline and endline are approximate and may be altered in accordance with implementation of the PRP and SRP interventions.

FIGURE 1: EVALUATION DESIGN



The same evaluation design was followed for both the Urdu- and Sindhi-medium schools. Districts for the full treatment group were pre-selected by the provincial DOE and USAID in Sindh. Districts for the light treatment group in Sindh were selected randomly prior to the implementation of the baseline. Since district-level selection for the two groups was not random, equivalence at baseline of the two treatment groups cannot be assured, and a quasi-experimental design was selected. In this design, any differences in scores at baseline (and midline and endline) will be statistically removed in the analysis, i.e., the two groups will be made statistically equivalent even though their average scores may be different. This will ensure fairness in the comparison of the full and light treatment groups.

For the full treatment schools in Sindh, the DOE and USAID chose the seven districts of Dadu, Jacobabad, Kambar-Shahdadkot, Kashmore, Khairpur Mir, Larkana, Sukkur, and the five towns in Karachi City (Kemari, Gadap, Bin Qasim, Orangi, and Leyari) for interventions in both Urdu- and Sindhi-medium schools. For Urdu-language, only Karachi City had enough full treatment schools for the full treatment sampling. For Sindhi-language, all full treatment districts and towns were used for the full treatment sampling. For both languages, the light treatment districts were selected randomly from the remaining districts of Sindh. Hyderabad, Mirpur Khas, and Sanghar were randomly selected as light treatment for the Urdu language baseline, and Badin, Tharparkar, and Thatta were randomly selected as light treatment districts for the Sindhi language baseline.

Sindh students were tested in either Urdu or Sindhi, depending on their main language of instruction. The same number of Urdu and Sindhi schools participated in the baseline study. For each language, equal numbers of male and female schools, i.e., 35 male and 35 female schools per treatment group, were

sampled for the EGRA testing. The sampling design met the USAID requirements of adequate sample size and equal gender representation (see the sampling section below).

Timeline

The Sindh baseline, like the other provinces, was conducted according to a timeline that started in January and ended in November 2013, with a first draft of reports completed by December. The reports may then be distributed to the DOE and other stakeholders as appropriate. (See Table 1 below.)

The process began in January with activity planning and design, including the creation of preliminary sampling designs, selection of model EGRA tasks, recruitment of staff, and budgeting/contracting. From February to August, the EGRA team, with participation from Sindh and other provinces, then prepared, piloted, and revised the EGRA tools and conducted the district/school sampling. The data collection in Sindh took place in September, followed by the data entry, analysis, and reporting in October and November. Presentations to Sindh and USAID were conducted in April 2014. This final report for Sindh was then submitted in May 2014.

TABLE I: TIMELINE (JANUARY 2013 TO MAY 2014)

Activity	2013												2014				
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Plan and design EGRA activities	X	X															
Debrief to USAID and the MOE	X	X	X			X	X										
Prepare EGRA tools		X	X														
Prepare test administration manuals			X														
Train master trainers and enumerators								X									
Select and verify sample schools							X	X									
Administer EGRA									X								
Enter data										X							
Analyze baseline data										X	X						
Produce draft reports												X	X				
Produce presentations														X	X		
Disseminate draft reports														X			
Make presentations																X	
Revise and finalize reports																X	X
Submit final reports to USAID																	X

Sampling

The sampling for Sindh was finalized in August 2013 following meetings with USAID. The EGRA team conducted the school sampling in July and August. This included developing the sampling requirements, verifying the sample in the field, and finalizing the sample. As mentioned in the above section on evaluation design, the sampling frame for the Urdu- and Sindhi-medium full treatment schools was comprised of seven full treatment districts and five full treatment towns in Karachi City. The sampling frame for light treatment was comprised of all other districts (n = 19), and three districts were selected at random for each of the languages. The sampling for Sindh, as detailed in the sampling report for USAID,¹ is briefly summarized in the following sub-sections of this report.

Sampling Requirements

Since the minimum requirement was 15 students per grade level in grades 3 and 5, only schools meeting that requirement were eligible for sampling. Within the treatment groups (full and light), as well as within each language group (Sindhi and Urdu), equal numbers of male and female schools (35 each) were selected.

Sampling Process and Field Verification

Due to requirements of SRP, all intervention districts (7) and intervention Karachi City towns (5) were selected for the full treatment group in Sindhi, and all intervention Karachi City towns (5) were selected for the full treatment group in Urdu. Three light treatment districts were randomly selected for each of the Sindhi and Urdu assessments. This resulted in a clustered sample. For the 35 male and 35 female schools in the light treatment group, the samples were divided among the selected districts according to the proportions of schools within those districts (stratified random sampling). For the full treatment group, an equal number of schools were selected in each of the districts. As much as possible, equal numbers of male and female schools were chosen within each of the districts for the full treatment group. For both groups, a second stratification was done at the “location” level, where schools were allocated by rural and urban. Table 2 shows the number of schools for the full treatment groups in Urdu and Sindhi per gender and location. Table 3 shows the total number of urban and rural schools in each district and the percent of these schools that were sampled for the light treatment groups.

¹ MSI (2013). *Pakistan EGRA Sampling Report*.

TABLE 2: URDU AND SINDHI - MEDIUM SAMPLE SCHOOLS BY DISTRICT, GENDER, AND LOCATION FOR FULL TREATMENT

District	Treatment	Location	Sample Schools		Replacement Schools	
			Boys	Girls	Boys	Girls
Karachi City	Full	Urban	26	32	5	10
Karachi City	Full	Rural	9	3	5	0
Total Urdu Full Treatment			35	35	10	10
Dadu	Full	Urban	1	2	2	0
Dadu	Full	Rural	4	2	0	2
Jacobabad	Full	Urban	1	7	2	0
Jacobabad	Full	Rural	0	1	0	2
Kambar-Shahdadkot	Full	Urban	0	2	1	1
Kambar-Shahdadkot	Full	Rural	4	3	1	0
Karachi City	Full	Urban	0	1	0	0
Karachi City	Full	Rural	5	1	0	0
Kashmore	Full	Urban	0	0	1	0
Kashmore	Full	Rural	5	4	1	0
Khairpur Mirs	Full	Urban	0	0	1	1
Khairpur Mirs	Full	Rural	5	4	0	0
Larkana	Full	Urban	0	0	0	2
Larkana	Full	Rural	5	4	1	0
Sukkur	Full	Urban	2	2	1	0
Sukkur	Full	Rural	3	2	1	2
Total Sindhi Urban	Full	Urban	4	14	8	4
Total Sindhi Rural	Full	Rural	31	21	3	7
Total Sindhi Full Treatment			35	35	11	11

TABLE 3: URDU AND SINDHI - MEDIUM SAMPLE SCHOOLS BY DISTRICT, GENDER, AND LOCATION FOR LIGHT TREATMENT

District	Treatment	Location	Schools	Pct.	Sample Schools		Replacement Schools	
					Boys	Girls	Boys	Girls
Hyderabad	Light	Urban	201	29	10	10	3	3
Mirpur Khas	Light	Urban	58	9	3	3	1	1
Mirpur Khas	Light	Rural	167	24	9	9	2	2
Sanghar	Light	Urban	53	8	3	3	1	1
Sanghar	Light	Rural	201	29	10	10	3	3
Total Urdu Urban	Light	Urban	312	46	16	16	5	5
Total Urdu Rural	Light	Rural	368	54	19	19	5	5
Total Urdu Light Treatment			680	100	35	35	10	10
Badin	Light	Urban	119	1	0	0	0	0
Badin	Light	Rural	2,885	27	10	10	3	3
Tharparkar	Light	Urban	26	0	0	0	0	0
Tharparkar	Light	Rural	4,212	40	14	14	4	4
Thatta	Light	Urban	128	1	0	0	0	0
Thatta	Light	Rural	3,276	31	11	11	3	3
Total Sindhi Urban	Light	Urban	273	3	0	0	0	0
Total Sindhi Rural	Light	Rural	10,373	97	35	35	10	10
Total Sindhi Light Treatment			10,646	100	35	35	10	10

Once the schools were sampled, the QCOs, supplemented by EGRA senior managers, verified the samples in the field. This step was necessary due to two factors: 1) some inaccuracies in the National Education Management Information System/Sindh Education Management Information System (NEMIS/SEMIS) data, and 2) changes in student numbers since the time period when the schools had submitted their data to NEMIS/SEMIS. If the original schools had fewer than 15 students in either grade 3 or 5, a replacement school was selected and verified. At times, schools were retained if their student numbers were near the minimum. Note that mixed schools may have been selected for some replacement schools due to not having enough options for replacement schools of strictly one gender. However, only students from the respective genders were included in those samples (i.e. if a mixed school was selected to replace a female school, only females were sampled).

Intended and Actual Samples

For the Urdu schools, 26 schools – 13 boys and 13 girls schools – were replaced due to lower than expected numbers of children in the original samples. Similarly, for the Sindhi schools, 37 schools – 18 boys and 19 girls schools – were replaced. The actual numbers of students, teachers, and head teachers in the survey are presented in the results section.

Instrument Development

A brief summary of the instrument development process is presented below. The full results from the trans-adaptation, which involved educationists from Sindh, were presented in a report to USAID.² This report is available to provincial education officials.

Trans-adaptation

In February, the EGRA team used tasks from recent EGRA administrations in other countries to develop a model test. Led by two international and two national assessment specialists, the EGRA team then organized a trans-adaptation workshop in Islamabad. A total of 17 English, Sindhi, and Urdu language specialists from the DOEs and teacher training institutes throughout Pakistan – including two subject specialists from Sindh – participated in the workshop.

The trans-adaptation process involved the following, with the local experts:

1. Discuss and choose reading tasks that would be of value to the baseline assessment in Pakistan;
2. Adapt each reading task using appropriate content in English, Urdu, and Sindhi; and
3. Ensure that the content would be suitable for grades 3 and 5 students.

The workshop resulted in a pilot EGRA test and pilot student, teacher, and head teacher questionnaires. The head teacher questionnaires included items about school characteristics.

Piloting

In March 2013, the EGRA English and Urdu tools were piloted in selected schools in AJK, ICT, and KP provinces, while the Sindhi tools were piloted in June in Sindh. Four tools were included in the pilot: 1) a student response booklet (including the student questionnaire), 2) a student stimuli booklet, 3) a teacher questionnaire, and 4) a head teacher questionnaire. The EGRA team conducted the pilot sampling, trained the enumerators, arranged the logistics, and supervised the piloting. The team then entered the pilot data into a database, analyzed the data, and developed preliminary recommendations for final tools in preparation for the revision workshop. They also prepared a piloting report for USAID.³

Revision and Finalization

The EGRA team held a revision workshop in March for the Urdu and English tools with a limited number of experts from the trans-adaptation workshop. The Sindhi tools were revised in July with Sindhi language experts. Changes were made to the instruments based on the pilot data and field observations. These changes were summarized in the piloting report. The team then finalized the four instruments for each language and submitted them to USAID. USAID made suggestions, particularly around the inclusion of reading- and library-related items into the questionnaires that would provide information for the PRP and SRP. The English and Urdu instruments were approved and then used in the training workshops in advance of Round 1 data collection in May, while the Sindhi tools were approved and then used in the training workshops in advance of Round 2 data collection in September. The baseline final instruments were comprised of the following:

- Students: 16 informational items, 8 tasks (one with 2 sub-tasks), and 34 questionnaire items
- Teachers: 15 informational items and 52 questionnaire items
- Head teachers: 17 informational items and 37 questionnaire items

These instruments are available for use by education officials.

² MSI (2013) *Pakistan EGRA Tools Trans-Adaptation Workshop Report*. June (Revised).

³ MSI (2013). *Pakistan EGRA Instrument Development and Pilot Data Analysis*.

Data Collection

Subcontractor Selection

The EGRA team, with the participation of USAID, issued a request for proposals and followed a set of criteria to select local subcontractors for the field data collection and data entry. In August, Voice Tel Tech (VTT) was chosen for data collection in Sindh, while Basic Education for Awareness, Reforms, and Empowerment (BEFARe) was selected for data entry. A joint team from MSI, STS, and VTT collaborated on the data collection in Sindh.

Data Collection

In August, EGRA senior managers trained MTs and QCOs during a two-week session in Islamabad. The MTs then spent one week in Karachi training the VTT data collection team, which was comprised of two regional coordinators (one per language), 11 field supervisors, and 128 enumerators (64 per language). The QCOs, coordinators, supervisors, and enumerators organized the logistics for the data collection. Following the training and logistical preparations, the QCOs and field supervisors conducted a four-day refresher course for the enumerators in Karachi just prior to commencing data collection in the schools.

Over a 10-day period in September, the enumerators spent a day in each of the 280 schools to collect the baseline data in Sindh. The enumerators were in regular communication with the EGRA senior manager, QCOs, coordinator, and field supervisors to check on the status of data collection and to troubleshoot any issues. After collecting the data from the schools, the enumerators submitted their booklets to the supervisors and QCOs for verification and feedback. At the end of data collection, all booklets were returned to Islamabad for data entry.

Data Entry

Data Entry

In May 2013, the EGRA team developed a customized data entry application so that 1) the exact data from the booklets and questionnaires could be entered into a database, and 2) the computers used for data entry could be networked with a server. The application was updated in August using lessons learned from the first round of data entry. In September, the team trained the BEFARe data coordinator, four supervisors, and 36 data entry operators. In October and November (Rounds 2 and 3), the EGRA and BEFARe teams completed data entry for over 21,000 student booklets, along with the questionnaires for the students, teachers, and head teachers. This included about 8,000 student booklets and over 600 questionnaires for Sindh.

Data Cleaning

In October and November, the EGRA and BEFARe teams conducted the data verification and reconciliation. Following USAID requirements, 100 percent of the data were entered twice (double data entry) and any discrepancies between the first and second databases were reconciled. A clean data file was then provided to the data analysis team.

Data Analysis

Methodology

In June, the EGRA statisticians and psychometrician from STS developed a research plan that included the following steps: 1) reliability estimates, 2) task and item statistics, 3) mean and grand mean scores (percent correct scores), 4) data plots, 5) timed and untimed task scores, and 6) questionnaire results. They used SPSS for the analysis. Some of the analyses were replicated to ensure that the calculations were accurate. Descriptive analyses and inferential statistical comparisons were conducted by grade level, gender, and for the three sets of questionnaire data.

Please note that the analyses were only performed at the provincial level. This is because the sampling was conducted at the provincial level, i.e., the sample is only accurate at the provincial level. The samples at the district or school level are too small for analysis purposes, and any results at those levels would be misleading.

Validity and Reliability

Validity evidence for the tests was derived from previous experiences with EGRA in other developing countries, as well as through the trans-adaptation process in Pakistan. The test developers targeted grade 3 for the level of the tasks. An assumption was that the grade 5 students should perform better than the grade 3 students on each of the tasks. The Urdu and Sindhi tests were analyzed separately because of the differences in these language structures. As mentioned earlier, there were separate samples for Urdu- and Sindhi-medium schools.

For reliability, a generally accepted method is to estimate the internal consistency reliability (Coefficient Alpha) of the test. The minimum reliability threshold is approximately 0.75 to 0.80 for tests of this nature. Reliability was estimated for each province and language. Table 4 shows the reliability estimates for grades 3 and 5 in Sindh for the Urdu and Sindhi tests, respectively. These reliabilities are excellent and lend credibility for the tests' internal consistency, indicating that the items are measuring generally the similar reading constructs for both grade levels and both languages.

TABLE 4: URDU AND SINDHI RELIABILITY ESTIMATES

Language	Grade Level	Tasks	N-count	Alpha
Urdu	Grade 3	9	1,966	0.88
	Grade 5	9	1,915	0.87
Sindhi	Grade 3	9	2,038	0.88
	Grade 5	9	2,038	0.88

Note that there were actually eight tasks, but one of the tasks (Task 7) was administered and scored in two parts, so the equivalent of nine tasks were used for the analysis.

Score Calculation

The EGRA data was analyzed three ways. First, p-values and item-total correlations were generated for assessing the difficulty and discrimination of the items and tasks. Second, the percent correct for each task provided an indication of the Sindh students' mastery of the tasks. Third, Sindh students' fluency was assessed. All testing in Sindh took place in either Urdu or Sindhi.

Item P-values and Item-Total Correlations

P-values and item-total correlations are classical test theory statistics that are used to evaluate the performance of individual items and the tasks they comprise. Item difficulty is measured by p-values, which range from 0.00 to 1.00. Higher p-values indicate easier items, because a higher percentage of students posted correct responses. The other classical statistic is the item-total correlation, and it ranges from -1.00 to +1.00. This statistic measures how close the item or task relates to the overall percent correct on the summary score. Values above 0.2 are an indication of a good item or task.

Percent Correct

The results of the EGRA testing were calculated using task and summary scores. Table 5 lists the tasks, stimuli, raw score ranges, and the method for calculating the task and summary scores on the test. For

each of the tasks, the stimuli (items) (i.e., questions, letters, sounds, words, and non-words) were worth one score point. The score points were added, and since the range of raw scores varies across the tasks, the percent of correct scores was used to report all results. No weighting was used with the tasks to calculate the summary scores. Each task summary score was calculated using the total number correct and dividing it by the number of items. The overall Reading Summary Score was calculated by adding all of the task summary scores and dividing by nine (total number of tasks) to arrive at the average.

Timed Tasks Scores

The scores on the timed tasks were calculated by taking the number of correct responses times 60 seconds then dividing that number by the number of seconds used to read the stimulus. For instance, if a student read 75 letters correctly in 30 seconds, their letters-correct-per-minute score would be 150 (75 words x 60 seconds/30 seconds). Given another example, if a student read 50 words correctly in 30 seconds, his or her timed task score would be 100 words per minute (50 words x 60 seconds/30 seconds). Table 5 lists the number of stimuli per task. Recall the percent correct scores ranged from zero to 100. The method for calculating phonics and fluency scores yielded much higher maximum values, upwards of 200 at baseline (see task box plots in Annex 2, Figures A1u-A2u and A1s-A2s).

TABLE 5: EGRA SCORE RANGES AND CALCULATIONS

Task (Subtest)	Stimuli	Score Range	Calculation
1. Orientation to print	5 questions (untimed)	0-5	Percent correct of answers
2. Letter name recognition	100 letters (timed)	0-100	Percent correct of letters
3. Phonemic awareness	10 questions (untimed)	0-10	Percent correct of words
4. Letter sound knowledge	100 sounds (timed)	0-100	Percent correct of sounds
5. Familiar word reading	50 words (timed)	0-50	Percent correct of words
6. Non-word reading	50 non-words (timed)	0-50	Percent correct of non-words
7a. Passage reading	60 words (timed)	0-60	Percent correct of words
7b. Passage comprehension	5 questions (untimed)	0-5	Percent correct of answers
8. Listening comprehension	3 questions (untimed)	0-3	Percent correct of answers
Reading Summary Score	-	-	Average of percent correct

An example of percent correct scores for each of the tasks and as a summary score is provided below. The raw score is divided by the maximum score (the highest score possible in the score range) to produce the percent correct score for each task. Then, the task scores are averaged to produce the summary score. Note that each of the task percent correct scores is weighted equally to provide the summary score.

TABLE 6: EXAMPLE OF EGRA PERCENT CORRECT AND SUMMARY SCORES

Task (Subtest)	Maximum Score	Raw Score	% Correct Score
1. Orientation to print	5	3	60.0%
2. Letter name recognition	100	68	68.0%
3. Phonemic awareness	10	5	50.0%
4. Letter sound knowledge	100	42	42.0%
5. Familiar word reading	50	34	68.0%
6. Non-word reading	50	25	50.0%
7a. Passage reading	60	50	83.3%
7b. Passage comprehension	5	2	40.0%
8. Listening comprehension	3	1	33.3%
Reading Summary Score	--	--	55.0%

An example of timed task scores (adjusted) is provided below for the five fluency tasks. The formula explained above is used (timed task score = raw score x 60 seconds/seconds used).

TABLE 7: EXAMPLE OF EGRA TIMED TASK SCORES

Task (Subtest)	Raw Score	Seconds Used	Timed Task Score
2. Letter name recognition	68	48	85.0
4. Letter sound knowledge	42	60	42.0
5. Familiar word reading	34	48	42.5
6. Non-word reading	25	40	37.5
7a. Passage reading	50	40	75.0

CHAPTER 3: URDU FINDINGS AND RESULTS

This chapter presents the findings and results from the Urdu EGRA baseline in Sindh. There are sections on the student sample, task and item statistics, score calculation, task and summary scores, timed task scores, and questionnaire findings.

Student Sample

The intended sample was 70 full and 70 light treatments schools. Within these schools, the target was to assess 15 students in each grade and gender per school, totaling 4,200 students; 2,100 for each gender, treatment, and grade. Table 8u shows the number of students in the sample by grade and gender for each language. For the full treatment group in grades 3 and 5, the actual samples were 97.8 percent and 97.2 percent of the intended sample, respectively. For the light treatment group, the actual sample size was 89.4 percent for grade 3 and 85.1 percent for grade 5. The entire Urdu grade 3 sample was 93.6 percent and 91.2 percent for grade 5. The boys' percent (85.6) was lower than the girls' (97.6). A small number of students in grade 3 ($n = 2$) and grade 5 ($n = 6$) did not complete the gender item on the questionnaire. The total actual sample in Urdu-medium schools in Sindh was 3,881 students, 92.4 percent of the intended 4,200 sample records.

TABLE 8U: URDU ACTUAL STUDENT SAMPLE BY GRADE AND GENDER

Treatment	Grade Level	Sample	Boys	Girls	Missing	Total
Full Treatment	Grade 3	Students	496	524	7	1,027
		% of Target	94.5%	99.8%	--	97.8%
	Grade 5	Students	468	543	10	1,021
		% of Target	89.1%	103.4%	--	97.2%
	Total	Students	964	1,067	17	2,048
		% of Target	91.8%	101.6%	--	97.5%
Light Treatment	Grade 3	Students	434	500	5	939
		% of Target	82.7%	95.2%	--	89.4%
	Grade 5	Students	399	483	12	894
		% of Target	76.0%	92.0%	--	85.1%
	Total	Students	833	983	17	1,833
		% of Target	81.7%	93.6%	--	87.3%
Full and Light Treatment	Grade 3	Students	930	1,024	5	1,966
		% of Target	88.6%	97.5%	--	93.6%
	Grade 5	Students	867	1,026	12	1,915
		% of Target	82.6%	97.7%	--	91.2%
	Total	Students	1,797	2,050	34	3,881
		% of Target	85.6%	97.6%	--	92.4%

Task and Item Statistics

Table 9u shows the statistics for the tasks on the Urdu test. Two classical statistics are provided: p-values and item-total correlations. P-values indicate the average score of the students on the tasks, or the difficulty of the tasks for the students. The item-total correlations in the table are actually task-total correlations, which indicate the degree to which the tasks can discriminate between low- and high-achieving students; this is an indicator of the quality of the items. P-values can range from 0.00 to 1.00, with higher values indicating easier items. Item-total correlations can range from -1.00 to +1.00, with values above +0.20 or +0.25 indicating that the item (or task) is of good quality.

In Table 9u below, task p-values, and item-total correlations are provided for the full and light treatment groups combined. For grade 3, task p-values ranged from 0.19 to 0.57, thus providing a spread on the lower and middle parts of the difficulty spectrum. The p-values for grade 5 ranged from 0.33 to 0.69 or in the middle part of the spectrum. All of the Urdu task scores in the grades 3 and 5 had item-total correlations of greater than 0.26, indicating good quality for these tasks. Note that the untimed tasks also have item statistics, since each of those tasks has multiple items. Full item statistics for the Urdu test items are provided in Annex 1 at the end of this report.

TABLE 9U: URDU TASKS STATISTICS (FULL AND LIGHT TREATMENT GROUPS)

Task (Subtest)	Grade 3		Grade 5	
	P-Value	Item-Total	P-Value	Item-Total
1. Orientation to print (untimed)	0.57	0.27	0.62	0.38
2. Letter name recognition (timed)	0.45	0.61	0.57	0.54
3. Phonemic awareness (untimed)	0.34	0.34	0.41	0.37
4. Letter sound knowledge (timed)	0.25	0.56	0.33	0.44
5. Familiar word reading (timed)	0.39	0.84	0.69	0.82
6. Non-word reading (timed)	0.29	0.82	0.53	0.81
7a. Passage reading (timed)	0.40	0.83	0.69	0.82
7b. Passage comprehension (untimed)	0.19	0.79	0.43	0.77
8. Listening comprehension (timed)	0.32	0.55	0.49	0.50

Task and Summary Scores

The next part of the analysis involved plotting the summary scores. Histograms of the summary scores (Figures 2u and 3u) show that the distributions are moving from left to right from grade 3 to grade 5, which is strong evidence that the children are learning basic skills at the primary school level. In addition, as with the task and item statistics, it also shows that there is room for growth at each grade level. The main goal of the intervention is to see movement of the score distributions to the right within the same grade level (i.e., grades 3 and 5) from the baseline to midline to endline.

FIGURE 2U: URDU GRADE 3 SUMMARY SCORES

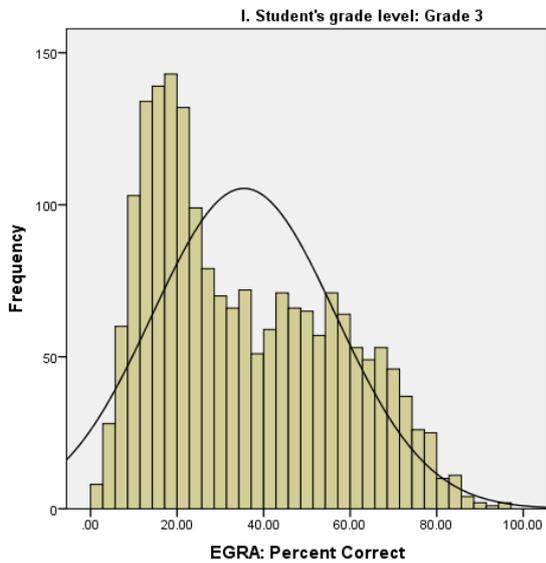


FIGURE 3U: URDU GRADE 5 SUMMARY SCORES

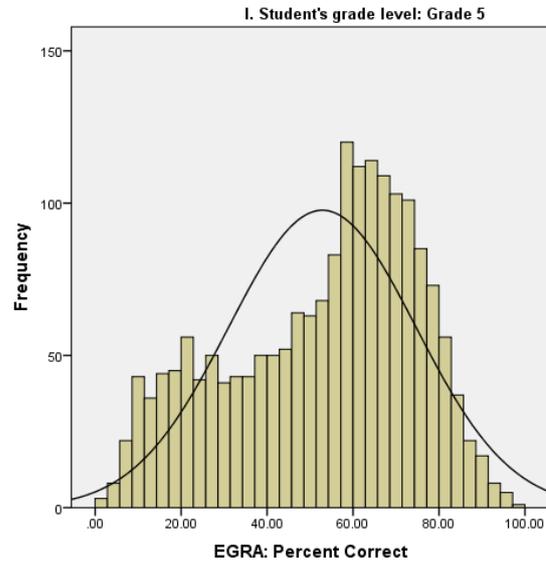


Table 11u and Figures 4u and 5u provide the average scores by task using percent correct scores. The score for each task was calculated using the total number correct and dividing by the number of items. For instance, a student who scored 3 out of 5 on Task 1 would receive a score of 60 percent. Averages were then calculated for all students on Task 1, which in Sindh in Urdu was 56.5 percent for grade 3 and 62.2 percent for grade 5. The same type of calculation was made for each student and each task. The table also includes the differences from grade 3 to grade 5, e.g., 56.5 percent minus 62.2 percent equals 6.7 percentage points.

The full treatment scores were higher for all tasks, especially in letter sound knowledge (10 percent and 13 percent for grades 3 and 5, respectively) and listening comprehension (8 percent and 12 percent for grades 3 and 5, respectively). This will be corrected statistically at the midline and endline. Grade 3 students did relatively better on the orientation to print, letter name recognition, and passage reading. The lowest scores were in phonics (non-word reading and letter sound knowledge) and comprehension (passage and listening). At grade 5, their best scores were in the two reading tasks (familiar word and passage). Both grades still had relatively low scores in phonics (non-word reading, letter sound knowledge) and comprehension (passage and listening).

There was also substantial progression from grade 3 to grade 5 on the summary score (17). The 17-point gain was consistent across genders and treatment groups. The tasks scores with the greatest gains were in familiar word reading (30.5), and passage reading (29) and comprehension (23).

TABLE 10U: URDU PERCENT CORRECT SCORES BY GRADE AND TASK (FULL AND LIGHT TREATMENT GROUPS)

Task (Subtest)	Grade 3	Grade 5	Difference (G5 – G3)
1. Orientation to print	56.5%	62.2%	6.7% points
2. Letter name recognition	44.6%	56.8%	12.2% points
3. Phonemic awareness	34.4%	40.5%	6.1% points
4. Letter sound knowledge	25.1%	33.0%	7.9% points
5. Familiar word reading	38.5%	69.0%	31.5% points
6. Non-word reading	28.5%	53.4%	24.9% points
7a. Passage reading	39.8%	69.0%	29.2% points
7b. Passage comprehension	19.3%	42.5%	23.2% points
8. Listening comprehension	32.2%	49.3%	17.1% points
Reading Summary Score	35.4%	52.9%	17.5% points

TABLE 11U: URDU SCORES BY GRADE, TASK, AND GROUP

Task (Subtest)	Full		Light	
	Grade 3	Grade 5	Grade 3	Grade 5
1. Orientation to print	58.1%	66.5%	54.9%	57.3%
2. Letter name recognition	47.2%	58.6%	42.1%	54.8%
3. Phonemic awareness	35.6%	44.3%	33.1%	36.2%
4. Letter sound knowledge	29.8%	38.8%	20.0%	26.4%
5. Familiar word reading	41.8%	73.4%	35.1%	64.2%
6. Non-word reading	31.3%	59.0%	25.4%	47.0%
7a. Passage reading	43.5%	73.8%	35.8%	63.7%
7b. Passage comprehension	21.0%	47.4%	17.5%	36.8%
8. Listening comprehension	35.8%	54.5%	28.2%	43.4%
Reading Summary Score	38.2%	57.4%	32.5%	47.8%

FIGURE 4U: URDU FULL TREATMENT SCORES BY GRADE AND TASK (FULL AND LIGHT TREATMENT GROUPS)

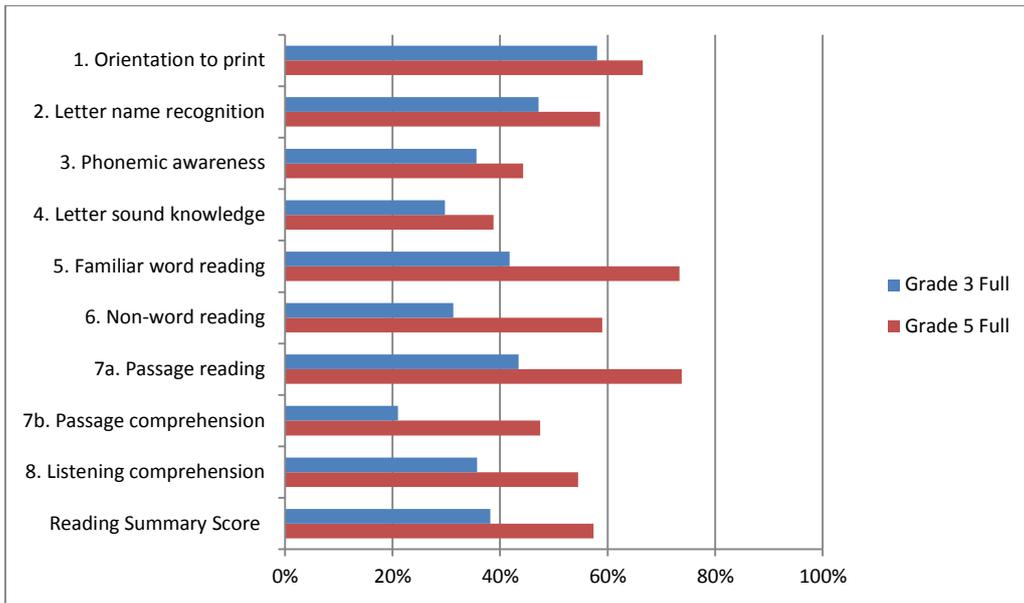
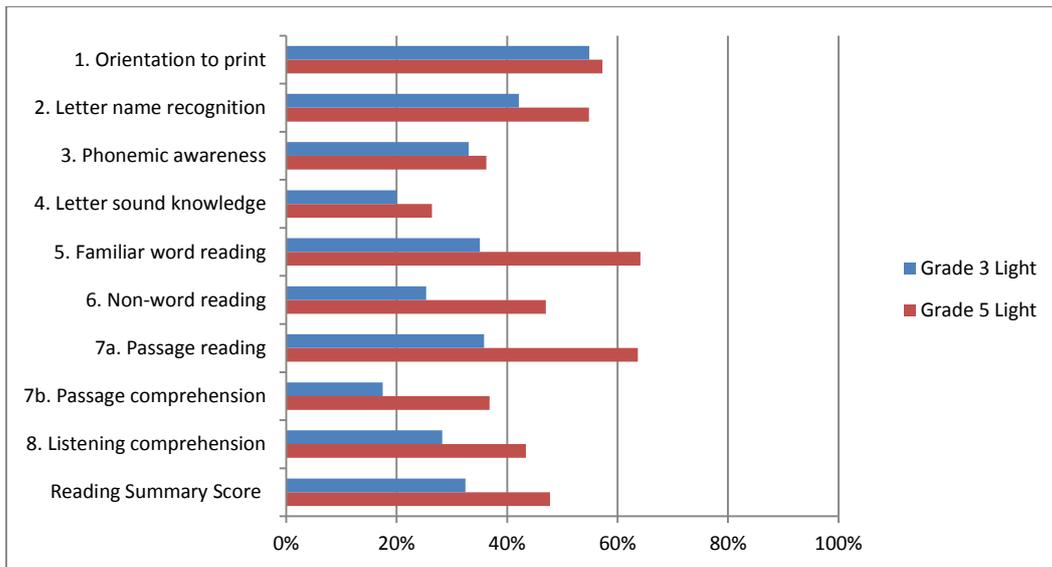


FIGURE 5U: URDU LIGHT TREATMENT SCORES BY GRADE AND TASK (FULL AND LIGHT TREATMENT GROUPS)



There were differences between boys and girls on the task and summary scores, but most of these differences were small (Table 12u). For both grades there was about an 11-point difference on the summary score favoring the girls. Girls had higher scores on familiar word reading, non-word reading, and passage reading and comprehension. Note that these are for the full and light treatment groups combined.

TABLE 12U: URDU SCORES BY GRADE, TASK, AND GENDER (FULL AND LIGHT TREATMENT GROUPS)

Task (Subtest)	Grade 3		Grade 5	
	Boys	Girls	Boys	Girls
1. Orientation to print	56.2%	56.9%	61.2%	63.2%
2. Letter name recognition	39.7%	49.5%	52.4%	60.8%
3. Phonemic awareness	32.5%	0.4%	37.2%	43.5%
4. Letter sound knowledge	21.1%	28.9%	27.8%	37.5%
5. Familiar word reading	28.8%	47.7%	60.4%	76.5%
6. Non-word reading	20.6%	35.9%	45.5%	60.3%
7a. Passage reading	30.1%	50.4%	61.1%	78.3%
7b. Passage comprehension	12.2%	26.1%	31.8%	51.7%
8. Listening comprehension	27.6%	36.4%	45.8%	52.7%
Reading Summary Score	29.9%	40.7%	46.9%	58.2%

FIGURE 6U: URDU GRADE 3 SCORES BY TASK AND GENDER (FULL AND LIGHT TREATMENT GROUPS)

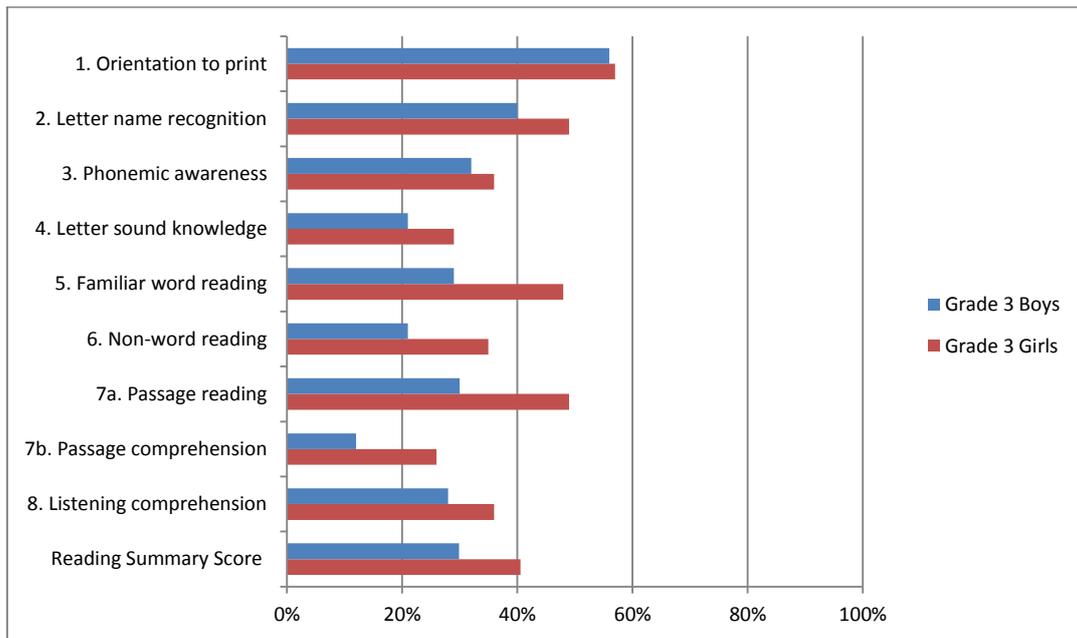
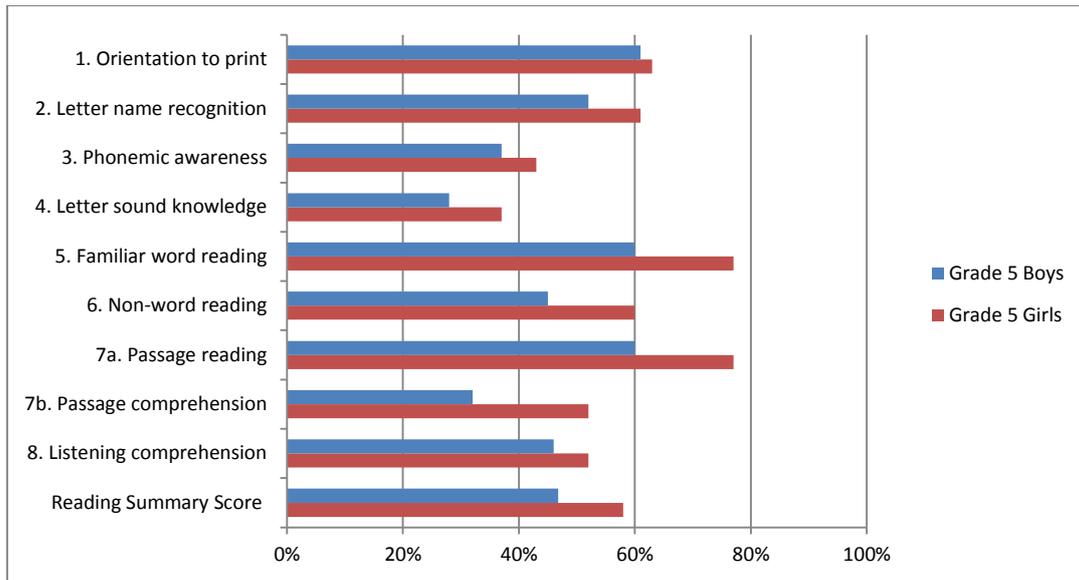


FIGURE 7U: URDU GRADE 5 SCORES BY TASK AND GENDER (FULL AND LIGHT TREATMENT GROUPS)



The final table in this section (Table 13u) further disaggregates the scores by treatment group, grade level, and gender. As seen in the tables above, the light treatment group scored higher on some of the tasks, which will be statistically corrected at the midline and endline. There were some variations in the scores by gender and treatment group. For instance, on many of the tasks, the girls scored higher than the boys in the full treatment group, but the boys scored higher than the girls in the light treatment group. Further investigation would be required to determine the reasons for this trend.

TABLE 13U: URDU PERCENT CORRECT SCORES BY GROUP, GRADE, GENDER AND TASK

Task (Subtest)	Full Treatment				Light Treatment			
	Grade 3		Grade 5		Grade 3		Grade 5	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
1. Orientation to print	56.7%	59.4%	64.7%	68.3%	55.6%	54.2%	57.2%	57.5%
2. Letter name recognition	42.7%	51.6%	54.7%	62.1%	36.3%	47.3%	49.7%	59.2%
3. Phonemic awareness	33.6%	37.5%	41.4%	46.8%	31.3%	34.8%	32.2%	39.8%
4. Letter sound knowledge	25.4%	34.0%	31.7%	44.9%	16.2%	23.5%	23.3%	29.4%
5. Familiar word reading	32.0%	51.2%	63.6%	81.7%	25.2%	44.0%	56.7%	70.7%
6. Non-word reading	23.7%	38.7%	50.3%	66.4%	17.0%	32.9%	39.8%	53.6%
7a. Passage reading	33.3%	53.6%	63.7%	82.5%	25.4%	45.3%	55.9%	70.9%
7b. Passage comprehension	13.5%	28.5%	35.9%	57.6%	10.7%	23.5%	27.0%	45.2%
8. Listening comprehension	31.1%	40.4%	50.4%	58.2%	23.7%	32.3%	40.4%	46.7%
Reading Summary Score	32.4%	43.7%	50.7%	63.2%	26.8%	37.5%	42.3%	52.3%

Timed Task: Phonics and Reading-Rate Fluency Scores

Fluency is a measure of reading efficiency. On the Pakistan EGRA baseline, there were two types of fluency measures: phonics and reading rate. The phonics-fluency subtest included letter name recognition, letter sound knowledge, and non-word reading, whereas, the reading-rate fluency subtest consisted of familiar word and passage reading.

Tables 14u to 18u below show scores in terms of raw scores (instead of the percent correct scores on the previous tables). Table 14u has the maximum raw scores attained by students on each task at each grade level. Tables 15u to 18u have mean scores for the students. In addition, adjustments were made to the raw scores for those students who finished the task before the end of one minute. For instance, if a student read 50 words correctly in 30 seconds, their words correct per minute score would be 100 (50 words x 60 seconds/30 seconds). This is a measure of reading fluency. Table 14u provides the baseline maximum scores at grades 3 and 5 for the five timed tasks.

TABLE 14U: URDU BASELINE MAXIMUM SCORES ON FLUENCY (TIMED) TASKS (FULL AND LIGHT TREATMENT GROUPS)

Phonics Fluency Subtest	Grade 3	Grade 5
2. Letter name recognition	143	390
4. Letter sound knowledge	150	125
6. Non-word reading	108	159
Reading-Rate Fluency Subtest	Grade 3	Grade 3
5. Familiar word reading	150	214
7a. Passage reading	240	277

The highest scores on the timed tasks at grade 3 were in letter name recognition (Table 15u). All other tasks had lower scores. At grade 5, the highest scores were in letter name recognition and familiar word reading. The areas of greatest progress from grade 3 to grade 5 were two reading tasks: familiar word reading and non-word reading.

TABLE 15U: URDU PHONICS AND READING-RATE FLUENCY TASK MEANS BY GRADE (FULL AND LIGHT TREATMENT GROUPS)

Phonics Fluency Subtest	Grade 3	Grade 5	Difference (G5 – G3)
2. Letter name recognition	45.4	58.7	13.3 points
4. Letter sound knowledge	33.5	43.8	10.3 points
6. Non-word reading	22.6	38.7	16.1 points
Reading-Rate Fluency Subtest	Grade 3	Grade 5	Difference (G5 – G3)
5. Familiar word reading	33.2	61.0	27.9 points
7a. Passage reading	26.2	54.4	28.2 points

At grade 3, the largest differences between the treatment groups were in letter name recognition and letter sound knowledge (Table 16u). At grade 5, the largest differences were in letter name recognition, letter sound knowledge and non-word reading. These differences will be corrected at midline and endline.

TABLE 16U: URDU PHONICS AND READING-RATE FLUENCY TASK MEANS BY GRADE AND GROUP (FULL AND LIGHT TREATMENT GROUPS)

Phonics Fluency Subtest	Grade 3		Grade 5	
	Full	Light	Full	Light
2. Letter name recognition	47.9	42.6	60.8	56.4
4. Letter sound knowledge	35.9	30.3	45.9	40.7
6. Non-word reading	23.7	21.3	40.2	36.7
Reading-Rate Fluency Subtest	Grade 3		Grade 5	
	Full	Light	Full	Light
5. Familiar word reading	34.4	31.7	62.8	58.7
7a. Passage reading	26.1	26.4	56.7	52.1

Girls in both grades had significantly higher fluency scores than the boys on each of the timed reading tasks (Table 17u). In grade 3 and 5, the greatest difference was in non-word and passage reading.

TABLE 17U: URDU PHONICS AND READING-RATE FLUENCY TASK MEANS BY GRADE AND GENDER (FULL AND LIGHT TREATMENT GROUPS)

Phonics Fluency Subtest	Grade 3		Grade 5	
	Boys	Girls	Boys	Girls
2. Letter name recognition	40.0	50.3	34.5	62.5
4. Letter sound knowledge	29.2	37.3	24.2	47.8
6. Non-word reading	17.8	26.3	22.8	42.7
Reading-Rate Fluency Subtest	Grade 3		Grade 5	
	Boys	Girls	Boys	Girls
5. Familiar word reading	25.5	39.0	33.7	70.0
7a. Passage reading	29.8	22.4	62.8	45.6

The final table in this section (Table 18u) further disaggregates the scores by treatment group, grade level, and gender. As with the percent correct scores, the light treatment group scored higher on some of the tasks, which will be statistically corrected at the midline and endline.

TABLE 18U: URDU PHONICS AND READING-RATE FLUENCY TASK MEANS BY GROUP, GRADE, AND GENDER

Phonics Fluency Subtest	Full Treatment				Light Treatment			
	Grade 3		Grade 5		Grade 3		Grade 5	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
2. Letter name recognition	42.8	52.5	56.8	63.6	36.4	47.5	50.0	60.0
4. Letter sound knowledge	25.4	34.4	31.9	45.1	16.9	23.5	23.5	29.6
6. Non-word reading	13.0	21.6	29.7	41.7	8.8	18.0	23.0	23.9
Reading-Rate Fluency Subtest	Full Treatment				Light Treatment			
	Grade 3		Grade 5		Grade 3		Grade 5	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
5. Familiar word reading	19.7	33.7	46.5	68.3	14.3	30.0	41.2	57.6
7a. Passage reading	25.3	44.8	58.2	92.5	17.5	37.8	51.3	76.2

Urdu Questionnaire Findings

Selected results are presented below for the Urdu questionnaires, including for those characteristics or items that showed significant differences in student scores. Due to the students having the same language (Urdu), the results were combined for the full and light treatment groups to increase the sample size and more accurately detect effects between the categories. Note that there were some students, teachers, and head teachers who did not respond to certain questionnaire items; they were labeled as missing. The total averages for the summary scores were calculated based on those who responded.

Since these are baseline data, reporting on the full and light treatment groups together will not affect the analyses at midline and endline. We combined the survey data for the groups since some of the questions led to reporting by relatively small categories (e.g., for teacher qualifications) and we wanted to know whether the survey results were associated with the student scores in general.

In addition, since the samples were by treatment group, the results will be generalized to the populations for each group. This will be done prior to the midline. The results will be generalized to by calculating sampling weights, applying the weights to the results, and then generalizing to the population by treatment group. We will also do this for the midline and endline. The current analyses only apply to the sampled districts.

Statistical significance was determined based on *t*-tests for indicators with two categories and analyses of variance for indicators with three or more categories (with post hoc pairwise comparisons). The significance value was set at $p < 0.05$; a 95 percent confidence level.

Urdu Student Questionnaires

Table 19u has summary scores by student age. According to the National Education Policy (2009), the official age of the students at the beginning of the different grade levels of primary education is 6 to 10 years old. Since the baseline took place during the school year, the normal ages for this analysis were set at 8 to 9 years old for grade 3 and 10 to 11 years old for grade 5. The students were placed into three categories: younger than normal age for their grade, normal age, and older than normal age. At grade 3, there were significant differences among all three age groups; younger students had lower average summary scores. Conversely, at grade 5, the scores were not significantly different among age groups, eradicating the older-student advantage by grade 5.

TABLE 19U: URDU SUMMARY SCORES BY STUDENT AGE

Age Group	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
Younger than normal age	131	27.7%*	120	48.6%
Normal age	725	32.8%*	785	53.3%
Older than normal age	1,092	38.3%*	989	53.0%
Missing	18	--	21	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

Table 20u shows the summary scores according to whether the student reads the Quran at home. There were significant differences in both grades in favor of students who read the Quran. Note that the difference was significant in spite of the large disparity in n-counts in the response categories.

TABLE 20U: URDU SUMMARY SCORES BY READING THE QURAN AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	124	25.4%	106	40.5%
Yes	1,784	36.6%*	1,759	53.9%*
Missing	58	--	50	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

Table 21u depicts the differences in scores based on whether there is a library at the school. There was a significant difference for the grade 5 students, but not a significant difference for third graders.

TABLE 21U: URDU SUMMARY SCORES BY THE PRESENCE OF A LIBRARY AT THE SCHOOL

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,334	35.9%	1,298	52.3%
Yes	822	35.0%	480	55.0%*
Missing	180	--	137	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

In Tables 22u to 24u, the data showed that the existence of newspapers and magazines generally made a difference in reading scores in most cases. On the other hand, the effect on scores of the presence of books at home was positive and only statistically significant for fifth graders.

TABLE 22U: URDU SUMMARY SCORES BY THE PRESENCE OF NEWSPAPERS AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,134	34.2%	1,001	51.3%
Yes	822	37.1%*	898	54.7%*
Missing	10	--	16	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

TABLE 23U: URDU SUMMARY SCORES BY THE PRESENCE OF MAGAZINES AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,740	34.7%	1,584	51.3%
Yes	216	41.5%*	315	61.0%*
Missing	10	--	16	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level.

TABLE 24U: URDU SUMMARY SCORES BY THE PRESENCE OF BOOKS AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,380	35.0%	1,292	51.9%
Yes	576	36.5%	607	55.0%*
Missing	0	--	0	--
Total	1,956	35.5%	1,899	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.01$ level

The final set of student questions (in Tables 25u to 27u) pertained to children's reading habits at home. In general, these habits made a difference in student scores in all cases in grade 5 and in grade 3. Having someone read to children at home, having children read to someone else at home, and children reading silently at home produced significantly higher reading scores than children who did not read aloud at home.

TABLE 25U: URDU SUMMARY SCORES BY CHILDREN HAVING SOMEONE READ TO THEM AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	881	32.8%	837	49.5%
Yes	1,051	37.7%*	1,049	55.6%*
Missing	47	--	29	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.01$ level

TABLE 26U: URDU SUMMARY SCORES BY CHILDREN READING TO SOMEONE ELSE AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,030	32.3%	831	48.7%
Yes	917	39.2%*	1,060	56.2%*
Missing	19	--	24	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.01$ level

TABLE 27U: URDU SUMMARY SCORES BY CHILDREN READING SILENTLY AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	403	31.3%	474	47.0%
Yes	1,540	36.6%*	1,412	55.0%*
Missing	23	--	29	--
Total	1,966	35.5%	1,915	52.9%

* Indicates that the performance of the group was significantly higher, $p < 0.01$ level

Urdu Teacher Questionnaires

With the smaller sample size, the analysis of the teacher questionnaires was limited to descriptive statistics, i.e., no group comparisons. Table 28u and Table 29u provide information on teacher academic and professional qualifications, neither of which showed consistent patterns in the student scores.

TABLE 28U: URDU SUMMARY SCORES BY TEACHER ACADEMIC QUALIFICATION

Academic Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.A./M.Sc./M.Phil.	20	35.0%	37	48.9%
B.A./B.Sc.	71	35.4%	62	50.3%
F.A./F.Sc.	25	37.6%	28	59.9%
Matric	11	31.7%	8	57.7%
Missing	0	--	0	--
Total	127	35.5%	135	52.4%

TABLE 29U: URDU SUMMARY SCORES BY TEACHER PROFESSIONAL QUALIFICATION

Professional Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.Ed./M.A.	8	38.2%	13	49.3%
B.Ed.	36	33.8%	45	49.7%
C.T.	11	40.7%	15	52.2%
P.T.C.	67	34.8%	57	54.3%
Missing	5	--	5	--
Total	127	35.3%	135	52.0%

In an analysis of student scores by teacher age and experience, there were no consistent patterns of younger or older teachers, or those with less or more experience, relating to lower or higher student scores (Tables 30u and 31u). Again, small teacher sample sizes made drawing conclusions difficult.

TABLE 30U: URDU SUMMARY SCORES BY TEACHER AGE

Age Group in Years	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
40 and less	41	36.5%	45	53.4%
Between 41 and 50	62	35.6%	65	51.1%
51 and more	22	38.5%	18	52.4%
Missing	2	--	3	--
Total	127	35.4%	135	52.1%

TABLE 31U: URDU SUMMARY SCORES BY TEACHER EXPERIENCE

Years of Experience	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
10 or less	26	38.2%	28	48.6%
Between 11 and 20	30	37.7%	41	56.5%
Between 21 and 30	57	32.4%	57	50.1%
31 or more	10	36.3%	5	55.2%
Missing	4	--	4	--
Total	127	35.2%	135	52.0%

For teachers, those who attended one or more in-service trainings had higher scores than those who never attended such trainings (Table 32u). Again, any differences should be interpreted with caution due to the small sample size.

TABLE 32U: URDU SUMMARY SCORES BY TEACHER IN-SERVICE TRAINING

Frequency of Training	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
None	85	34.4%	103	51.6%
One time	18	38.2%	16	57.8%
Two times	13	37.5%	7	57.4%
Three times	7	38.0%	6	49.5%
Missing	4	--	3	--
Total	127	35.5%	135	52.6%

Urdu Head Teacher Questionnaires

Similarly to the teachers, the sample size for the head teacher questionnaires was small, so data interpretations should be treated with caution. Tables 33u and 35u show head teacher academic and professional qualifications. In general, the results show that higher teacher academic and professional qualifications show no definitive pattern in student scores.

TABLE 33U: URDU SUMMARY SCORES BY HEAD TEACHER ACADEMIC QUALIFICATION

Academic Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.A./M.Sc./M.Phil.	48	33.2%	48	52.1%
B.A./B.Sc.	68	34.6%	68	51.1%
F.A./F.Sc.	18	34.1%	18	51.2%
Matric	5	54.6%	5	65.0%
Missing	1	--	1	--
Total	140	34.8%	140	52.0%

TABLE 34U: URDU SUMMARY SCORES BY HEAD TEACHER PROFESSIONAL QUALIFICATION

Professional Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.Ed./M.A.	25	35.4%	25	54.2%
B.Ed.	49	34.7%	49	51.2%
C.T.	21	36.1%	21	55.1%
P.T.C.	43	35.0%	43	50.1%
Missing	2	--	2	--
Total	140	35.5%	140	52.0%

Tables 35u and 36u provide information on head teachers' experience and in-service training. Teachers with two years or less experience had slightly lower reading scores than the more experienced head teachers. In terms of in-service trainings, head teachers who attended one or two in-service trainings had higher scores than those did not attend the trainings. This is the same pattern shown by teachers. Again, any differences should be interpreted with caution due to the small sample size.

TABLE 35U: URDU SUMMARY SCORES BY HEAD TEACHER EXPERIENCE

Years of Experience	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
2 or less	28	32.5%	28	47.9%
3 to 5	27	36.2%	27	54.6%
6 to 10	33	36.8%	33	53.4%
11 or more	43	36.8%	43	53.5%
Missing	9	--	9	--
Total	140	35.8%	140	52.6%

TABLE 36U: URDU SUMMARY SCORES BY HEAD TEACHER IN-SERVICE TRAINING

Frequency of Training	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
None	97	34.3%	97	52.6%
1 time	24	42.6%	24	54.1%
2 times	11	43.6%	11	59.2%
More than 2 times	6	33.7%	6	51.2%
Missing	2	--	2	--
Total	140	34.9%	140	52.0%

Tables 37u and 38u provide data on head teachers' support to teachers in reading and the training that head teachers received in teaching reading. There were too few head teachers that reported not supporting teachers in reading (9) to make valid conclusions. However, there were slightly higher reading scores shown when the head teacher received training in teaching reading.

TABLE 37U: URDU SUMMARY SCORES BY HEAD TEACHER SUPPORT TO TEACHERS IN READING

Support to Teachers	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	9	39.3%	9	56.1%
Yes	129	34.9%	129	52.1%
Missing	2	--	2	--
Total	140	35.5%	140	52.9%

TABLE 38U: URDU SUMMARY SCORES BY HEAD TEACHER TRAINING IN TEACHING READING

Support to Teachers	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	70	32.7%	70	50.6%
Yes	66	37.0%	66	53.4%
Missing	4	--	4	--
Total	140	35.5%	140	52.9%

Urdu School Characteristics

The final section provides information on school characteristics (from the head teacher questionnaires) by student summary scores. As with the teacher and head teacher characteristics, most patterns appeared to be inconclusive (Tables 39u to 42u). The female schools performed better than the male and mixed schools. Schools with libraries (only 10) had significantly better grade 5 scores than those without libraries, but this result must be interpreted with caution due to the low sample size. Better infrastructure seemed to have a positive relationship with student scores, but these results were not statistically significant ($p > 0.05$).

TABLE 39U: URDU SUMMARY SCORES BY SCHOOL GENDER

School Gender	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
Male school	39	30.5%	39	48.0%
Female school	49	40.2%	49	57.8%
Mixed school	50	32.8%	50	49.3%
Missing	2	--	2	--
Total	140	34.8%	140	52.0%

TABLE 40U: URDU SUMMARY SCORES BY PTA/SMC/PTSMC/PTC

Parent Teacher Committee	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1	57.8%	1	69.8%
Yes	138	34.9%	138	52.1%
Missing	1	--	1	--
Total	140	35.0%	140	52.1%

TABLE 41U: URDU SUMMARY SCORES BY PRESENCE OF A SCHOOL LIBRARY

School Library	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	130	34.5%	130	51.4%
Yes	10	41.1%	10	61.4%
Missing	0	--	0	--
Total	140	35.0%	140	52.1%

TABLE 42U: URDU SUMMARY SCORES BY INFRASTRUCTURE (DRINKING WATER, ELECTRICITY, TOILETS)

Number of Infrastructures (Water, Electricity, Toilets)	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
None	10	32.9%	10	48.0%
1	19	32.1%	19	50.4%
2	29	36.0%	29	50.1%
3	82	35.0%	82	53.7%
Missing	0	--	0	--
Total	140	35.0%	140	52.1%

CHAPTER 4: SINDHI FINDINGS AND RESULTS

This chapter presents the findings and results from the Sindhi EGRA baseline in Sindh. There are sections on the student sample, task and item statistics, score calculation, task and summary scores, timed task scores, and questionnaire findings.

Student Sample

The intended sample was 70 full and 70 light treatment schools. Within these schools, the target was to assess 15 students in each grade and gender per school; totaling 4,200 students; 2,100 for each gender, treatment, and grade. Table 8s shows the number of students in the sample by grade and gender for each language. For the full treatment group in both grades 3 and 5, the actual samples were 99.3 percent and 97.0 percent of the intended sample, respectively. For the light treatment group the actual sample size was 94.8 percent for grade 3 and 97.0 percent for grade 5. The entire Sindhi grade 3 and grade 5 sample was 97.0 percent. The boys' percent (100 percent) was higher than the girls' (93 percent). It is important to note that in some schools, boys were also enrolled in girls' schools. In these cases, if 15 girls were not available for the EGRA test, the enumerators chose boys to reach a total of 15 students per grade. As such, the number of boys reached exceeded the target number in the full treatment Sindhi group. A small number of students in grade 3 (n = 1) and grade 5 (n = 1) did not complete the gender item on the questionnaire. The total actual sample in Sindhi-medium schools in Sindh was 4,076 students, 97.0 percent of the intended 4,200 sample students.

TABLE 8S: SINDHI ACTUAL STUDENT SAMPLE BY GRADE AND GENDER

Sindhi Treatment	Grade Level	Sample	Boys	Girls	Missing	Total
Full Treatment	Grade 3	Students	576	464	3	1,043
		% of Target	109.7%	88.4%	--	99.3%
	Grade 5	Students	564	448	7	1,019
		% of Target	107.4%	85.3%	--	97.0%
	Total	Students	1,140	912	10	2,062
		% of Target	108.6%	86.9%	--	98.2%
Light Treatment	Grade 3	Students	485	508	2	995
		% of Target	92.4%	96.8%	--	94.8%
	Grade 5	Students	474	533	12	1,019
		% of Target	90.3%	101.5%	--	97.0%
	Total	Students	959	1041	14	2,014
		% of Target	94.0%	99.1%	--	95.9%
Full and Light Treatment	Grade 3	Students	1,061	972	5	2,038
		% of Target	101.0%	92.6%	--	97.0%
	Grade 5	Students	1,038	981	19	2,038
		% of Target	98.9%	93.4%	--	97.0%
	Total	Students	2,099	1,953	24	4,076
		% of Target	100.0%	93.0%	--	97.0%

Task and Item Statistics

In Table 9s below, the task p-values for grade 3 ranged from 0.22 to 0.57, thus providing a spread on the lower and middle section of the difficulty spectrum. The p-values for grade 5 ranged from 0.32 to 0.63, or in the upper-lower half and middle parts of the spectrum. All of the Sindhi task scores in grades 3 and 5 had item-total correlations of greater than 0.30, indicating good quality for these tasks. Note that the untimed tasks also have item statistics, since each of those tasks has multiple items. Full item statistics for the Sindhi test items are provided in Annex 1 at the end of this report.

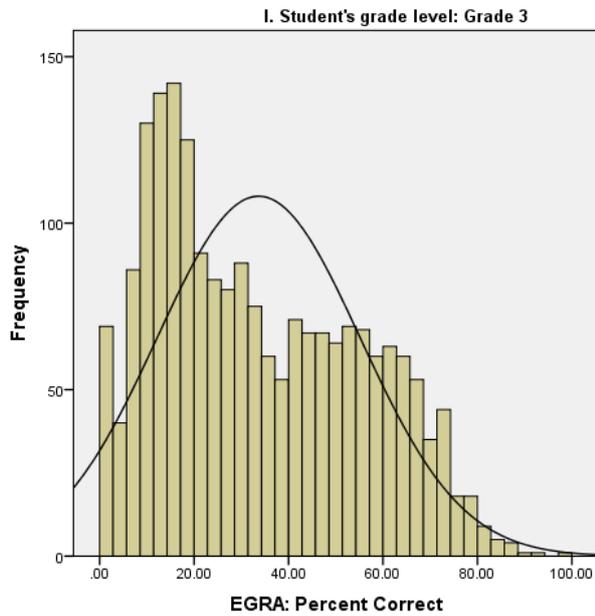
TABLE 9S: SINDHI TASK STATISTICS (FULL AND LIGHT TREATMENT GROUPS)

Task (Subtest)	Grade 3		Grade 5	
	P-Value	Item-Total	P-Value	Item-Total
1. Orientation to print (untimed)	0.57	0.39	0.61	0.35
2. Letter name recognition (timed)	0.34	0.63	0.44	0.55
3. Phonemic awareness (untimed)	0.28	0.34	0.32	0.30
4. Letter sound knowledge (timed)	0.26	0.61	0.40	0.64
5. Familiar word reading (timed)	0.39	0.84	0.63	0.85
6. Non-word reading (timed)	0.20	0.78	0.42	0.80
7a. Passage reading (timed)	0.37	0.85	0.62	0.86
7b. Passage comprehension (untimed)	0.22	0.79	0.41	0.79
8. Listening comprehension (timed)	0.41	0.44	0.52	0.45

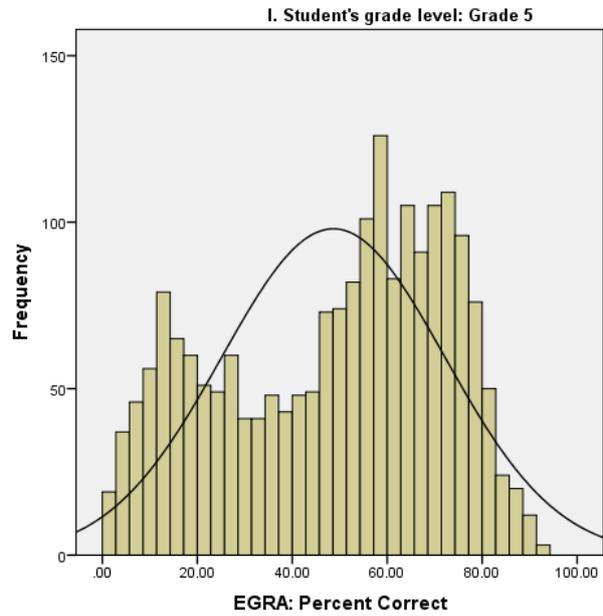
Task and Summary Scores

The next part of the analysis involved plotting the scores. Histograms of the summary scores (Figures 2s and 3s) show that the distributions are moving from left to right from grade 3 to grade 5, which is strong evidence that the children are learning basic skills at the primary school level. In addition, as with the task and item statistics, it also shows that there is room for growth at each grade level. The main goal of the intervention is to see movement of the score distributions to the right within the same grade level (i.e., grades 3 and 5) from the baseline to midline to endline.

**FIGURE 2S: SINDHI GRADE 3
SUMMARY SCORES**



**FIGURE 3S: SINDHI GRADE 5
SUMMARY SCORES**



Tables 10s and 11s, and Figures 4s and 5s, provide the average scores by task using percent correct scores. The score for each task was calculated using the total number correct and dividing by the number of items. For instance, a student who scored 3 out of 5 on Task 1 would receive a score of 60 percent. Averages were then calculated for all students on Task 1, which in Sindh in Sindhi was 56.6 percent for grade 3 and 61.4 percent for grade 5. The same type of calculation was made for each student and each task. The table also includes the differences from grade 3 to grade 5, e.g., 61.4 percent minus 56.6 percent equals 4.8 percentage points.

There were no consistent differences among the reading tasks between the full and light treatment groups. The light group had higher scores on some tasks while the full group had higher scores on others. Most of these differences were slight with the exception of letter name recognition (12.5 percent) and listening comprehension (12.5 percent) favoring the full treatment group. These differences will be corrected statistically at the midline and endline.

Grade 3 and 5 showed the same patterns on the reading tasks. Students did relatively better on orientation to print, listening comprehension, and familiar word and passage reading. The lowest scores were in phonics (non-word reading and letter sound knowledge) and reading comprehension.

There was also substantial progression from grade 3 to grade 5 on the summary score (15 points). The light treatment group had a 13-point increase whereas the full treatment had a 17-point increase.

TABLE 10S: SINDHI PERCENT CORRECT SCORES BY GRADE AND TASK (FULL AND LIGHT TREATMENT GROUPS)

Task (Subtest)	Grade 3	Grade 5	Difference (G5 – G3)
1. Orientation to print	56.6%	61.4%	4.8% points
2. Letter name recognition	34.1%	43.7%	9.6% points
3. Phonemic awareness	27.5%	32.4%	4.9% points
4. Letter sound knowledge	25.6%	40.3%	14.7% points
5. Familiar word reading	39.1%	63.4%	24.3% points
6. Non-word reading	20.1%	41.7%	21.6% points
7a. Passage reading	37.3%	61.6%	24.3% points
7b. Passage comprehension	21.7%	40.9%	19.2% points
8. Listening comprehension	41.1%	52.3%	11.2% points
Reading Summary Score	33.7%	48.6%	14.9% points

TABLE 11S: SINDHI SCORES BY GRADE, TASK, AND GROUP

Task (Subtest)	Full		Light	
	Grade 3	Grade 5	Grade 3	Grade 5
1. Orientation to print	55.3%	57.8%	57.9%	65.1%
2. Letter name recognition	38.5%	50.0%	29.4%	37.5%
3. Phonemic awareness	28.2%	32.9%	26.7%	31.9%
4. Letter sound knowledge	26.2%	42.0%	25.0%	38.6%
5. Familiar word reading	39.7%	67.7%	38.6%	59.0%
6. Non-word reading	18.9%	43.0%	21.25	40.4%
7a. Passage reading	37.6%	65.7%	37.0%	57.6%
7b. Passage comprehension	21.85	44.2%	21.65	37.6%
8. Listening comprehension	44.3%	58.6%	37.7%	46.1%
Reading Summary Score	34.5%	51.3%	32.8%	46.0%

FIGURE 4S: SINDHI FULL TREATMENT SCORES BY GRADE AND TASK

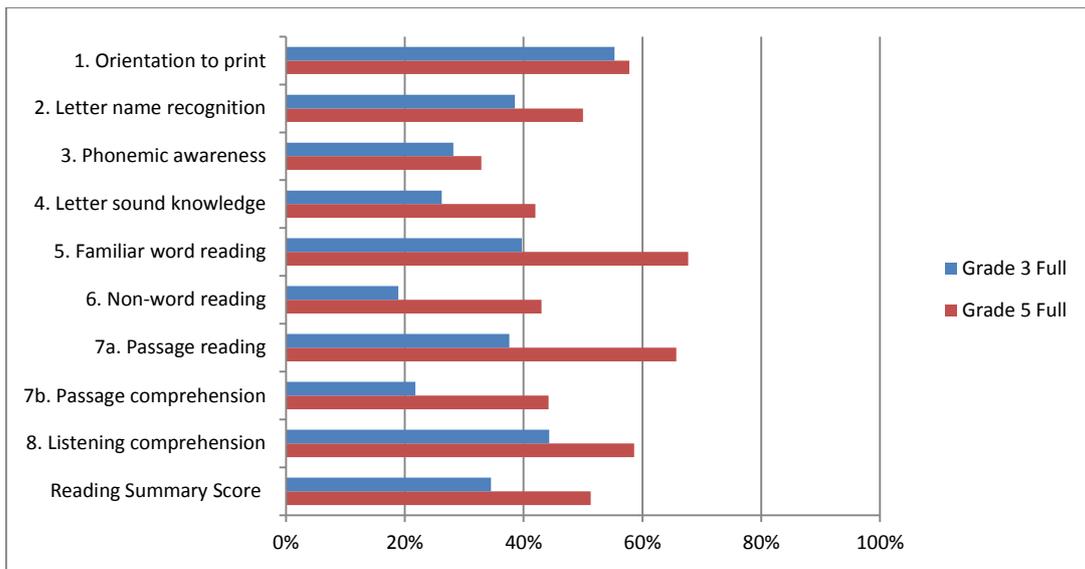
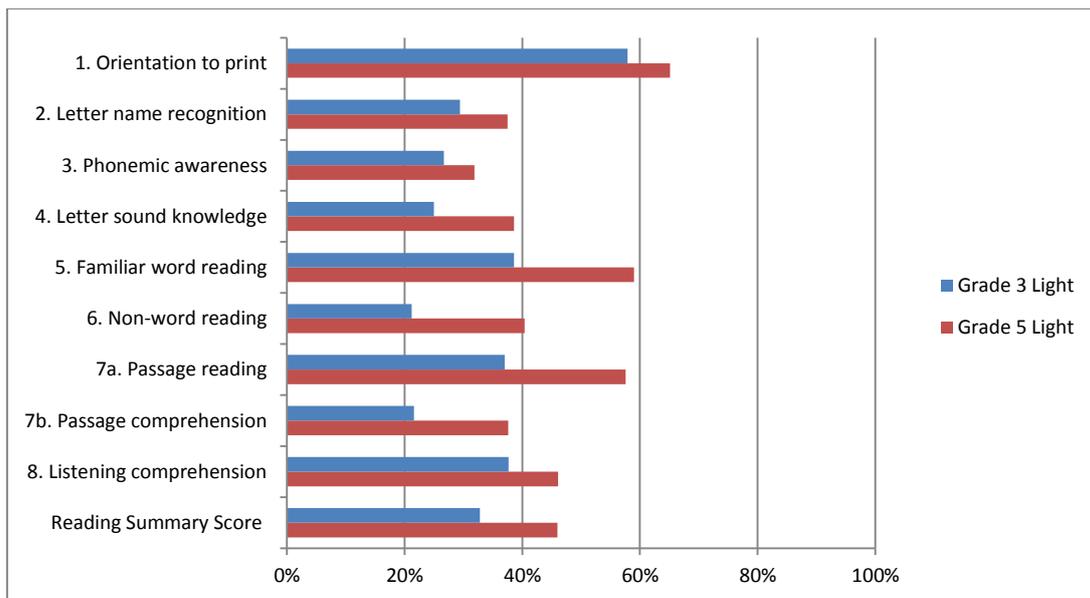


FIGURE 5S: SINDHI LIGHT TREATMENT SUMMARY SCORES BY GRADE AND TASK



The scores increased by 12 and 18 points for girls and boys, respectively. The task scores with the greatest gains were in familiar word reading (24), passage reading (24), non-word reading (22), and passage comprehension (19).

There were differences between boys and girls on the task and summary scores (Table 12s). For the task and summary scores, boys had statistically higher ($p < 0.01$) scores than girls. The summary score difference was 7 and 12 points, respectively for grades 3 and 5. The tasks with the largest differences were in familiar word reading and passage reading.

TABLE 12S: SINDHI SCORES BY GRADE, TASK, AND GENDER (FULL AND LIGHT TREATMENT GROUPS)

Task (Subtest)	Grade 3		Grade 5	
	Boys	Girls	Boys	Girls
1. Orientation to print	59.9%	52.9%	64.4%	58.3%
2. Letter name recognition	35.7%	32.3%	46.1%	41.3%
3. Phonemic awareness	29.0%	25.8%	33.5%	31.2%
4. Letter sound knowledge	28.1%	23.0%	44.7%	35.7%
5. Familiar word reading	44.6%	33.2%	73.3%	52.8%
6. Non-word reading	23.1%	16.9%	49.5%	33.5%
7a. Passage reading	42.7%	31.6%	71.4%	51.2%
7b. Passage comprehension	25.0%	18.3%	47.9%	33.5%
8. Listening comprehension	45.6%	36.2%	59.6%	44.6%
Reading Summary Score	37.0%	30.3%	54.5%	42.4%

FIGURE 6S: SINDHI GRADE 3 SCORES BY TASK AND GENDER (FULL AND LIGHT TREATMENT GROUPS)

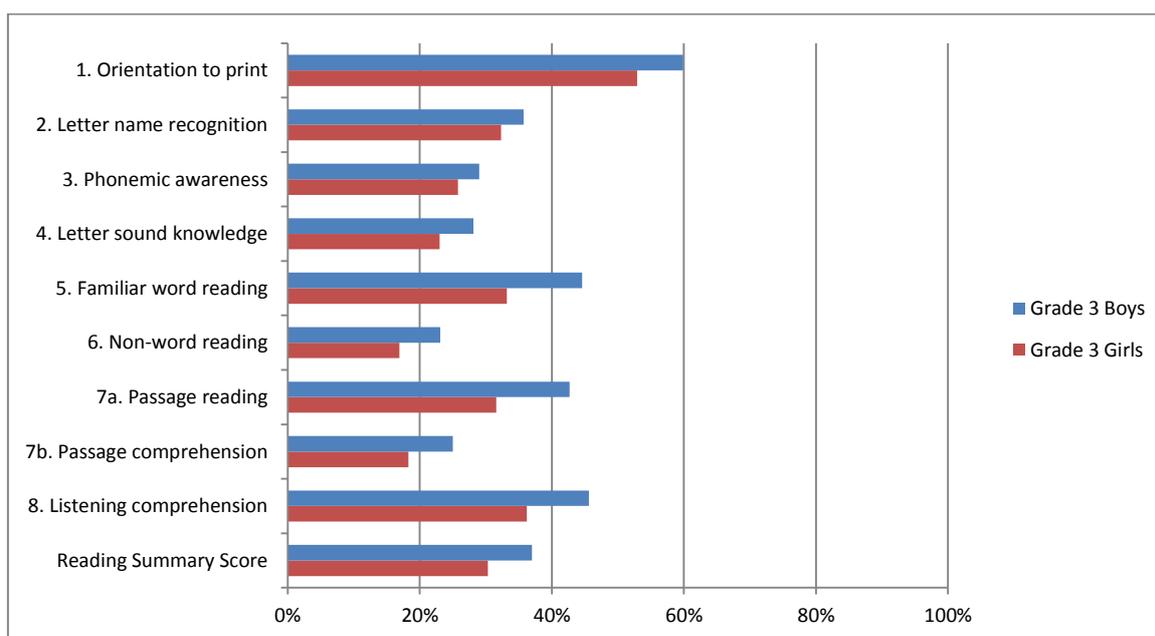
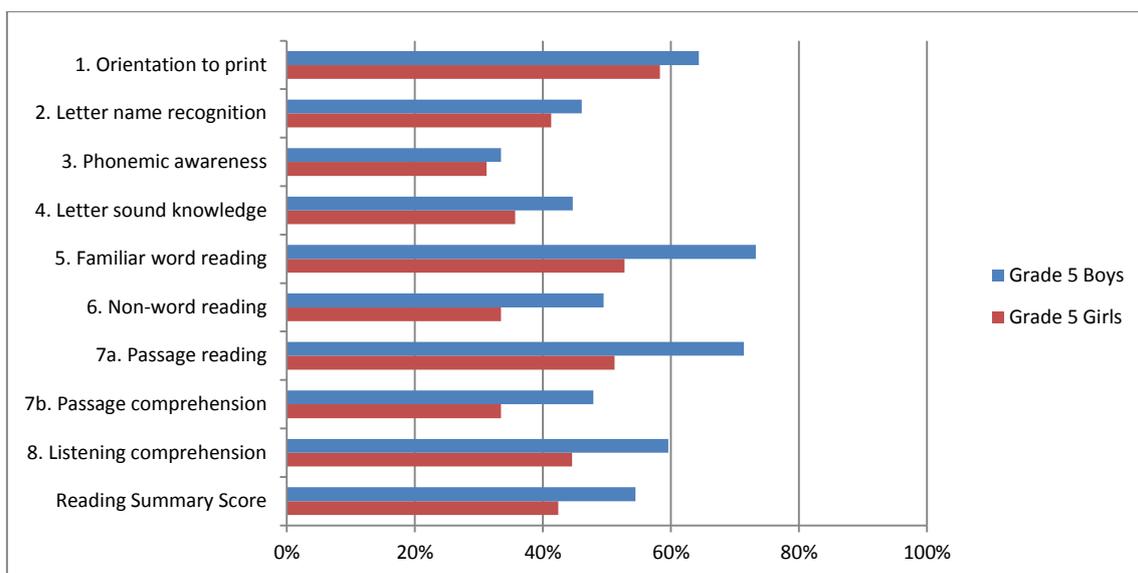


FIGURE 7S: SINDHI GRADE 5 SCORES BY TASK AND GENDER (FULL AND LIGHT TREATMENT GROUPS)



The final table in this section (Table 13s) further disaggregates the scores by treatment group, grade level, and gender. As seen in the tables above, the light treatment group scored higher on some of the tasks, which will be statistically corrected at the midline and endline. There were some variations in the scores by gender and treatment group. For instance, on many of the tasks, the girls scored higher than the boys in the full treatment group, but the boys scored higher than the girls in the light treatment group. Further investigation would be required to determine the reasons for this trend.

TABLE 13S: SINDHI PERCENT CORRECT SCORES BY GROUP, GRADE, GENDER AND TASK

Task (Subtest)	Full Treatment				Light Treatment			
	Grade 3		Grade 5		Grade 3		Grade 5	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
1. Orientation to print	56.9%	53.4%	58.9%	56.2%	63.5%	52.6%	70.8%	60.0%
2. Letter name recognition	34.5%	43.7%	48.0%	52.4%	37.2%	21.9%	43.8%	31.9%
3. Phonemic awareness	28.8%	27.5%	33.2%	32.4%	29.3%	24.3%	34.0%	30.2%
4. Letter sound knowledge	25.0%	27.9%	42.5%	41.6%	31.9%	18.5%	47.4%	30.7%
5. Familiar word reading	36.7%	43.5%	68.4%	66.6%	54.1%	23.9%	79.2%	41.1%
6. Non-word reading	17.3%	21.1%	43.9%	42.0%	30.1%	13.0%	56.1%	26.3%
7a. Passage reading	34.9%	41.2%	66.1%	64.8%	52.0%	22.8%	77.6%	39.7%
7b. Passage comprehension	20.1%	24.3%	44.4%	44.0%	30.9%	12.8%	52.1%	24.7%
8. Listening comprehension	46.4%	42.0%	61.4%	54.9%	44.7%	31.0%	57.6%	35.8%
Reading Summary Score	33.3%	36.1%	51.8%	50.6%	41.5%	24.5%	57.7%	35.6%

Timed Task: Phonics and Reading-Rate Fluency Scores

Tables 14s to 18s below show scores in terms of raw scores (instead of the percent correct scores on the previous tables). Table 14s has the maximum raw scores attained by students on each task at each grade level. Tables 15s to 18s have mean scores for the students. In addition, adjustments were made to the raw scores for those students who finished the task before the end of one minute. For instance, if a student read 50 words correctly in 30 seconds, their words correct per minute score would be 100 (50 words x 60 seconds/30 seconds). This is a measure of reading fluency. *T*-tests were generated to determine group differences. The significance for these tests is denoted in the tables. Table 14s provides the baseline maximum scores at grades 3 and 5 for the five timed tasks.

TABLE 14S: BASELINE MAXIMUM SCORES ON FLUENCY (TIMED) TASKS (FULL AND LIGHT TREATMENT GROUPS)

Phonics Fluency Subtest	Grade 3	Grade 5
2. Letter name recognition	120	159
4. Letter sound knowledge	119	156
6. Non-word reading	164	107
Reading-Rate Fluency Subtest	Grade 3	Grade 3
5. Familiar word reading	143	167
7a. Passage reading	166	221

The highest scores on the timed tasks at grade 3 were in letter name recognition (Table 15s). All other tasks had lower scores. At grade 5, the highest scores were in letter name recognition and familiar word reading and passage reading. All fluency tasks showed improvement from grades 3 to 5 ($p < 0.01$). The tasks with the largest increases were in familiar word reading and passage reading.

TABLE 15S: SINDHI PHONICS AND READING-RATE FLUENCY TASK MEANS BY GRADE (FULL AND LIGHT TREATMENT GROUPS)

Phonics Fluency Subtest	Grade 3	Grade 5	Difference (G5 – G3)
2. Letter name recognition	34.1	44.0	9.1 points
4. Letter sound knowledge	25.6	40.8	15.2 points
6. Non-word reading	10.3	23.0	12.7 points
Reading-Rate Fluency Subtest	Grade 3	Grade 5	Difference (G5 – G3)
5. Familiar word reading	23.3	46.7	23.4 points
7a. Passage reading	26.2	54.4	28.2 points

At third grade, the full treatment group had higher scores for letter name recognition; conversely the light treatment had higher scores in non-word reading (Table 16s). At fifth grade, the largest differences were in letter name recognition and familiar word reading, favoring the full treatment group. These differences will be statistically corrected in the midline and endline evaluations.

TABLE 16S: SINDHI PHONICS AND READING-RATE FLUENCY TASK MEANS BY GRADE AND GROUP

Phonics Fluency Subtest	Grade 3		Grade 5	
	Full	Light	Full	Light
2. Letter name recognition	38.6	29.5	50.4	37.6
4. Letter sound knowledge	25.6	25.1	42.5	39.1
6. Non-word reading	9.5	11.2	23.3	22.7
Reading-Rate Fluency Subtest	Grade 3		Grade 5	
	Full	Light	Full	Light
5. Familiar word reading	23.7	22.9	49.4	44.0
7a. Passage reading	26.1	26.4	56.7	52.1

Boys in both grades had significantly higher ($p < 0.001$) fluency scores than the girls on each of the timed reading tasks (Table 17s). For both grades, the largest differences were in passage reading and familiar word reading. The gender differences were greater in grade 5.

TABLE 17S: SINDHI PHONICS AND READING-RATE FLUENCY TASK MEANS BY GRADE AND GENDER (FULL AND LIGHT TREATMENT GROUPS)

Phonics Fluency Subtest	Grade 3		Grade 5	
	Boys	Girls	Boys	Girls
2. Letter name recognition	35.8	32.2	46.4	41.4
4. Letter sound knowledge	28.0	22.9	45.0	36.3
6. Non-word reading	12.0	8.5	27.2	18.6
Reading-Rate Fluency Subtest	Grade 3		Grade 5	
	Boys	Girls	Boys	Girls
5. Familiar word reading	26.5	20.1	53.9	39.0
7a. Passage reading	29.8	22.4	62.8	45.6

The final table in this section (Table 18s) further disaggregates the scores by treatment group, grade level, and gender. As with the percent correct scores, the light treatment group scored higher on some of the tasks, which will be statistically corrected at the midline and endline.

TABLE 18S: SINDHI PHONICS AND READING-RATE FLUENCY TASK MEANS BY GROUP, GRADE, AND GENDER

Phonics Fluency Subtest	Full Treatment				Light Treatment			
	Grade 3		Grade 5		Grade 3		Grade 5	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
2. Letter name recognition	34.5	43.7	48.3	52.8	37.2	22.0	44.1	31.9
4. Letter sound knowledge	24.6	27.7	42.6	42.3	31.9	18.6	47.7	31.3
6. Non-word reading	8.7	10.5	23.5	23.0	15.8	6.7	31.6	14.8
Reading-Rate Fluency Subtest	Full Treatment				Light Treatment			
	Grade 3		Grade 5		Grade 3		Grade 5	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
5. Familiar word reading	21.4	26.7	49.4	49.3	31.9	14.4	59.2	30.4
7a. Passage reading	23.8	29.0	55.96	57.7	36.8	16.5	70.9	35.4

Sindhi Questionnaire Findings

Selected results are presented below for the Sindhi questionnaires, including for those characteristics or items that showed significant differences in student scores. Due to the students having the same language (Sindhi), the results were combined for the full and light treatment groups to increase the sample size and more accurately detect effects between the categories. Note that there were some students, teachers, and head teachers who did not respond to certain questionnaire items; they were labeled as missing. The total averages for the summary scores were calculated based on those who responded.

Since these are baseline data, reporting on the full and light treatment groups together will not affect the analyses at midline and endline. We combined the survey data for the groups since some of the questions led to reporting by relatively small categories (e.g., for teacher qualifications) and we wanted to know whether the survey results were associated with the student scores in general.

In addition, since the samples were by treatment group, the results will be generalized to the populations for each group. This will be done prior to the midline. The results will be generalized to by calculating sampling weights, applying the weights to the results, and then generalizing to the population by treatment group. We will also do this for the midline and endline. The current analyses only apply to the sampled districts.

Statistical significance was determined based on *t*-tests for indicators with two categories and analyses of variance for indicators with three or more categories (with post hoc pairwise comparisons). The significance value was set at $p < 0.05$; a 95 percent confidence level.

Sindhi Student Questionnaires

Table 19s has summary scores by students' relative age to their classmates. According to the National Education Policy (2009), the official age of the students at the beginning of the different grade levels of primary education is 6 to 10 years old. Since the baseline took place during the school year, the normal ages for this analysis were set at 8 to 9 years old for grade 3 and 10 to 11 years old for grade 5. The students were placed into three categories: younger than normal age for their grade, normal age, and older than normal age. At grade 3, there were significant differences among all age group comparisons; younger students had significantly lower average summary scores. Conversely, at grade 5, the scores were not significantly different among age groups, eradicating the older-student advantage by grade 5.

TABLE 19S: SINDHI SUMMARY SCORES BY STUDENT AGE

Age Group	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
Younger than normal age	139	25.3%	114	49.1%
Normal age	1,026	31.8%*	877	49.8%
Older than normal age	871	37.2%*	1,026	47.5%
Missing	1	--	21	--
Total	2,038	33.7%	2,038	48.6%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

Table 20s shows the summary scores according to whether the student reads the Quran at home. There were significant differences in both grades favoring students who read the Quran at home.

TABLE 20S: SINDHI SUMMARY SCORES BY READING THE QURAN AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	254	25.2%	212	34.1%
Yes	1,525	35.3%*	1,563	51.0%*
Missing	259	--	263	--
Total	2,038	33.7%	2,038	48.6%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

Table 21s depicts the differences in scores based on whether there is a library at the school. The large number of missing responses is due to the numerous students who responded that they had no knowledge of a school library. The presence of a school library was not indicative of higher EGRA scores for grade 3; however it was significant for grade 5.

TABLE 21S: SINDHI SUMMARY SCORES BY THE PRESENCE OF A LIBRARY AT THE SCHOOL

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,391	34.1%	1,300	48.4%
Yes	438	34.9%	546	51.3%*
Missing	209	--	192	--
Total	2,038	33.7%	2,038	48.6%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

In Tables 22s to 24s, the data showed that the existence of newspapers and magazines generally made a positive difference in reading scores in most cases.

TABLE 22S: SINDHI SUMMARY SCORES BY THE PRESENCE OF NEWSPAPERS AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,377	31.5%	1,192	46.8%
Yes	661	38.1%*	830	51.3%*
Missing	0	--	16	--
Total	2,038	33.7%	2,038	48.6%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

TABLE 23S: SINDHI SUMMARY SCORES BY THE PRESENCE OF MAGAZINES AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,954	33.5%	1,876	47.8%
Yes	84	36.5%	146	58.9%*
Missing	0	--	16	--
Total	2,038	33.7%	2,038	48.6%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

TABLE 24S: SINDHI SUMMARY SCORES BY THE PRESENCE OF BOOKS AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,025	31.9%	1,138	46.4%
Yes	1013	35.5%*	884	51.5%*
Missing	0	--	16	--
Total	2,038	33.7%	2,038	48.6%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

The final set of student questions (in Tables 25s to 27s) pertained to children’s reading habits at home. In general, students had higher reading scores when they practiced reading at home. Having someone read to children did not show significantly higher reading scores, however, having children read to someone else and children reading silently produced significantly higher reading scores than children who did not read aloud or silently at home.

TABLE 25S: SINDHI SUMMARY SCORES BY CHILDREN HAVING SOMEONE READ TO THEM AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,021	33.0%	948	48.1%
Yes	980	34.8%	1,040	49.5%
Missing	37	--	50	--
Total	2,038	33.7%	2,038	48.6%

TABLE 26S: SINDHI SUMMARY SCORES BY CHILDREN READING TO SOMEONE ELSE AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	1,030	32.6%	887	46.2%
Yes	986	35.1%*	1,120	50.6%*
Missing	22	--	31	--
Total	2,038	33.7	2,038	48.6

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

TABLE 27S: SINDHI SUMMARY SCORES BY CHILDREN READING SILENTLY AT HOME

Response	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	539	29.6%	578	40.5%
Yes	1,480	35.4%*	1,441	51.9%*
Missing	19	--	19	--
Total	2,038	33.7%	2,038	48.6%

* Indicates that the performance of the group was significantly higher, $p < 0.05$ level

Sindhi Teacher Questionnaires

With the smaller sample size, the analysis of the teacher questionnaires was limited to descriptive statistics, i.e., no group comparisons. Tables 28s and 29s provide information on Sindhi-medium teacher academic and professional qualifications, neither of which showed consistent patterns in the student scores.

TABLE 28S: SINDHI SUMMARY SCORES BY TEACHER ACADEMIC QUALIFICATION

Academic Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.A./M.Sc./M.Phil.	23	37.4%	35	53.7%
B.A./B.Sc.	56	36.4%	66	51.2%
F.A./F.Sc.	26	32.7%	11	49.2%
Matric	11	32.7%	6	39.2%
Missing	0	--	2	--
Total	116	35.3%	120	51.1%

TABLE 29S: SINDHI SUMMARY SCORES BY TEACHER PROFESSIONAL QUALIFICATION

Professional Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.Ed./M.A.	8	35.9%	21	53.3%
B.Ed.	39	35.1%	44	52.9%
C.T.	5	40.7%	6	48.7%
P.T.C.	62	34.7%	45	50.1%
Missing	2	--	4	--
Total	116	35.2%	120	51.1%

In an analysis of student scores by teacher age and experience, there were no consistent patterns of younger or older teachers, or those with less or more experience, relating to higher or lower student scores (Tables 30s and 31s). Again, small teacher sample sizes made drawing conclusions difficult.

TABLE 30S: SINDHI SUMMARY SCORES BY TEACHER AGE

Age Group in Years	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
40 and less	36	34.9%	57	50.3%
Between 41 and 50	62	36.8%	48	53.4%
51 and more	17	29.2%	11	46.6%
Missing	1	--	4	--
Total	116	35.1%	120	51.2%

TABLE 31S: SINDHI SUMMARY SCORES BY TEACHER EXPERIENCE

Years of Experience	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
10 or less	27	36.3%	34	49.7%
Between 11 and 20	25	33.5%	34	51.6%
Between 21 and 30	56	35.9%	45	52.1%
31 or more	5	25.9%	5	50.5%
Missing	3	--	2	--
Total	116	35.0%	120	51.2%

There were no significant differences in EGRA scores for teachers who did or did not attend in-service training sessions (Table 32s).

TABLE 32S: SINDHI SUMMARY SCORES BY TEACHER IN-SERVICE TRAINING

Frequency of Training	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
None	83	35.3%	82	51.3%
One time	20	31.2%	14	52.0%
Two times	5	36.1%	6	42.2%
Three times	5	54.1%	7	62.3%
Missing	3	--	8	--
Total	116	35.4%	120	50.9%

Sindhi Head Teacher Questionnaires

Similarly to the teachers, the sample size for the head teacher questionnaires was small, so data interpretations should be treated with caution. Tables 33s and 34s show head teacher academic and professional qualifications. In general, the results show that higher teacher academic and professional qualifications show no definitive pattern in student scores.

TABLE 33S: SINDHI SUMMARY SCORES BY HEAD TEACHER ACADEMIC QUALIFICATION

Academic Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.A./M.Sc./M.Phil.	36	34.5%	36	50.1%
B.A./B.Sc.	77	35.4%	77	49.7%
F.A./F.Sc.	21	30.0%	21	45.9%
Matric	6	15.9%	6	26.2%
Missing	0	--	0	--
Total	140	33.4%	140	48.5%

TABLE 34S: SINDHI SUMMARY SCORES BY HEAD TEACHER PROFESSIONAL QUALIFICATION

Professional Qualification	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
M.Ed./M.A.	21	37.2%	21	55.3%
B.Ed.	54	35.4%	54	49.8%
C.T.	5	35.9%	5	46.8%
P.T.C.	56	31.2%	56	46.5%
Missing	4	--	4	--
Total	140	33.9%	140	49.2%

Tables 35s and 36s provide information on head teachers' experience and in-service training. Teachers with five years or less experience had slightly higher reading scores compared to the more experienced head teachers. In terms of in-service training, there was no clear pattern in EGRA scores and the amount of head teacher training. This was the same pattern shown by teachers.

TABLE 35S: SINDHI SUMMARY SCORES BY HEAD TEACHER EXPERIENCE

Years of Experience	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
2 or less	27	37.1%	27	52.5%
3 to 5	20	37.4%	20	52.7%
6 to 10	24	30.1%	24	41.9%
11 or more	63	32.3%	63	47.8%
Missing	6	--	6	--
Total	140	33.6%	140	48.8%

TABLE 36S: SINDHI SUMMARY SCORES BY HEAD TEACHER IN-SERVICE TRAINING

Frequency of Training	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
None	14	34.5%	14	52.6%
1 time	16	36.1%	16	45.1%
2 times	5	35.0%	5	40.0%
More than 2 times	102	33.2%	102	49.2%
Missing	3	--	3	--
Total	140	33.7%	140	48.7%

Tables 37s and 38s provide data on head teachers' support to teachers in reading and the training they received in teaching reading. Only 13 head teachers reported not supporting teachers in reading, therefore the results showing higher reading scores with this support should be interpreted with caution. The EGRA scores were similar for head teachers who did and did not attend training sessions on teaching reading.

TABLE 37S: SINDHI SUMMARY SCORES BY HEAD TEACHER SUPPORT TO TEACHERS IN READING

Support to Teachers	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	13	24.6%	13	35.9%
Yes	126	34.5%	126	50.0%
Missing	1	--	1	--
Total	140	33.7%	140	48.6%

TABLE 38S: SINDHI SUMMARY SCORES BY HEAD TEACHER TRAINING IN TEACHING READING

Support to Teachers	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	59	34.2%	59	49.5%
Yes	74	33.1%	74	48.9%
Missing	5	--	5	--
Total	140	33.7%	140	48.6%

Sindhi School Characteristics

The final section provides information on school characteristics (from the head teacher questionnaires) by student summary scores. As with the teacher and head teacher characteristics, most patterns appeared to be inconclusive (Tables 39s to 42s). The female schools performed better than the male and mixed schools. Schools with libraries (only 12) had significantly better grade 3 scores than those without libraries, but this result must be interpreted with caution due to the low sample size. Better infrastructure seemed to have a positive relationship for grade 5 students ($p > 0.05$).

TABLE 39S: SINDHI SUMMARY SCORES BY SCHOOL GENDER

School Gender	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
Male school	58	29.6%	58	41.7%
Female school	49	37.0%	49	53.3%
Mixed school	31	35.5%	31	53.7%
Missing	2	--	2	--
Total	140	33.5%	140	48.5%

TABLE 40S: SINDHI SUMMARY SCORES BY PTA/SMC/PTSMC/PTC

Parent Teacher Committee	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	3	25.5%	3	30.2%
Yes	135	33.8%	135	49.1%
Missing	2	--	2	--
Total	140	%	140	%

TABLE 41S: SINDHI SUMMARY SCORES BY PRESENCE OF SCHOOL LIBRARY

School Library	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
No	125	32.6%	125	48.2%
Yes	12	41.8%	12	50.4%
Missing	3	--	3	--
Total	140	%	140	%

TABLE 42S: SINDHI SUMMARY SCORES BY INFRASTRUCTURE (DRINKING WATER, ELECTRICITY, TOILETS)

Number of Infrastructures (Water, Electricity, Toilets)	Grade 3		Grade 5	
	n-count	Sum. Score	n-count	Sum. Score
None	17	35.3%	17	43.7%
1	38	30.4%	38	45.3%
2	35	31.9%	35	46.3%
3	50	36.2%	50	53.9*%
Missing	0	--	0	--
Total	140	33.4%	140	48.5%

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

This final chapter provides conclusions and recommendations from the Sindh EGRA baseline. The conclusions are organized according to the two main sections in the report: 1) design and methodology, and 2) findings and results. There are also recommendations based on the instrument development, data collection, data entry, and analysis.

Design and Methodology

1. The design followed USAID evaluation guidelines for a cross-sectional approach. This will allow for an examination of the progress of students in grades 3 and 5 over the life of the PRP and SRP. The province also has two languages, Urdu and Sindhi. In addition, Sindh has two treatment groups: full and light. This will allow for an evaluation of the full treatment effects above and beyond those of the light treatment.
2. The sampling issues were addressed as well as could have been expected. In a limited number of schools, there was an issue of a lack of the requisite number of students per grade level. The actual sample of schools was 100 percent and the actual sample of students reached 92.4 percent of the intended Urdu language sample and 97 percent of the intended Sindhi language sample.
3. The EGRA tests in Urdu and Sindhi were of good quality. The reliability estimates were in the high part of the acceptable range. The task statistics were acceptable, with an appropriate range of p-values and item-total correlations that were at an acceptable level of quality. The characteristics of the tests were such that it should be sensitive to potential progress over time due to project-led interventions. As with any test, there may be ways to improve on the task and item statistics for the midline and endline.
4. The field implementation was successful, though there were difficulties to overcome, including the law and order situation in Karachi and Hyderabad, the logistical challenges with the difficult terrain in some areas of Sindh, and the low actual enrollment of students in some of the Urdu-medium schools. There was a high level of standardization reported by the quality control officers, which they attributed to the effective training process established by the EGRA team. The team paid careful attention to detail in the logistics and test administration, which was reflected in the low error rates in the booklets and in the data entry.

Findings and Results

Sindh schools teaching in Urdu and Sindhi were evaluated separately. Both evaluations will involve two kinds of comparisons: 1) a comparison of full and light treatment groups to determine the effects of full treatment above and beyond that of the light treatment, and 2) a comparison of each group to itself at the baseline, midline, and endline. Please see Figure 1 and the accompanying text for a fuller description of the evaluation design.

Several key findings emerged from the baseline assessment in Sindh. These are as follows:

1. EGRA was administered to a robust sample at each grade level (3 and 5) and in each group (full and light treatment). Test reliabilities were very good, showing that the EGRA tasks and items worked well in measuring reading constructs at both grade levels. The task and item statistics showed that EGRA discriminates well between low- and high-achieving students in both grades. They also showed that there is adequate room for growth by students in each grade level.
2. The EGRA in Urdu and Sindhi were used for the Sindh province. For each language group a total of 140 schools, with 70 schools from each group (full and light treatment), were randomly selected for the baseline.

3. The baseline data were collected in the schools across Sindh province for both Urdu-medium and Sindhi-medium schools. A random sample of male and female schools was selected, followed by a random sample of grades 3 and 5 students within those schools. The results from this sample are presented in this report as a generalized view of the reading levels for students in Sindh in Urdu and Sindhi.
4. The EGRA tools, which have been administered in various forms in over 40 countries, were successfully adapted for use in Pakistan. These included individually administered reading tests for students, along with questionnaires for students, teachers, and head teachers. The Urdu version of the tools was piloted in AJK, ICT, and KP. The Sindhi version was piloted in Sindh.
5. The target sample was 140 Urdu and 140 Sindhi schools. The Urdu assessment was successfully administered in (with a percentage of the target reached in parentheses) 140 schools (100.0 percent), to 3,881 students (92.4 percent), 262 teachers (93.6 percent – with some teachers covering both grades 3 and 5, thus reducing the total number of teachers in the survey), and 140 head teachers (100.0 percent). The Sindhi assessment was successfully administered in 140 schools (100.0 percent) to 4,071 students (96.9 percent), 236 teachers (84.3 percent), and 140 head teachers (100.0 percent). The total number of Sindh province students sampled was 7,952, which is 97 percent of the intended sample.
6. The validity and reliability of the tools was acceptable. Validity was assured through the adaptation process, which involved more than 17 educationists from throughout the country who participated in a workshop in Islamabad. The attention paid to the standardized administration of the tools improved the reliability in the Sindh districts. The data entry and data cleaning process followed international standards. All student data were entered twice into two separate databases. These databases were then compared. All data were reconciled across the two databases and with the assessment booklets. A clean data file was produced for analysis.

Urdu-Medium Schools Analysis

Several key findings emerged from the Urdu baseline assessment in Sindh. These are as follows:

1. EGRA was administered to 1,956 grade 3 and 1,899 grade 5 students. The reliability was very good, with Coefficient Alpha reliability estimates of 0.88 and 0.84 for grades 3 and 5, respectively. The task and item statistics showed that the EGRA can discriminate between low- and high-achieving students in both grades.
2. Grade 3 children did relatively better on the orientation to print, letter name recognition, and passage reading. The lowest scores were in phonics (non-word reading and letter sound knowledge) and comprehension (passage and listening). At grade 5, their best scores were in the two reading tasks (familiar word and passage). They still had relatively low scores in phonics (non-word reading, letter sound knowledge) and comprehension (passage and listening).
3. There was also substantial progression from grade 3 to grade 5 on the summary score (17 points) and on some of the tasks scores – especially in familiar word, and passage reading and comprehension.
4. There were differences between boys and girls on the task and summary scores, but most of these differences were small. For both grades, there was about an 11-point difference on the summary score favoring the girls. Girls had higher scores on familiar word reading, non-word reading, and passage reading and comprehension. In addition girls also had higher fluency scores on the timed reading tasks.
5. The summary score for students from full treatment schools was about six points higher in grade 3 and nine points higher in grade 5 than in light treatment schools. The full and light treatment groups had similar scores for most of the tasks, though the full treatment scores were higher. The largest differences were in the phonics areas of letter sound knowledge and listening comprehension. These differences will be corrected statistically during midline and endline evaluations.

6. Students were timed on five tasks as they read words or passages. These tasks were categorized into phonics fluency (letter name recognition, letter sound knowledge, and non-word reading) and reading-rate fluency (familiar word and passage reading). Students at both grades had lower phonics fluency scores than reading-rate fluency. Passage reading (fluency) was approximately 30 points higher in grade 5 (69 points) than in grade 3 (39 points). Although the passage was designed for grade 3, this difference shows that the reading levels in grade 3 are low, but that children can make substantial progress in the early grades if expectations are high enough and if they are provided with the opportunity to learn. Specifically, the mastery of phonics, such as letter sound knowledge, phonemic awareness, and non-word reading, should help the students become better overall readers. It is clear that these types of knowledge and skills are not receiving an appropriate emphasis in Sindh Urdu-medium schools.
7. The student questionnaire revealed two interesting findings. The first was that having reading materials and opportunities to read at home seemed to have a positive effect on reading outcomes for both the third and fifth grades. Second, third grade summary scores increased with relative age (younger than normal, normal, older than normal age); older children in the grade had higher reading scores. However, by fifth grade that advantage was no longer significant.
8. Student, teacher, and head teacher questionnaire findings were mostly inconclusive, due to small sample sizes and the lack of variation in the scores that were related to their characteristics. For teachers, those who attended one or more in-service trainings had higher scores than those who never attended such trainings. For head teachers, attending one or two in-service trainings, or having received training – specifically in-service training in teaching reading – tended to relate to higher reading scores for students. For the schools, the presence of a library and better infrastructure was associated with better student reading scores.

Sindhi-Medium Schools Analysis

1. EGRA was administered to 2,039 grade 3 and 2,022 grade 5 students. The reliability was very good with Coefficient Alpha reliability estimates of 0.88 and 0.88 for grade 3 and grade 5, respectively. The task and item statistics showed that the EGRA is able to effectively discriminate between low- and high-achieving students in both grades.
2. Grades 3 and 5 showed the same patterns on the reading tasks. Students did relatively better on the orientation to print, listening comprehension, and familiar word and passage reading. The lowest scores were in phonics (non-word reading and letter sound knowledge) and reading comprehension.
3. There was also substantial progression from grade 3 to grade 5 on the summary score (15 points) and on some of the tasks scores – especially in fluency (familiar word reading, passage reading, non-word reading) and passage comprehension.
4. There were differences between boys and girls on the task and summary scores. For the task scores and the summary score, boys had statistically higher ($p < 0.01$) scores than girls. The summary score difference was seven and 12 points, respectively, for grades 3 and 5. The tasks with the largest differences were in familiar word reading and passage reading.
5. Summary scores for students from full treatment schools were about two points higher in grade 3 and five points higher in grade 5. There were no consistent differences among the reading tasks between the full and light treatment groups. The light group had higher scores on some tasks while the full treatment group had higher scores on others. Most of these differences were slight, with the exception of letter name recognition and listening comprehension favoring the full treatment group. These differences will be corrected statistically at the midline and endline.
6. Passage reading (fluency) was approximately 18 points higher in grade 5 than in grade 3. Boys had higher scores than girls. However, no large discrepancies were found between the full and light treatment groups. Although the passage was designed for grade 3, this difference shows that the reading levels in grade 3 are low, but that children can make substantial progress in the early grades if expectations are high enough and if they are provided with the opportunity to learn.

Specifically, the mastery of phonics, such as letter sound knowledge, phonemic awareness, and non-word reading, should help the students become better overall readers. It is clear that these types of knowledge and skills are not receiving an appropriate emphasis in Sindh Sindhi-medium schools.

7. The student questionnaire revealed two interesting findings. The first positive finding was that having reading materials and opportunities to read in the home seemed to have a positive effect on reading outcomes for grade 3 and grade 5. Second, third grade summary scores increased with relative age (younger than normal, normal, older than normal age); older children in the grade had higher reading scores. However, by grade 5 that advantage was no longer significant.
8. Student, teacher, and head teacher questionnaire findings were mostly inconclusive, due to small sample sizes and the lack of variation in the scores that were related to their characteristics. For teachers and head teachers, reading scores were not significantly higher for those who attended in-service trainings. For the schools, better reading scores were associated with the presence of a library (grade 3 only) and better infrastructure (grade 5 only).

Evaluation Recommendations

Given the success of the baseline assessment in Sindh (and in the other provinces), the methods used in 2013 should be repeated as much as possible for the midline and endline assessments in future years. This should be conducted as follows:

1. The EGRA instruments proved to be of high quality, and equivalent versions of those tools should be developed – through trans-adaptation, piloting, and revision – for the midline and endline assessments so that progress can be accurately measured over time.
2. The EGRA items and tasks had good reliability values and covered the low-to-middle difficulty range. At baseline, the reading scores were relatively low for both grades, and show room for growth. In addition, histograms and box plots provided evidence that the tool is expected to measure higher levels of reading that are anticipated due to project-led interventions. Therefore, the baseline data indicates that the EGRA is appropriate for measuring increases in reading ability at midline and endline.
3. The sampling was reasonable in terms of finding a balance between the resources available, the required sample size, and the geographic coverage. It should be maintained in the midline and endline, i.e., keep the same districts and schools, along with the methods at the school level.
4. The systems developed for field data collection should be repeated. The different layers of management, coordination, supervision, and quality control contributed to successful planning, implementation, and problem solving. The QCOs were particularly important in maintaining standards and providing support for the local subcontractors.
5. The data entry process took time to develop, but it eventually proved to be advantageous in terms of having the data entry operators connect to a central server. This better facilitated the two rounds of data entry and the reconciliation process. This system should also be repeated in subsequent data entry activities.
6. The methods for analysis took some time to develop, but it was important to create templates and agree on a methodology due to the volume of analysis and reporting that needs to be done for eight provinces. Again, the investment of time and effort in this process should pay dividends for the midline and endline assessments.
7. Reading proficiency levels should be created to provide educators and other stakeholders with meaningful results. Most parents and educators better understand reading achievement in useful terms or levels, such as emerging, proficient, or advanced, rather than interpreting a percent-correct test score that may differ by test or reading passage difficulty. Education officials are encouraged to select specific EGRA scores to serve as levels of reading proficiency for both

grades. Percent correct for each task, summary score, as well as fluency rates are recommended for this purpose. The baseline EGRA data can be used for establishing these reading proficiency levels.

8. Finally, it may be advisable to add items to the student, teacher, and head teacher questionnaires to collect data on SRP- and PRP-supported interventions so that student scores can be correlated with these indicators.

In general, the Sindh baseline was successful in providing accurate data on which to base decisions for implementation of the SRP and PRP interventions, and also for tracking student reading progress over time. It provides a solid foundation for the midline and endline assessments.

ANNEXES

Annexes 1 to 4 provide additional information on the EGRA baseline. Specifically, the annexes have the following:

Annex 1 gives complete item statistics – p-values (the difficulty of the items) and item-total correlations (the quality of the items) by grade – for the items associated with the various tasks. These are more detailed than the task statistics presented in Chapter 3 of the report. Measurement specialists often request these kinds of item statistics for the purposes of quality control, analysis, and test equating.

Annex 2 provides box plots for the fluency tasks. The box plots are more task-specific than the overall score distributions (histograms) presented in the report. They show the median (middle score), the range (highest and lowest scores), and the distribution of scores (by quartiles) for each task. The task-specific distributions are useful to EGRA specialists who place emphasis on the fluency tasks.

Annex 3 gives two examples of categorizing passage reading fluency scores using performance levels. The categorizations – along with raw scores and scale scores -- are often used to interpret test scores. The first example combines reading speed with comprehension, while the second example only uses reading speed. Each example uses a set of cut-scores for placing the students into performance categories.

Annex 4 provides detailed information on the second example, with results for each category of fluency and each level of comprehension. These data can be used as evidence on the reliability of using a combined measure of fluency and comprehension for setting performance cut-scores. The validity of combining these scores is more of an issue for reading experts.

Annex I: Complete Item Statistics by Grade

Tables A1 (Urdu) and A2 (Sindhi) present item statistics for the untimed tasks, each of which have multiple items. For instance, task 1 (orientation to print) has item statistics for its five items (Q1 to Q5). The timed tasks are lists of letters, sounds, and words, i.e., not items, so it is not necessary to calculate item statistics for them.

Previously, we presented task statistics (Chapter 3, Table 8) with explanations for how they are calculated. These item statistics are calculated in the same way. They show the difficulty and quality of the items. Recall that when constructing a test, we strive for tasks and items that have difficulty values (p-values) that are spread across the range from about 0.05 to 0.90 and quality values (item-total correlations) of at least 0.20. For Urdu, the difficulty values ranged from 0.11 to 0.79 for grade 3 and 0.17 to 0.84 for grade 5, indicating a good range of item difficulties. A total of 22 items for both grades 3 and 5 out of the 23 items for each grade had item-total correlations of at least 0.20, indicating high quality items.

TABLE AIU: URDU COMPLETE ITEM STATISTICS BY GRADE

Task (Subtest)	Item	Grade 3		Grade 5	
		P-Value	Item-Total	P-Value	Item-Total
1. Orientation to print (untimed)	Q1	0.77	0.38	0.75	0.42
	Q2	0.79	0.43	0.84	0.43
	Q3	0.52	0.28	0.52	0.29
	Q4	0.13	0.11	0.22	0.14
	Q5	0.63	0.20	0.78	0.20
2. Letter name recognition (timed)	--				
3. Phonemic awareness (untimed)	Q1	0.51	0.34	0.61	0.40
	Q2	0.30	0.33	0.39	0.42
	Q3	0.33	0.28	0.38	0.35
	Q4	0.25	0.27	0.33	0.37
	Q5	0.37	0.29	0.42	0.39
	Q6	0.41	0.33	0.48	0.40
	Q7	0.25	0.32	0.29	0.39
	Q8	0.26	0.32	0.33	0.40
	Q9	0.27	0.34	0.30	0.36
	Q10	0.50	0.36	0.54	0.42
4. Letter sound knowledge (timed)	--				
5. Familiar word reading (timed)	--				
6. Non-word reading (timed)	--				
7a. Passage reading (timed)	--				
7b. Passage comprehension (untimed)	Q1	0.19	0.58	0.44	0.58
	Q2	0.20	0.55	0.36	0.45
	Q3	0.11	0.46	0.28	0.47
	Q4	0.23	0.60	0.54	0.62
	Q5	0.22	0.59	0.52	0.64
8. Listening comprehension (untimed)	Q1	0.38	0.32	0.56	0.32
	Q2	0.12	0.27	0.17	0.24
	Q3	0.46	0.31	0.76	0.31

For Sindhi, the difficulty values ranged from 0.02 to 0.87 for grade 3 and 0.07 to 0.92 for grade 5, indicating a strong range of item difficulties. A total of 22 items for both grades 3 and 5 out of the 23 items per grade had item-total correlations of at least 0.20, indicating high quality items.

TABLE A1S: SINDHI COMPLETE ITEM STATISTICS BY GRADE

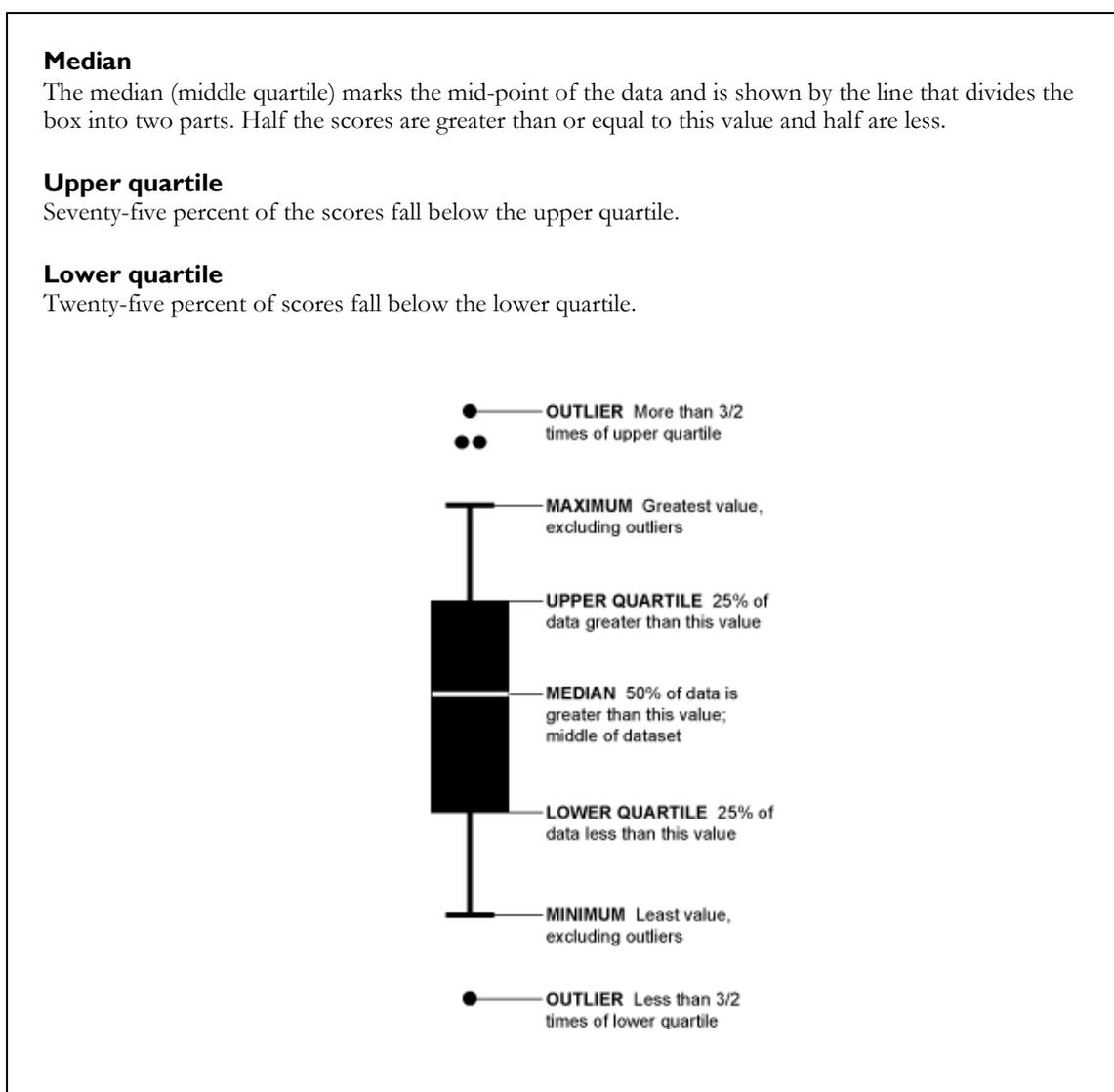
Task (Subtest)	Item	Grade 3		Grade 5	
		P-Value	Item-Total	P-Value	Item-Total
1. Orientation to print (untimed)	Q1	0.82	0.36	0.87	0.29
	Q2	0.87	0.35	0.92	0.33
	Q3	0.45	0.30	0.46	0.25
	Q4	0.07	0.13	0.12	0.13
	Q5	0.63	0.29	0.71	0.26
2. Letter name recognition (timed)	--				
3. Phonemic awareness (untimed)	Q1	0.33	0.29	0.37	0.41
	Q2	0.25	0.24	0.32	0.31
	Q3	0.30	0.33	0.32	0.39
	Q4	0.24	0.22	0.32	0.30
	Q5	0.31	0.28	0.30	0.30
	Q6	0.26	0.33	0.32	0.44
	Q7	0.29	0.34	0.36	0.35
	Q8	0.18	0.23	0.25	0.30
	Q9	0.32	0.35	0.35	0.32
	Q10	0.28	0.31	0.33	0.41
4. Letter sound knowledge (timed)	--				
5. Familiar word reading (timed)	--				
6. Non-word reading (timed)	--				
7a. Passage reading (timed)	--				
7b. Passage comprehension (untimed)	Q1	0.19	0.50	0.35	0.52
	Q2	0.31	0.71	0.57	0.73
	Q3	0.28	0.72	0.50	0.69
	Q4	0.29	0.71	0.57	0.75
	Q5	0.02	0.23	0.07	0.24
8. Listening comprehension (untimed)	Q1	0.39	0.59	0.49	0.61
	Q2	0.45	0.36	0.57	0.39
	Q3	0.40	0.57	0.50	0.62

Annex 2: Box Plots for Phonics and Reading-Rate Fluency Tasks

EGRA places a high emphasis on fluency (timed) tasks. In addition to the descriptive statistics in Table 9 (percent correct scores) and Table 14 (fluency task means), we show box plots for the different fluency tasks. Widely used since their development in the 1960s, box plots are a convenient way for graphically presenting numerical data.

Box plots have two characteristics: 1) central tendency (i.e., the median, or the middle score in the data) and 2) variation (i.e., the range, with scores grouped by quartile). The boxes (which are actually rectangles) represent the two middle quartiles of the scores and the “whiskers” represent the upper and lower quartiles. The small circles on the ends of the whiskers represent outliers. The figure below provides a more detailed explanation for interpreting box plots.

FIGURE A1: UNDERSTANDING BOXPLOTS



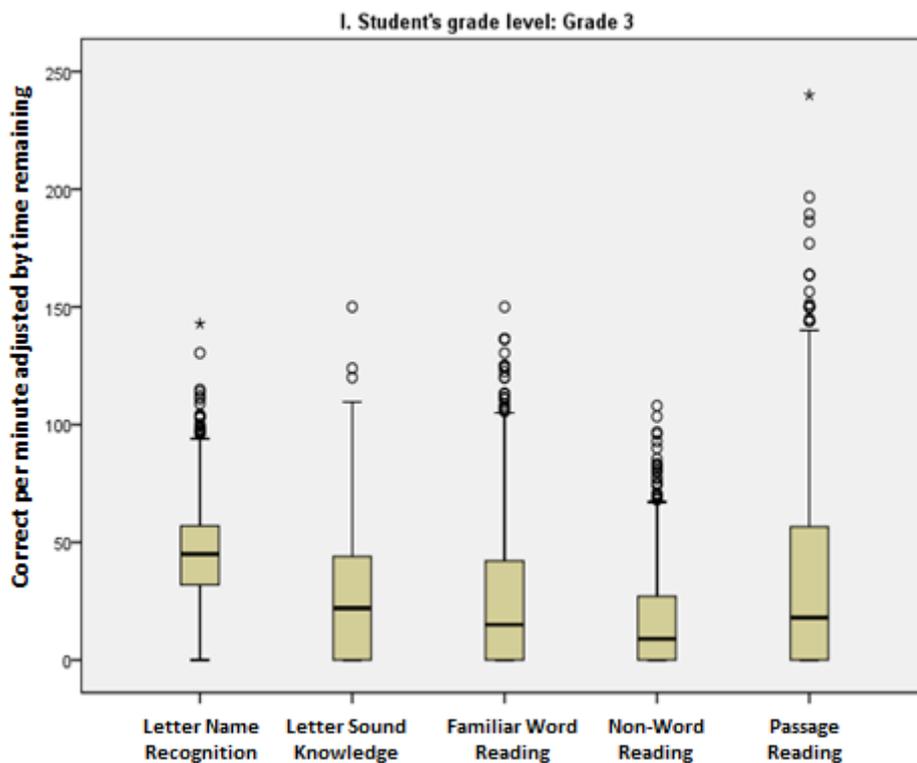
Box plots are presented below (Figures A2u-A3u and A2s-A3s) for the results by language and grade level on the five fluency (timed) tasks: letter name recognition (task 2), letter sound knowledge (task 4), familiar word reading (task 5), non-word reading (task 6), and passage reading (task 7a).

Grade 3, Urdu

For Urdu grade 3, the central tendency (i.e., the median speed, or the line in the middle) for each of the tasks ranged from about 10 (non-word reading) to about 50 (letter name recognition) items per minute. It shows that the students had a better knowledge of letter names than of grapheme-morpheme correspondence.

The variation (i.e., the range of scores, without outliers) for each of the tasks varied from about 70 (non-word reading) to about 150 (passage reading). It shows that the scores were more spread out when reading connected words than sounding out pseudo-words.

FIGURE A2U: PHONICS AND READING-RATE FLUENCY BOX PLOTS FOR GRADE 3, URDU



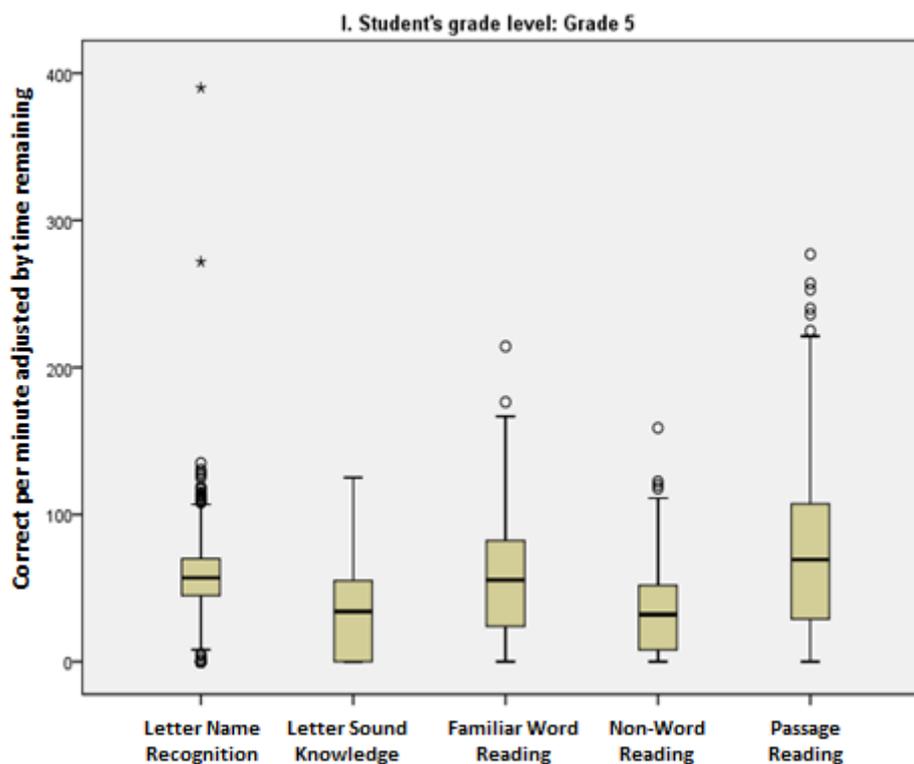
Grade 5, Urdu

For Urdu grade 5, the central tendency (the median speed) for each of the tasks ranged from about 20 (letter sound knowledge) to about 80 (passage reading) items per minute. It shows that the students had more fluency with reading connected words than phonics.

The variation (range of scores) for each of the tasks varied from about 110 (non-word reading) to about 200 (passage reading). It shows that the scores were more spread out when reading connected words than sounding out pseudo-words.

Note also that the medians and the ranges increased from grade 3 to grade 5 for all fluency tasks. Many students are becoming more fluent readers at grade 5, but there are also those students who are either non-readers or very low readers. These children lack of knowledge of letter names, sight words, connected text, and (especially) phonics.

FIGURE A3U: PHONICS AND READING-RATE FLUENCY BOX PLOTS FOR GRADE 5, URDU

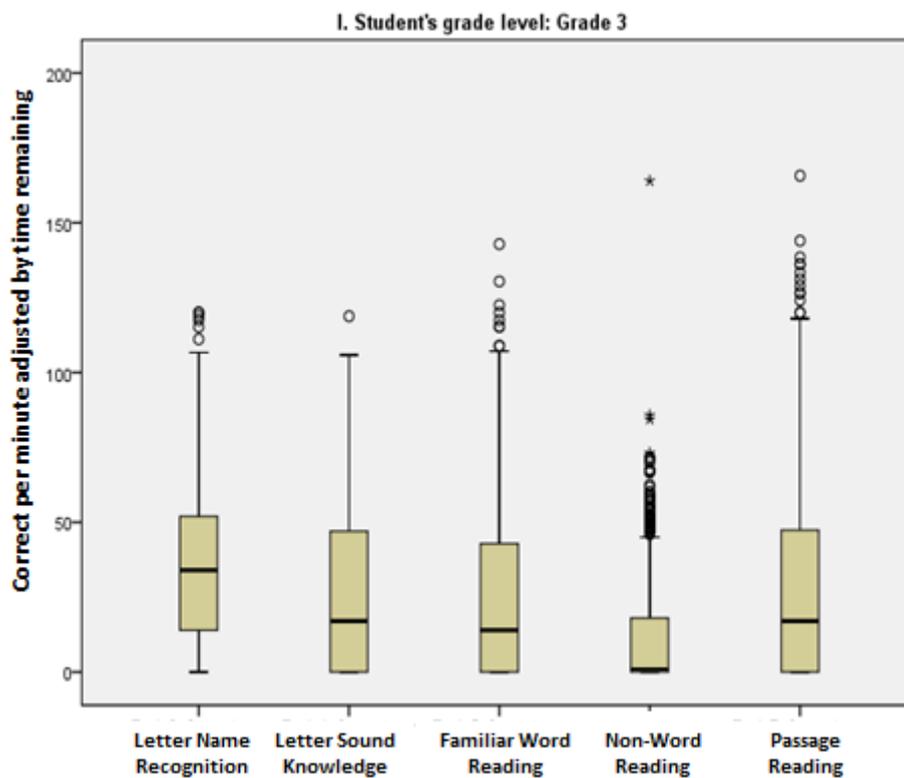


Grade 3, Sindh

For Sindhi grade 3, the central tendency (i.e., the median speed, or the line in the middle) for each of the tasks ranged from about 0 (non-word reading) to about 40 (letter name recognition) items per minute. It shows that the students had much better knowledge of letter names than of grapheme-morpheme correspondence.

The variation (i.e., the range of scores, without outliers) for each of the tasks varied from about 50 (non-word reading) to about 120 (passage reading). It shows that the scores were more spread out when reading words than sounding out pseudo-words.

FIGURE A2S: PHONICS AND READING-RATE FLUENCY BOX PLOTS FOR GRADE 3, SINDHI



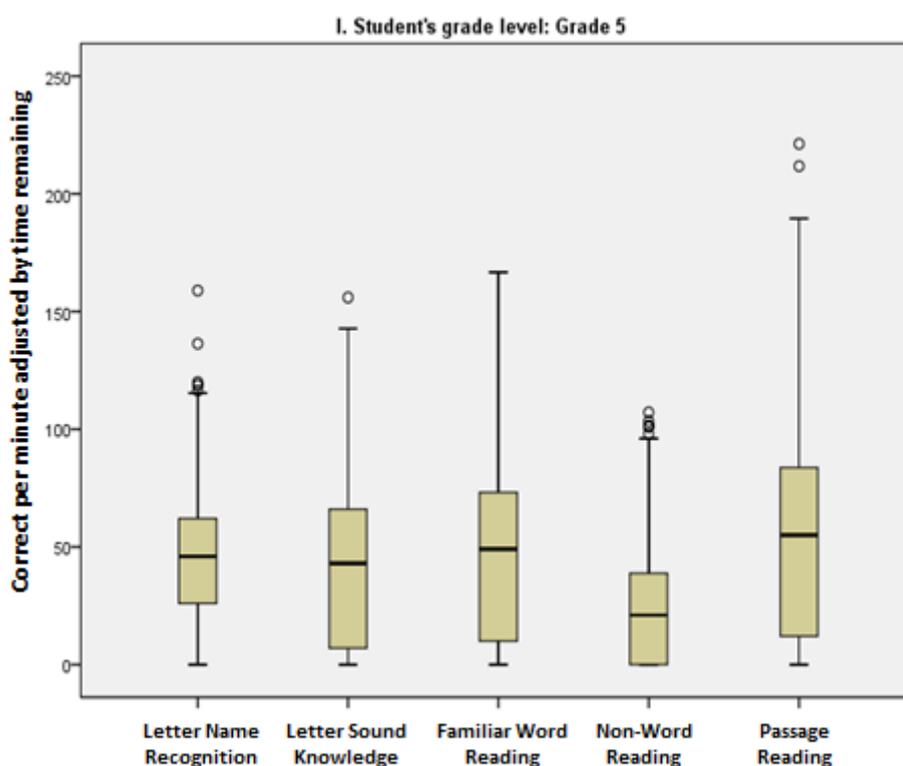
Grade 5, Sindh

For Sindhi grade 5, the central tendency (the median speed) for each of the tasks ranged from about 30 (non-word reading) to about 50 (passage reading) items per minute. It suggests that the students had more fluency reading connected words than conducting grapheme-morpheme correspondence.

The variation (range of scores) for each of the tasks varied from about 100 (non-word reading) to about 200 (passage reading). It shows that the scores were more spread out when reading connected words than sounding out pseudo-words.

Note also that the medians and the ranges increased from grade 3 to grade 5 for all fluency tasks. Many students are becoming more fluent readers at grade 5, but there are also those students who are either non-readers or very low readers. These children lack of knowledge of letter names, sight words, connected text, and (especially) phonics.

FIGURE A3S: PHONICS AND READING-RATE FLUENCY BOX PLOTS FOR GRADE 5, SINDHI



Annex 3: Example of Fluency Score Threshold Calculations

There are different ways of interpreting test scores. Three of the main ways are 1) raw scores (e.g., number correct), 2) scale scores (e.g., percent correct), and 3) percentile scores (e.g., rank in relation to other students). In the report, we presented scores in terms of number correct (for the fluency tasks) and percent correct (for all tasks). We could also calculate the percentile scores for each student, though this is not normally done with EGRA. Note that these kinds of calculations do not change or affect the actual results, but they do involve issues of interpretability.

A fourth main way of interpreting scores is through performance categories, e.g., low, middle, and high. This requires setting cut-scores, or thresholds, to separate the student scores into categories, e.g., two cut-scores lead to three performance categories. The following analysis shows two examples of calculating thresholds for passage reading scores (CWPM), which allows us to place the student scores into different performance categories. Note that performance categories are often accompanied by performance level descriptors (PLDs), which give a text-based explanation of the meaning of the scores in each category. We have not developed PLDs for these examples since 1) the threshold setting is at a preliminary stage and 2) reading specialists with knowledge of local curricula and context generally develop the PLDs.

Fluency using an 80 percent comprehension threshold

In the first example, we used a method that has been suggested by some EGRA specialists. It involves calculating the mean reading speed associated with 80 percent comprehension for those that can read at least one word correctly and then applying it as a fluent cut-score. In other words, the mean reading speed for these students signifies whether the students are fluent readers through using both passage reading speed *and* comprehension in the calculation; the fluent cut-score separates the fluent readers from the non-fluent readers. To establish a second threshold, we again followed the suggested method and used the lowest level of reading (1 CWPM) as the non-fluent cut-score. The two cut-scores resulted in three performance levels: non-readers (low), non-fluent readers (middle), and fluent readers (high).

Urdu

At grade 3, the mean reading speed on the passage reading task (Task 7a) for students who scored 80 percent on the passage comprehension task (Task 7b) was 85.5 (rounded to 86). With this method, 86 CWPM becomes a threshold for grade 3 students who are proficient at passage reading *and* comprehension. At grade 5, the mean speed on the passage reading task (Task 7a) for students who scored 80 percent on the passage comprehension task (Task 7b) was 108.0 (or 108). Then 108 CWPM becomes a threshold for grade 5 students who are proficient at passage reading and comprehension.

The definitions of the three categories in terms of CWPM and the percentages of grades 3 and 5 students in the categories are shown in Table A2u below.

TABLE A2U: URDU THRESHOLDS FOR CWPM WITH 80 PERCENT COMPREHENSION

Category (Performance Level)	Grade 3		Grade 5	
	CWPM	% of Students	CWPM	% of Students
Non-Reader	0	39.8%	0	16.1%
Non-Fluent Reader	1 to 85	50.0%	1 to 107	59.3%
Fluent Reader	86 and above	10.2%	108 and above	24.7%
Total	--	100.0%	--	100.0%

Sindhi

At grade 3, the mean reading speed on the passage reading task (Task 7a) for students who scored 80 percent on the passage comprehension task (Task 7b) was 69.2 (rounded to 69). With this method, 69 CWPM becomes a threshold for grade 3 students who are proficient at passage reading *and* comprehension. At grade 5, the mean speed on the passage reading task (Task 7a) for students who scored 80 percent on the passage comprehension task (Task 7b) was 90.0 (or 90). Then 90 CWPM becomes a threshold for grade 5 students who are proficient at passage reading and comprehension.

The definitions of the three categories in terms of CWPM and the percentages of grades 3 and 5 students in the categories are shown in Table A2s below.

TABLE A2S: SINDHI THRESHOLDS FOR CWPM WITH 80 PERCENT COMPREHENSION

Category (Performance Level)	Grade 3		Grade 5	
	CWPM	% of Students	CWPM	% of Students
Non-Reader	0	38.5%	0	20.8%
Non-Fluent Reader	1 to 68	52.7%	1 to 89	58.0%
Fluent Reader	69 and above	8.8%	90 and above	21.2%
Total	--	100.0%	--	100.0%

Note that for both languages the majority of the students are in the middle category at each grade level. This is due the large range of scores for this category, i.e., from the students who score just above non-readers to those who score just below fluent readers are in the non-fluent reader (middle) category.

Fluency using fixed interval thresholds

In the second example, we used fixed intervals of CWPM for the performance levels. This reduced the problem of having a large range of students in the middle category by creating early reader and intermediate reader categories. It also follows common practice when setting performance categories of having between three and five levels for student scores. We used an interval of 40 CWPM to produce five performance levels, along with a category for the non-readers. The five levels were: non-readers (0 CWPM); early readers (1-40 CWPM); intermediate readers (41-80 CWPM); fluent readers (81-120 CWPM); and advanced readers (121 and above CWPM). Results by language are displayed below.

TABLE A3U: URDU THRESHOLDS FOR CWPM WITH FIXED INTERVALS

Category (Performance Level)	CWPM	% of Students	
		Grade 3	Grade 5
Non-Reader	0	39.8%	16.1%
Early Reader	1 to 40	25%	13.9%
Intermediate Reader	41 to 80	23.3%	27.8%
Fluent Reader	81 to 120	9.2%	24.6%
Advanced Reader	121 and above	2.7%	17.7%
Total	--	100.0%	100.0%

TABLE A4S: SINDHI THRESHOLDS FOR CWPM WITH FIXED INTERVALS

Category (Performance Level)	CWPM	% of Students	
		Grade 3	Grade 5
Non-Reader	0	38.5%	20.8%
Early Reader	1 to 40	30.7%	18.9%
Intermediate Reader	41 to 80	26.3%	32.6%
Fluent Reader	81 to 120	3.9%	21.1%
Advanced Reader	121 and above	0.6%	6.6%
Total	--	100.0%	100.0%

At both grades 3 and 5, the fixed interval method allowed for more distribution of the scores across the categories. We can also see a shift in percentages of students in each category from grade 3 to grade 5 in each language; the performance categories allow for a score interpretation showing that students are improving across the grade levels, with more scores in the lower categories at grade 3 and more scores in the higher categories at grade 5.

Remarks

While it is possible to use such percentages to set cut-scores for interpretation purposes at the baseline, midline and endline, this analysis should be taken as preliminary. For instance, more well-known and accepted method of setting thresholds – which is commonly called “standard setting” by measurement specialists – involve holding a workshop with local reading experts to set the cut-scores according to the experts’ conceptions of what students should know and be able to do in order to be classified into a performance category. There are several well-known methods, e.g., Angoff and Bookmark, which have been judged as valid and reliable for this purpose.⁴ Further discussions on setting thresholds involving local reading experts are recommended.

⁴ References include: Zieky, M. & Perie, M. (2006). *A primer on setting cut-scores on tests of educational achievement*. Princeton, New Jersey: Educational Testing Service; Cizek, G. (1996). *Standard-setting guidelines*. Educational Measurement: Issues and Practices, Spring 1996, p. 13-21; Cizek, G., Bunch, M., & Koons, H. (2004). *Setting performance standards: Contemporary methods*. Educational Measurement: Issues and Practices, Winter 2004.

Annex 4: Distribution of Reading Fluency and Comprehension Scores using Fixed Intervals

In this last annex, we provide more information on the relationship between reading fluency (speed) and comprehension using information from the fixed interval method. While the data show a positive relationship between speed and comprehension, there are sizeable numbers of “fluent” readers with little comprehension. Our conclusion is that setting a cut-score using a less than reliable indicator, such as the mean speed of students with 80 percent comprehension (i.e., using *both* speed and comprehension), can be problematic. The result is categorizing some students as fluent readers who in fact, according to the definition, are not, i.e., they have high reading speed but low comprehension. It may be better to set thresholds based solely on a single indicator – reading speed – rather than mixing it with comprehension.

The figures and tables below (Tables A4u-A5u, A4s-A5s, and Figures A4u-A5u, A4s-A5s) expand on the data in Tables A2u and A2s. They show the results for reading fluency (in terms of speed) by comprehension level for grades 3 and 5. We used the categories based on intervals of 40 CWPM, along with a category for the CWPM non-readers (0 CWPM). Comprehension levels were calculated in terms of percent correct scores (e.g., 20 percent is the same as correctly answering one question out of five total questions). For instance, at grade 3 in Urdu, 100 percent of the non-readers have 0 percent comprehension, 34 percent of the advanced readers have 80 percent comprehension and 47 percent (34% + 13%) have at least 80 percent comprehension.

Urdu

TABLE A5U: GRADE 3 READING FLUENCY AND COMPREHENSION, URDU

Category (Performance Level)	CWPM	% of Students by Comprehension Level						
		0%	20%	40%	60%	80%	100%	Total
Non-Reader	0	100%	0%	0%	0%	0%	0%	100%
Early Reader	1 to 40	67%	18%	11%	3%	1%	0%	100%
Intermediate Reader	41 to 80	16%	20%	26%	18%	13%	7%	100%
Fluent Reader	81 to 120	8%	10%	30%	22%	20%	11%	100%
Advanced Reader	121 and above	2%	2%	13%	36%	34%	13%	100%

FIGURE A4U: GRADE 3 READING FLUENCY AND COMPREHENSION, URDU

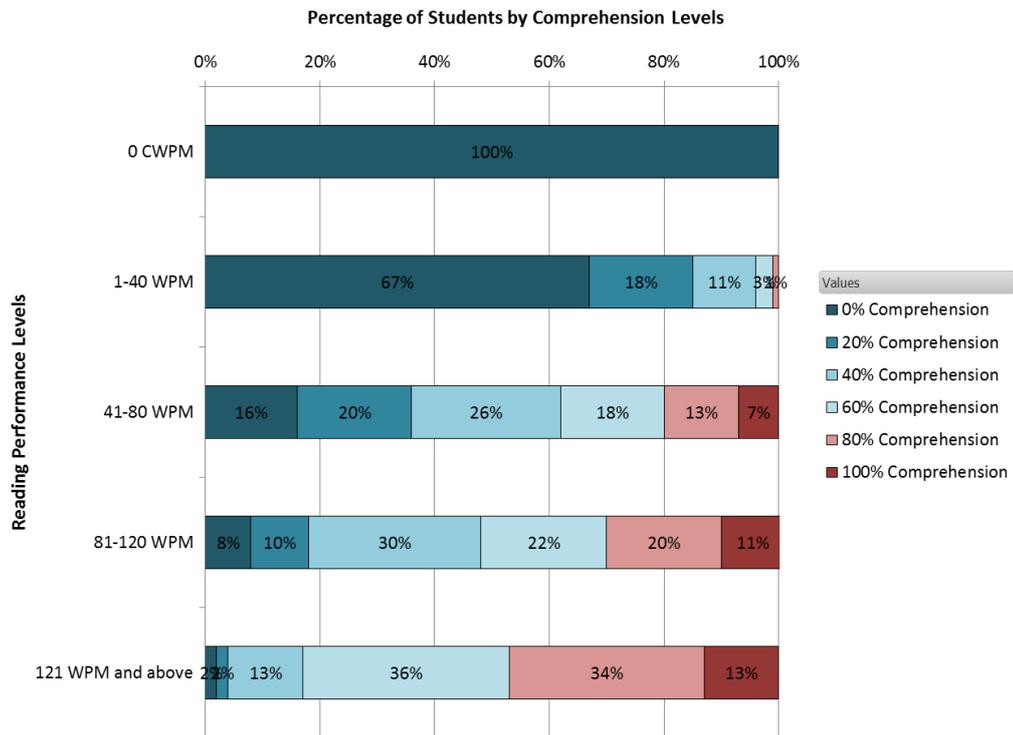
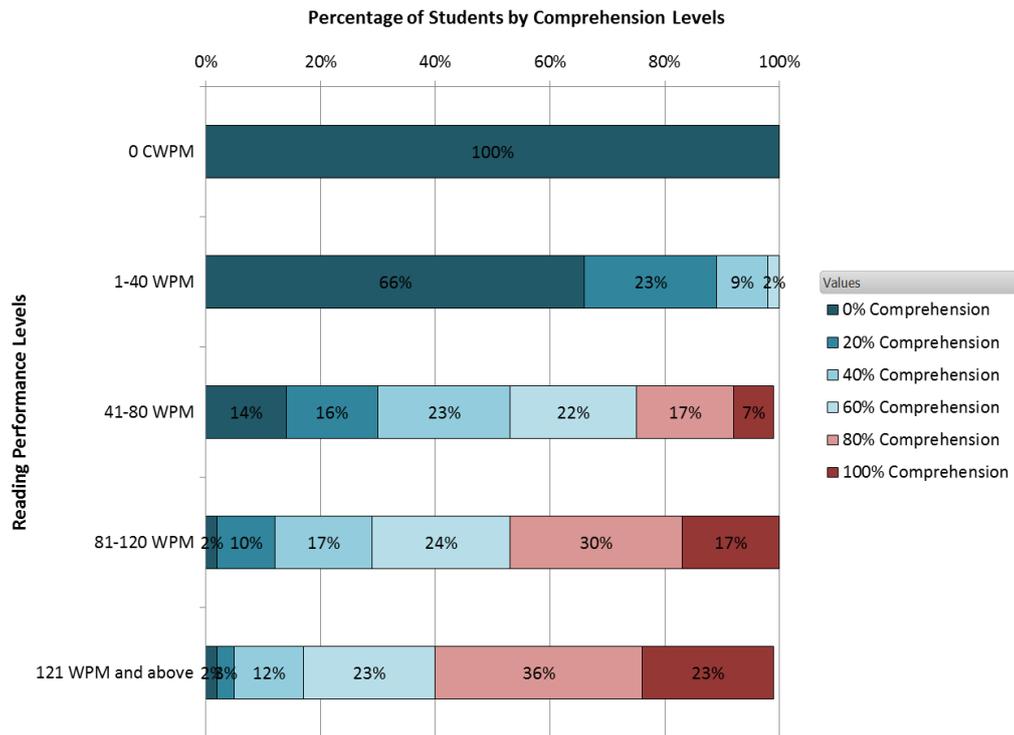


TABLE A6U: GRADE 5 READING FLUENCY AND COMPREHENSION, URDU

Category (Performance Level)	CWPM	% of Students by Comprehension Level						Total
		0%	20%	40%	60%	80%	100%	
Non-Reader	0	100%	0%	0%	0%	0%	0%	100%
Early Reader	1 to 40	66%	23%	9%	2%	0%	0%	100%
Intermediate Reader	41 to 80	14%	16%	23%	22%	17%	7%	100%
Fluent Reader	81 to 120	2%	10%	17%	24%	30%	17%	100%
Advanced Reader	121 and above	2%	3%	12%	23%	36%	23%	100%

FIGURE A5U: GRADE 5 READING FLUENCY AND COMPREHENSION, URDU



Sindhi

TABLE A7S: GRADE 3 READING FLUENCY AND COMPREHENSION, SINDHI

Category (Performance Level)	CWPM	% of Students by Comprehension Level						Total
		0%	20%	40%	60%	80%	100%	
Non-Reader	0	100%	0%	0%	0%	0%	0%	100%
Early Reader	1 to 40	57%	19%	12%	11%	1%	0%	100%
Intermediate Reader	41 to 80	11%	12%	18%	32%	24%	3%	100%
Fluent Reader	81 to 120	1%	8%	11%	34%	42%	4%	100%
Advanced Reader	121 and above	0%	0%	0%	25%	42%	33%	100%

FIGURE A4S: GRADE 3 READING FLUENCY AND COMPREHENSION, SINDHI

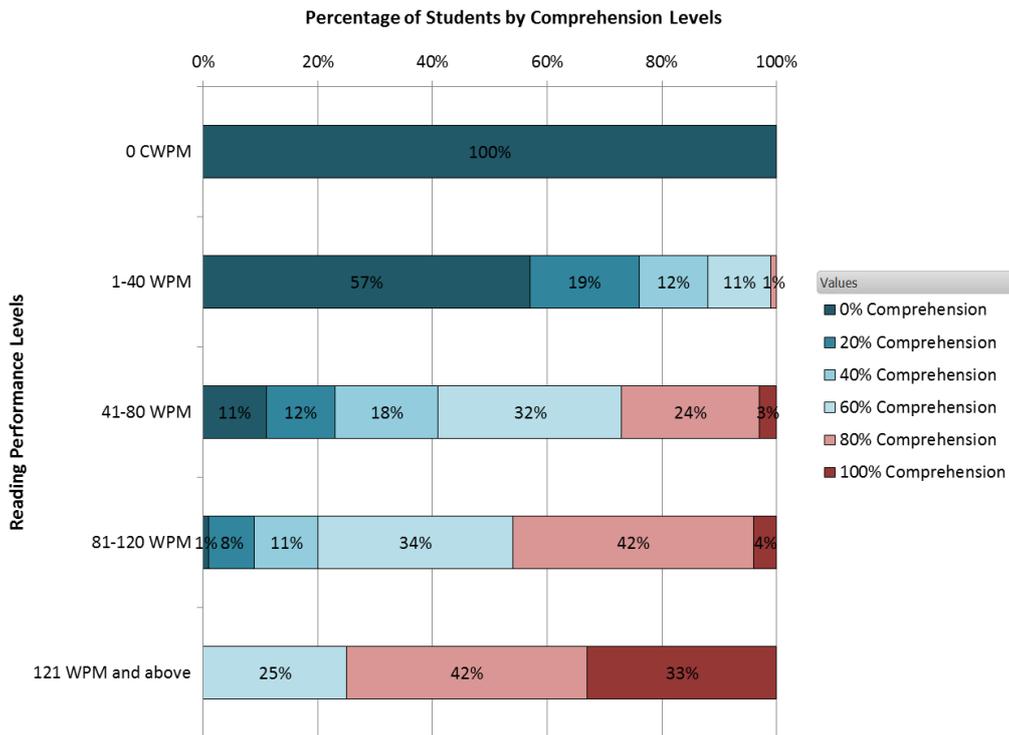
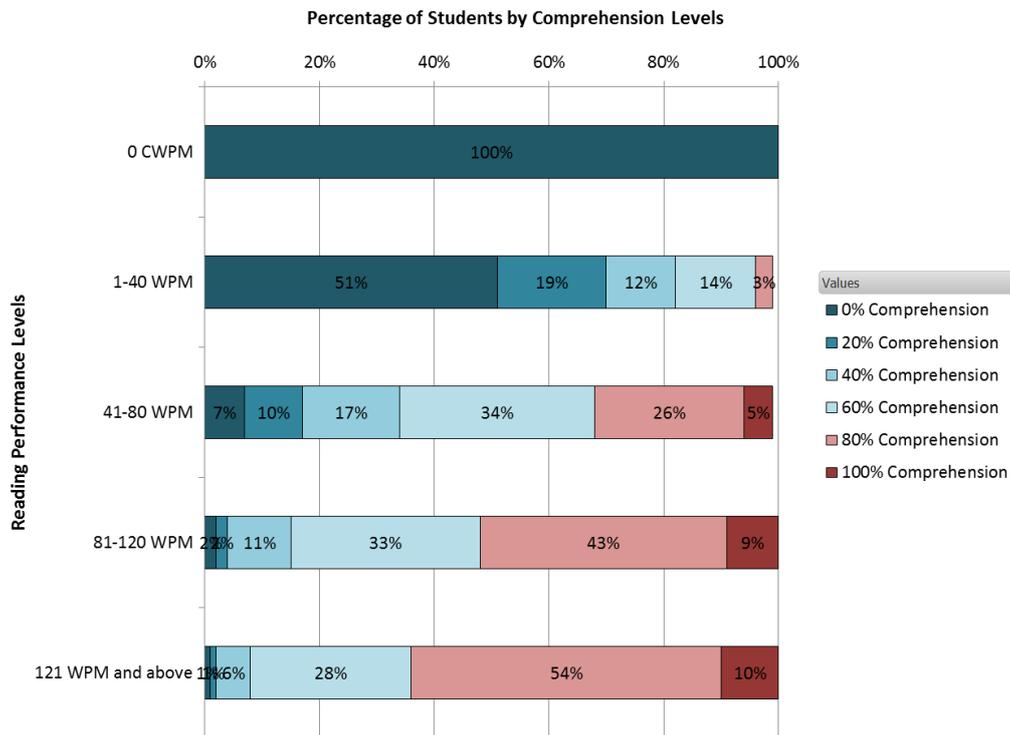


TABLE A8S: GRADE 5 READING FLUENCY AND COMPREHENSION, SINDHI

Category (Performance Level)	CWPM	% of Students by Comprehension Level						Total
		0%	20%	40%	60%	80%	100%	
Non-Reader	0	100%	0%	0%	0%	0%	0%	100%
Early Reader	1 to 40	51%	19%	12%	14%	3%	0%	100%
Intermediate Reader	41 to 80	7%	10%	17%	34%	26%	5%	100%
Fluent Reader	81 to 120	2%	2%	11%	33%	43%	9%	100%
Advanced Reader	121 and above	1%	1%	6%	28%	54%	10%	100%

FIGURE A55: GRADE 5 READING FLUENCY AND COMPREHENSION, SINDHI



Main Results

The main results for the categories of reading speed (from non-readers to advanced readers) in relation to comprehension levels (from 0 percent to 100 percent) for grades 3 and 5 are summarized as follows:

- Non-Readers (0 CWPM) – All of the non-readers (in grades 3 and 5 and Urdu and Sindhi) had 0 percent comprehension.
- Early Readers (1-40 CWPM) – Most of the early readers (67 percent at grade 3 and 66 percent at grade 5 in Urdu and 57 percent at grade 3 and 51 percent at grade 5 in Sindhi) had 0 percent comprehension. Almost none of them achieved 80 percent comprehension.
- Intermediate Readers (41-80 CWPM) – About one in every ten intermediate readers had 0 percent comprehension (16 percent in grade 3 and 14 percent in grade 5 in Urdu and 11 percent for grade 3 and 7 percent for grade 5 in Sindhi). More than two out of every five achieved at least 80 percent comprehension (20 percent at grades 3 and 24 percent at grade 5 in Urdu and 27 percent at grade 3 and 31 at grade 5 in Sindhi).
- Fluent Readers (81-120 CWPM) – A minority of the fluent readers had 0 percent comprehension (in Urdu, 8 percent at grade 3 and 2 percent at grade 5 and in Sindhi 1 percent in grade 3 and 2 percent in grade 5).
- Advanced Readers (121 CWPM and above) – A small percentage of the advanced readers had 0 percent comprehension. For grade 3 in Urdu about half (47 percent) achieved 80 comprehension. Most students in the other grades and languages achieved at least 80 percent comprehension. (59 percent at grade 5 in Urdu and 75 percent at grade 3 and 64 percent in grade 5 in Sindhi).

The key point from the data is that most of the fluent and advanced readers – at both grade levels – did not reach 80 percent comprehension. Setting a threshold under the assumption that fluent readers (in terms of speed) have a high level of comprehension can be misleading. Conversely, using a single indicator, i.e., reading speed, to set thresholds can be a more reliable way of interpreting the results.