

SYRIA EDUCATION PROGRAMME



Syria Education Programme Learning Assessment 2019

MAY 2020

Disclaimer

This document has been redacted to protect the individuals involved in the Syria Education Programme. All names of people and locations have either been altered or removed, as has any information that may identify people or locations.

Project Description

The Syria Education Programme (SEP), also known as Manahel, provides access to safe, inclusive, and quality learning opportunities. Across its lifecycle the project will reach half a million primary-school-age children in Syria.

SEP enables teachers, school staff, and education sector leadership to deliver quality education. In response to the ever-changing landscape of conflict and crisis in Syria, SEP invests in and applies research to respond to the educational, psychological, and protection needs of Syria's children.

From the specialised requirements of disabled children to the psychological demands of childhood within conflict, students' needs are as diverse as they are urgent. SEP takes a broad and nuanced approach to the myriad needs of individual children and groups. By broadening educational access, promoting a safe and secure environment, and creating quality learning opportunities, SEP strives to meet children's holistic needs at scale.



The Syria Education Programme is funded by UK aid from the UK government.

Table of Contents

Executive Summary	6
Introduction	24
Methodology	29
Findings	39
Results and Recommendations by Skills	65
Conclusions and Recommendations	82
Annexes	88

LIST OF ACRONYMS

AO	Access Only Intervention
CAPPM	Correct addition problems per minute
CLSPM	Correct letter sounds per minute
CNPM	Correct numbers per minute
CNWPM	Correct nonwords per minute
CSPPM	Correct subtraction problems per minute
CWPM	Correct words per minute
DFID	United Kingdom's Department for International Development
ED	Education Directorates
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EU	European Union
FCDO	United Kingdom's Foreign, Commonwealth & Development Office
ORF	Oral reading fluency
QE	Quality Education Intervention
SIG	Syrian Interim Government
STS	School-to-School International
TOT	Training-of-Trainers
USAID	United States Agency for International Development

List of Annexes

Annex A: Evaluation Framework	88
Annex B: Assessment and Survey Tools	91
Annex C: EGRA Tool Changes	109
Annex D: Proficiency Band Results	111
Annex E: Mean Fluency Rates and Accuracy Scores	118
Annex F: Zero Scores	131
Annex G: 2017 versus 2019 Analysis	136
Annex H: Manahel Levels Analysis	139
Annex I: Accuracy Scores for Low, Moderate, and High Performers	142
Annex J: Correlations Between Subtasks	155
Annex K: Bibliography	159

List of Figures

Figure 1. Percentage of Students by Reading Proficiency Level	10
Figure 2. Percentage of Students by Reading Proficiency Levels by Gender	10
Figure 3. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Gender	11
Figure 4. Reading Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender	14
Figure 5. Mathematics Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender	14
Figure 6. Mathematics Percentage of Items Correct (Accuracy Scores) in AO Schools by Gender	15
Figure 7. Percentage of Students in Province A and Province B by Reading Proficiency Level	15
Figure 8. Reading Percentage of Items Correct (Accuracy Scores) in Province A by Gender	16
Figure 9. Reading Percentage of Items Correct (Accuracy Scores) in Province B by Gender	17
Figure 10. Mathematics Percentage of Items Correct (Accuracy Scores) in Province A by Gender	17
Figure 11. Mathematics Percentage of Items Correct (Accuracy Scores) in Province B by Gender	18
Figure 12. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Displacement Group	21
Figure 14. Percentage of Students by Reading Proficiency Level	41
Figure 15. Percentage of Students by Reading Proficiency Levels by Gender	41
Figure 16. Reading Items Correct Per Minute (Fluency Rates) by Subtask and Gender	41
Figure 17. Reading Percentage of Items Correct (Accuracy Scores) by Subtask and Gender	42
Figure 18. Mathematics Items Correct Per Minute (Fluency Rates) by Subtask and Gender	43
Figure 19. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Gender	43
Figure 20. Percentage of Students by Reading Proficiency Level and School Type	45
Figure 21. Percentage of Non-Readers by Wave in QE Schools	45
Figure 22. Reading Items Correct Per Minute (Fluency Rates) by Subtask and School Type	46
Figure 23. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and School Type	46
Figure 24. Mathematics Items Correct Per Minute (Fluency Rates) by Subtask and School Type	47
Figure 25. Percentage of Students in QE Schools by Reading Proficiency Level and Gender	48
Figure 26. Reading Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender	48
Figure 27. Mathematics Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender	49

Figure 28. Reading and Mathematics Items Correct Per Minute (Fluency Rates) QE Schools by Subtask and Gender	49
Figure 29. Percentage of Students in AO Schools by Reading Proficiency Level and Gender	50
Figure 30. Reading Percentage of Items Correct (Accuracy Scores) in AO Schools by Gender	50
Figure 31. Mathematics Percentage of Items Correct (Accuracy Scores) in AO Schools by Gender	51
Figure 32. Reading and Mathematics Items Correct Per Minute (Fluency Rates) AO Schools by Subtask and Gender	51
Figure 33. Percentage of Students in Province A and Province B by Reading Proficiency Level.	52
Figure 34. Reading Percentage of Items Correct (Accuracy Scores) by Subtask and Province	52
Figure 35. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Province	53
Figure 36. Percentage of Province A Students by Reading Proficiency Level and Gender	54
Figure 37. Reading Percentage of Items Correct (Accuracy Scores) in Province A by Gender	54
Figure 38. Mathematics Percentage of Items Correct (Accuracy Scores) in Province A by Gender	55
Figure 39. Reading and Mathematics Items Correct Per Minute (Fluency Rates) for Province A by Subtask and Gender	55
Figure 40. Percentage of Province B Students by Reading Proficiency Level and Gender	56
Figure 41. Reading Percentage of Items Correct (Accuracy Scores) in Province B by Gender	56
Figure 42. Mathematics Percentage of Items Correct (Accuracy Scores) in Province B by Gender	56
Figure 43. Reading and Mathematics Items Correct Per Minute (Fluency Rates) for Province B by Subtask and Gender	57
Figure 44. Manahel Reading Levels	57
Figure 45. Number of Students by Times They Moved Due to the War	61
Figure 46. Percentage of Students by Reading Level, 1-5 Moves	62
Figure 47. Percentage of Students by Reading Level, 6+ Moves	62
Figure 48. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Displacement Rate	62
Figure 49. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Attendance Rate	63
Figure 50. Distribution of Letter Sound Identification Accuracy Scores and Fluency Rates for Beginning and Proficient Readers	67
Figure 51. Distribution of Letter Sound Identification Accuracy Scores by Gender	67
Figure 52. Distribution of Oral Reading Fluency Accuracy Scores and Fluency Rates for Beginning and Proficient Readers	67
Figure 53. Distribution of Oral Reading Fluency Accuracy Scores by Gender	68
Figure 54. Distribution of Listening Comprehension Accuracy Scores for Beginning and Proficient Readers	70
Figure 55. Distribution of Listening Comprehension Accuracy Scores by Gender	70
Figure 56. Distribution of Reading Comprehension Accuracy Scores for Beginning and Proficient Readers	71
Figure 57. Distribution of Reading Comprehension Accuracy Scores by Gender	71
Figure 58. Distribution of Number Recognition Accuracy Scores and Fluency Rates for Beginning and Proficient Readers	73
Figure 59. Distribution of Number Recognition Accuracy Scores by Gender	73
Figure 60. Distribution of Number Discrimination Accuracy Scores for Beginning and Proficient Readers	73
Figure 61. Distribution of Number Discrimination Accuracy Scores by Gender	73
Figure 62. Distribution of Missing Number Accuracy Scores for Beginning and Proficient Readers	74
Figure 63. Distribution of Missing Number Accuracy Scores by Gender	74
Figure 64. Distribution of Addition 1 Accuracy Scores and Fluency Rates for Beginning and Proficient Readers	76
Figure 65. Distribution of Addition 1 Accuracy Scores by Gender	76

Figure 66. Distribution of Subtraction 1 Accuracy Scores and Fluency Rates for Beginning and Proficient Readers	77
Figure 67. Distribution of Subtraction 1 Accuracy Scores by Gender	77
Figure 68. Distribution of Addition Level 2 Accuracy Scores for Beginning and Proficient Readers	78
Figure 69. Distribution of Addition 2 Accuracy Scores by Gender	78
Figure 70. Distribution of Subtraction Level 2 Accuracy Scores for Beginning and Proficient Readers	78
Figure 71. Distribution of Subtraction 2 Accuracy Scores by Gender	78
Figure 72. Distribution of Word Problems Accuracy Scores for Beginning and Proficient Readers	80
Figure 73. Distribution of Word Problems Accuracy Scores by Gender	80
Figure 74. Percentage of Students by Reading Proficiency Level and Year	136
Figure 75. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask	137

List of Tables

Table 1. Target and Final Sample	7
Table 2. Summary of Key Conclusions and Recommendations	21
Table 3. Manahel Intervention Levels	26
Table 4. Description of Manahel Intervention Waves	27
Table 5. Target and Final Sample	30
Table 6. Summary of EGRA and EGMA Subtasks	33
Table 7. Manahel Reading Levels Mapped to EGRA Subtasks and Proportion of Students Mastering	58
Table 8. Student Profiles for Mechanics of Reading Skills by Performance Category	68
Table 9. Student Profiles for Understanding Skills by Performance Category	70
Table 10. Student Profiles for Reading Comprehension Skills by Performance Category	72
Table 11. Student Profiles for Whole Number Skills by Performance Category	75
Table 12. Student Profiles for Operations Skills by Performance Category	79
Table 13. Student Profiles for Real-World Problems Skills by Performance Category	81
Table 14. Summary of Key Conclusions and Recommendations	85
Table 15. Proposed Evaluation Years and Grade	89
Table 16. Changes between 2017 and 2019 Assessment	138
Table C.1. Summary of EGRA Item-Level Changes	109
Table C.2. Details of Nonword Reading Subtask Item-Level Changes	109
Table D.1. Proportion of Students Per Proficiency Band, by Gender, Province, and School Type	111
Table D.2. Proportion of Students Per Proficiency Band, by Province and Gender	111
Table D.3. Proportion of Students Per Proficiency Band, by School Type and Gender	112
Table D.4. Proportion of Students Per Proficiency Band, by Province and School Type	112
Table D.5. Proportion of Students Per Proficiency Band by District	112
Table D.6. Proportion of Students Per Proficiency Band, by District and Gender	113
Table D.7. Proportion of Students Per Proficiency Band, by School Type and Wave	113
Table D.8. Proportion of Students Per Proficiency Band, by Number of Moves and New Schools Due to Conflict, Attendance in Past Week, and Province	114
Table D.9. Proportion of Students Per Proficiency Band, by District and Number of Moves Due to Conflict	115
Table D.10. Proportion of Students Per Proficiency Band, by Feelings of Safety	115

Table D.11. Proportion of Students Per Proficiency Band, by Province and Time to Travel to School	116
Table D.12. Proportion of Students Per Proficiency Band, by Feelings of Hunger and Tiredness	116
Table D.13. Proportion of Students Per Proficiency Band between Provinces	116
Table D.14. Proportion of Students Per Proficiency Band between School Types	117
Table D.15. Proportion of Students Per Proficiency Band between Districts	117
Table D.16. Proportion of Students Per Proficiency Band, comparing 2017 and 2019 Results by Province and Gender	117
Table E.1. EGRA Fluency Rates and Accuracy Scores, by Province, Gender and School Type	118
Table E.2. EGMA Fluency Rates and Accuracy Scores, by Province, Gender and School Type	119
Table E.3. EGRA Fluency Rates and Accuracy Scores, by Province and Gender	120
Table E.4. EGMA Fluency Rates and Accuracy Scores, by Province and Gender	121
Table E.5. EGRA Fluency Rates and Accuracy Scores, by School Type and Gender	122
Table E.6. EGMA Fluency Rates and Accuracy Scores, by Province and Gender	123
Table E.7. EGRA Fluency Rates and Accuracy Scores, by District	125
Table E.8. EGMA Fluency Rates and Accuracy Scores, by District	125
Table E.9. EGRA Fluency Rates and Accuracy Scores, by District and Gender	127
Table E.10. EGMA Fluency Rates and Accuracy Scores, by District and Gender	128
Table E.11. EGMA Fluency Rates and Accuracy Scores, Comparing 2017 and 2019 by Province and Gender	129
Table F.1. Proportion of EGRA and EGMA Zero Scores by Subgroups	131
Table F.2. Proportion of EGRA and EGMA Zero Scores by Province and Gender	132
Table F.3. Proportion of EGRA and EGMA Zero Scores, by School Type and Gender	133
Table F.4. Proportion of EGRA and EGMA Zero Scores by District	134
Table F.5. Proportion of EGRA and EGMA Zero Scores by District and Gender	135
Table 16. Changes between 2017 and 2019 Assessment	138
Table H.1. Summary of Results by Manahel Level and Subgroup	139
Table I.1. Proportion of Students in Low, Moderate and High-Performing Groups for Reading Mechanics, by Subtask and Subgroup	142
Table I.2. Proportion of Students in Low, Moderate and High-Performing Groups for Understanding and Comprehension, by Subtask and Subgroup	144
Table I.3. Proportion of Students in Low, Moderate and High-Performing Groups for Whole Numbers, by Subtask and Subgroup	145
Table I.4. Proportion of Students in Low, Moderate and High-Performing Groups for Addition, by Subtask and Subgroup	147
Table I.5. Proportion of Students in Low, Moderate and High-Performing Groups for Subtraction and Word Problems, by Subtask and Subgroup	149
Table I.6. Comparisons of Column Proportions for EGRA Subtasks, by Subgroup b, c	151
Table I.7. Comparisons of Column Proportions for EGMA Subtasks, by Subgroup b, c	153
Table J.1. Score Correlations with EGRA and EGMA Fluency Rates	155
Table J.2. Score Correlations with EGRA Accuracy Scores	156
Table J.3. Score Correlations with EGMA Accuracy Scores	157

Executive Summary

Introduction

This report presents the results of a study of learning outcomes for students served by the Syria Education Programme, also known as Manahel, in Northwest Syria. Manahel is a three-year project funded by the United Kingdom's Foreign, Commonwealth and Development Office (FCDO) and implemented by Chemonics International. The project seeks to expand access to education and strengthen the capacity of educational actors to improve the quality of learning in Northwest Syria.

As a follow-on to the previous iteration of the project which ran from 2014-2018, Manahel focuses on four core components:

1. Stipend payments to teachers and education staff along with a teacher payment and attendance monitoring system;
2. Evidence-driven quality education initiatives;
3. Inclusion interventions;
4. Safeguarding and psycho-social support activities.

Schools supported by Manahel receive two levels of these interventions:

- A. Access Only (AO) schools that receive only the teacher stipend and attendance support.
- B. Quality Education (QE) schools which receive the full package of intervention activities.

Study Purpose and Research Questions

The study examines the performance of Grade 3 students in reading and mathematics across Manahel-supported schools. The results will provide Manahel with insights to ensure the project meets the needs of the schools and the students it serves. They will also serve as a point of reference for comparison to an assessment conducted under the

previous iteration of the Manahel project in 2017 and a 2021 Manahel endline assessment. The study was conducted by Manahel partner, School-to-School International (STS).

Six research questions guided the study:

1. What proportion of Grade 3 students are classified as ‘progressing’ and as ‘proficient’ readers? How do these proportions compare by subgroups?
2. How do Manahel students’ learning outcomes at the beginning of Grade 3 in reading and mathematics compare to those for students assessed under Manahel’s predecessor project in 2017 at the end of Grade 3?
3. How do beginning of Grade 3 students’ learning outcomes in reading and mathematics in QE schools compare with that of AO schools? Within QE schools, how do schools with fixed libraries compare with those with mobile libraries?
4. How do beginning of Grade 3 students’ learning outcomes in reading and mathematics compare between Province A and Province B?
5. How do beginning Grade 3 students’ EGRA findings relate to the eight reading levels that Manahel- supported teachers are using to track learners’ progress? What is the relationship between Manahel- developed reading levels and EGRA performance?
6. What do we know about children’s current levels of stress and their relationship with their learning?

Methodology

The study draws on data collected from Grade 3 students at the start of the 2019/2020 academic year. Manahel captured student data using three tools: the Early Grade Reading Assessment (EGRA), which measures student performance on the foundational reading skills; the Early Grade Mathematics Assessment (EGMA), which measures student performance on the foundational skills of mathematics; and a Student Stressors Survey, which examines students’ experiences and feelings of safety during the conflict. Head teachers also provided school enrolment and attendance data.

The study estimated an overall required sample size of one-third of intervention schools by intervention type. This sample is larger than required to report results but necessary to accommodate comparisons between the 2019 assessment and 2021 endline if the ongoing conflict results in schools being dropped from the sample. Table 1 provides a summary of the target and final sample by province and school type.

Table 1. Target and Final Sample

Province	School type	Total schools	Target sample		Final sample	
			Schools	Students	Schools	Students
Province A	Quality Education	216	72	720	77	763
	Access Only	108	36	360	30	301
Province B	Quality Education	84	28	280	29	286
	Access Only	42	14	140	13	129
Total		450	150	1,500	149	1,479

Between 19 October 2019 and 6 November 2019, twenty enumerators visited each of the sampled schools. They assessed 10 randomly selected students — five girls and five boys — to complete the EGRA, EGMA and the Student Stressors Survey. Enumerators collected data electronically via the Tangerine® data collection software. Enumerators uploaded data daily from their tablets via wi-fi to a secure server monitored and maintained by STS.

The STS team applied sampling weights for more representative estimates, produced descriptive results analysed for statistically significant differences by gender, province and school type, and analysed associations between students' characteristics and scores to identify predictors of student performance.

Challenges and Limitations

- **High levels of displacement within the population.** With the population in flux due to the war, it is important to remember that the assessment tested individual students, not schools. Students may have transitioned between, or into, QE and AO schools. Whilst students reported moving frequently as a result of the war, the survey did not collect data on the length of time students had been in a Manahel-supported school, or which interventions they had received at that school. It also did not examine the displacement levels of teachers, as this was not the mandate. Therefore, the division of school types is not a clean representation of the interventions that students and teachers may have received.
- **Impact of heightened violence on students' performance and participation.** Enumerators reported nearby airstrikes occurred during school visits. One school needed to close early for security concerns, and 55 students refused to participate in some subtasks or the Student Stressor Survey.¹
- **Conflict made some areas inaccessible for enumerators.** Schools in these areas were excluded from the sample by necessity.
- **Lack of comparability between the assessment results of the previous iteration of the Manahel project, and Manahel's results.** Despite the original intention to compare the two assessments, several barriers arose during the design and implementation of the study. These included the large difference in the academic year progression between the sampled students within the two studies, the revisions made to items within the EGRA and EGMA subtasks, the move from paper-based to electronic scoring, and the change from remote to in-person enumerator training. Consequently, the results of the previous project and Manahel assessments are not comparable.
- **Structural bias in the comparison of scores of students in 'Access Only' schools to students in 'Quality Education'.** For the most effective use of donor funding, the Manahel project targets potentially struggling schools — based on a number of

¹ Under the standard EGRA student assent protocol, students may choose not to participate in the full assessment, a subtask of the assessment, or individual items. During this data collection, 88 students chose not to complete the oral reading fluency subtask and 55 students chose not to complete the student stressors survey.

contextual factors, such as existing resources, accessibility and community need – for their robust QE intervention. The AO and QE group designations were not created for evaluation purposes. However, at legacy DFID's direction, evaluators compared results for students in these two existing school types to provide a proxy measure of the difference in learning gains between the robust QE interventions and the minimal support AO interventions. The extent to which this comparison does provide a proxy measure is confounded with the selection criteria of which schools receive AO interventions and which receive QE interventions. Upon further discussion with the Manahel team, STS determined that the comparison between AO and QE provides a very limited view into how these two interventions compare. Any differences in the groups' results cannot be attributed solely to the different interventions, as students and schools are likely also influenced by the varied contextual factors – including those Manahel used to assign the schools' interventions in the first place.

- **Inability to pilot revised tools.** Manahel and STS conducted a thorough technical review of the tools to improve the assessments' accuracy. This review found that several items within the subtasks lacked Arabic modifiers, thus allowing multiple correct answers from children. Manahel staff made item-level revisions within the letter sound identification, nonword reading, oral reading fluency (ORF) and word problems subtasks. In the case of the ORF passage, these modifiers did not change the word or its meaning, but only clarified the pronunciation for the students. Under typical circumstances, the number of changes made to the tools would have prompted a pilot of the revised versions. However, due to constraints within the context and timeline, revisions to the tools could not be piloted with students. Additionally, the original 2017 War Stressor Survey was deemed too blunt to ask children, so was revised to adhere to the principles of Do No Harm.
- **Potential bias of using project and partner staff as enumerators.** Members of the Manahel staff and their partners at partner organisation 1 served as enumerators for this assessment. In a normal setting, an outside data collection firm would be used to limit potential bias within the assessment. However, the decision was made to use internal program staff and partners as they already had the requisite permissions needed to access schools, as well as familiarity with child safe-guarding practices. This allowed for efficiency within the short timeline and cost restrictions for the assessment. Furthermore, STS monitored the incoming data daily throughout data collection and felt confident in its quality and accuracy.

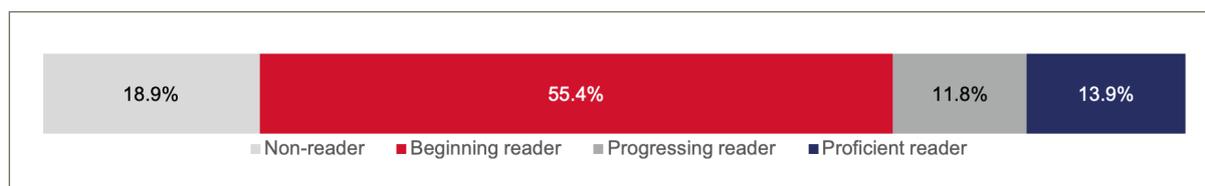
Findings by Research Question

RESEARCH QUESTION 1: STUDENT PERFORMANCE IN 2019

Overall, most students showed some reading ability (Figure 1). The majority of students (55.4%) were categorised as beginning readers; 11.8% as progressing readers; and 13.9% as proficient readers. Fewer than one in five students (18.9%) were non-readers.²

² Non-readers could not read a single word of the story reading passage. Beginning readers read between 1 and 22 CWPM but scored less than 80% on the comprehension subtask. Progressing readers read 23 CWPM or more but scored less than 80% on the comprehension subtask. Proficient readers are students who scored 80% or more on the reading comprehension subtask.

Figure 1. Percentage of Students by Reading Proficiency Level

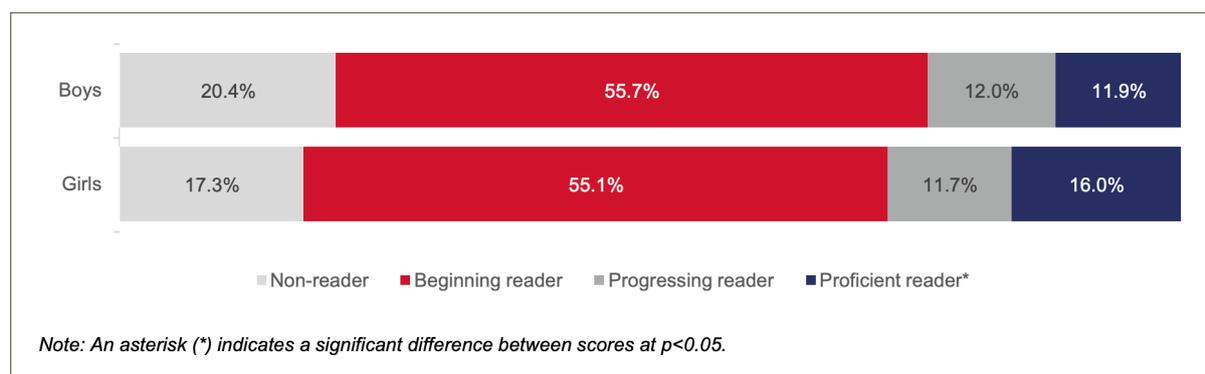


These reading categories were taken from a study conducted by the previous iteration of the Manahel project of 2017 and applied to learning data in 2019. As such, these categories are useful within the Syrian context but have not been aligned with international frameworks. At the country level, tracking the proportion of students who move in and out of these categories at endline will be instrumental in understanding the changes in student’s learning outcomes between 2019 and 2021. Manahel may consider engaging in an exercise of aligning the categories reported above with those described in the Global Proficiency Framework for students in Grade 2.³ This alignment exercise will reveal the proximity of the categories to the study conducted by Manahel’s processor project to internationally used categories, as well as underlying differences in expectations of students in each domain of reading.

Reading Results by Gender

The proportion of girls who are proficient readers (16.0%) was significantly higher than the proportion of boys (11.9%). The proportions of girls and boys were comparable in all other reading proficiency categories (Figure 2).

Figure 2. Percentage of Students by Reading Proficiency Levels by Gender



To identify students within these proficiency bands, STS computed fluency and accuracy scores for three reading subtasks: letter-sound identification, nonword reading and ORF. Girls’ fluency rates were significantly higher than boys’ rates on two subtasks: letter-sound identification and ORF. Girls also had higher accuracy scores in reading subtasks compared to boys. On average, girls answered 54.9% of the letter sound identification items correctly compared with boys, who answered 49.0% of the items correct; this difference was statistically significant. On the ORF subtask, girls answered significantly more ORF items correctly than boys: 33.5% correct compared to 29.8%, respectively. Girls also performed better on the reading comprehension subtask, where they averaged 31.8% of items correctly, whilst boys averaged 27.0%.

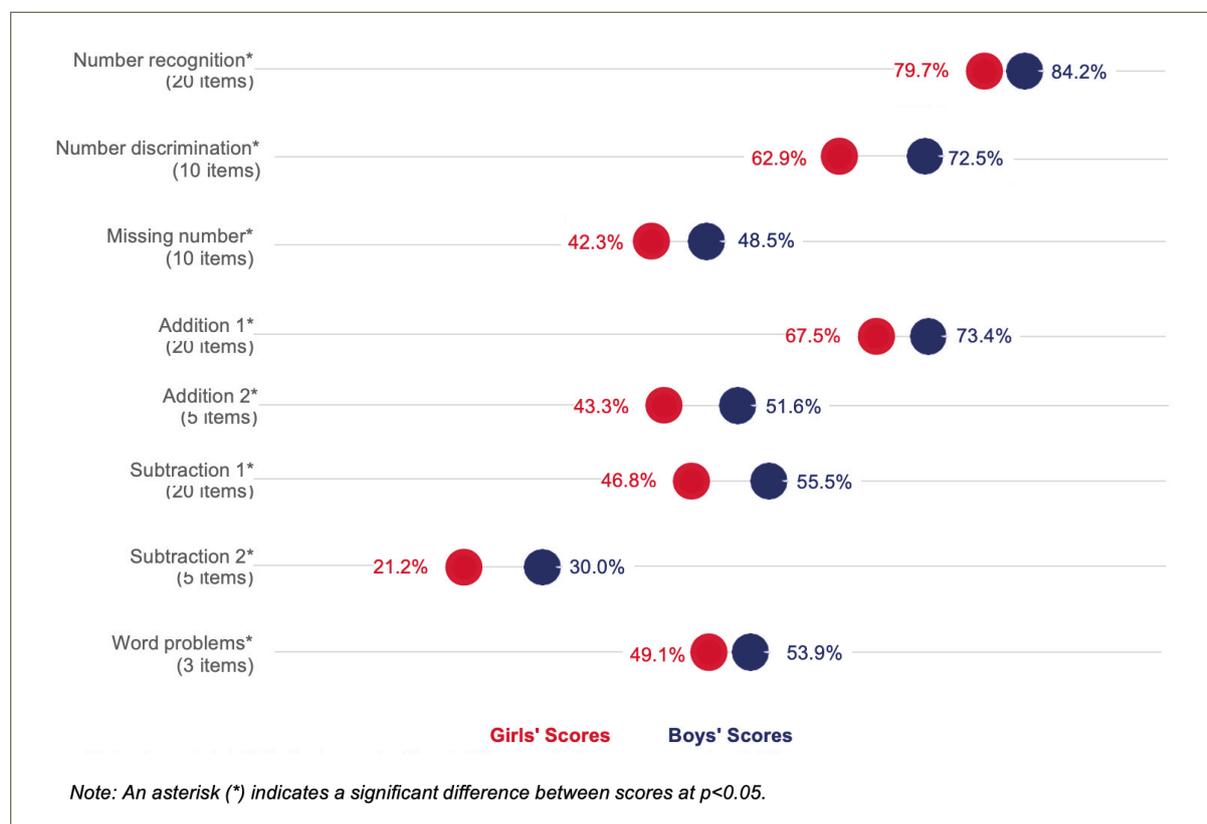
³ Global Alliance for Monitoring Learning, “Global Proficiency Framework for Reading and Mathematics - Grades 2 to 6,” October 2019, http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/Global-Proficiency-Framework-18Oct2019_KD.pdf

Notably, boys had significantly higher accuracy scores on the listening comprehension subtask than girls. Boys, on average, answered 81.9% of the six listening comprehension questions correctly, compared to girls, who answered 77.2% correctly, on average. Boys' higher performance was also reflected in the proportion of zero scores — almost twice as many girls as boys did not answer a single item correctly — 3.2% of girls compared to 1.5% of boys.

Mathematics Results by Gender

Additionally, STS conducted analyses on mathematics outcomes by gender. In contrast to reading, boys outperformed girls in all mathematics subtasks. Boys' higher performance was reflected in their significantly higher accuracy scores and lower percentages of zero scores. Figure 3 shows the average accuracy score for boys and girls by mathematics subtask. On every subtask, boys had a statistically significantly higher accuracy score than girls.

Figure 3. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Gender



RESEARCH QUESTION 2: COMPARISON OF 2017 AND 2019 STUDENT PERFORMANCE

One of the outcomes in Manahel's logframe is 'improved learning outcomes', meaning that the project's performance, ideally, would be assessed as a percentage change in EGRA scores, using the 2017 results as a baseline. However, in consultation with the Manahel team, STS assessed the validity of such comparison and determined that the results of the Manahel project's predecessor study in 2017 and Manahel's 2019 results are not comparable. Therefore, a comparison of results is not presented in the body of the report.

Comparison between the 2017 and 2019 results is tenuous given the difference in timepoints — end of Grade 3 in 2017 and beginning of Grade 3 in 2019 — as well as the changes in the EGRA tools. During the design phase, Manahel and STS chose not to assess students at the beginning of Grade 4, which would have been more comparable to the 2017 assessment. However, assessing Grade 4 students would not be properly aligned with the project’s logical framework. Instead, Manahel chose to align the midline study with the logical framework, to ensure a comparable endline study in 2021.⁴

The decision to compare students’ performance in 2017 and 2019 was made solely due to the lack of a better alternative. It is recommended that a comparison to the 2017 results be removed from the logframe and the endline study. Instead, endline 2021 results should be compared to the 2019 results—as long as the samples are still comparable.

The comparisons between 2017 and 2019 are provided in Annex G because the current logframe requires comparing results to 2017. STS recommends that the logframe be updated to remove comparisons to 2017 for the previously stated reasons.

RESEARCH QUESTION 3: OUTCOMES BY SCHOOL TYPE

The Manahel program team included a research question examining school type results to see if it was possible to determine the value add of QE interventions. Under the program’s objectives, both AO and QE schools provide children with more stability and psychosocial support, which should have a positive impact on their wellbeing and learning. Beyond this, the Manahel team hypothesised that the additional academic support to QE schools should result in larger gains in children’s learning.

However, it was uncovered during the validation process that the labels, AO and QE, do not accurately reflect what has occurred in a school, and in turn, comparisons of results by these labels are not reliable. With the reported high levels of displacement within the Student Stressors Survey⁵, students and teachers who have benefited from the Manahel program may have relocated, making it challenging to determine which beneficiaries have received which interventions.

Additionally, the role of other NGOs, funders and education organisations in Manahel-supported schools is not tracked and therefore cannot be accounted for in the analyses. Lastly, the AO and QE assignments were made for programmatic purposes, not evaluation purposes, and comparison is confounded with the selection criteria for which schools receive the more robust QE interventions. As such, the analyses by these labels — AO and QE — cannot be used to draw conclusions about the comparative impact of QE schools over and above AO schools at this time. At endline, it is recommended that a survey of teachers, headteachers and/or students be considered to capture the types and duration of Manahel interventions that these individuals have experienced. This data can instead be used to generate labels, or classifications, for analysis at endline.

⁴ This was discussed with legacy DFID and approved by them.

⁵ Two-thirds of students (63.4%) reported being forced to move due to the war. The majority of those students reported moving two times. Overall, 40.1% of students reported attending a new school due to the war.

As with the prior research question, results by school type are still included in this report to answer the stated research question. However, results should be interpreted with caution for previously stated reasons.

Overall, students' reading proficiency levels were comparable between schools labelled as AO and schools labelled as QE, but results for individual subtasks show some significant differences. AO schools have higher scores in the reading and listening comprehension subtasks than QE schools, but they also had higher proportions of zero scores on letter sound identification and nonword reading subtasks. AO students had significantly higher accuracy scores in ORF and listening comprehension subtasks than peers in QE schools; however, both groups had comparable fluency rates.⁶ QE schools had smaller proportions zero scores than AO schools on foundational skills subtasks — letter sound identification and nonword reading.⁷

On EGMA subtasks, students in AO schools had higher average accuracy scores than QE peers on all subtasks except addition level 2.⁸ AO schools also had higher average fluency rates on all three timed subtasks.

Province and school type also played a factor in reading performance by gender. STS conducted a multinomial logistic regression examining province, school type, and gender on reading and mathematics performance to understand these factors' impact. This analysis showed that being a girl is associated with a 0.4 increase in the odds of being a proficient reader compared to a non-reader, accounting for province and school type. Being from Province B is associated with a 0.7 increase in the odds of being a proficient reader, accounting for gender and school type. Being from a QE school is associated with a -0.4 decrease in the odds of being a proficient reader, accounting for gender and region. Results by gender are further discussed below and by province in the next research question.

Results by Gender

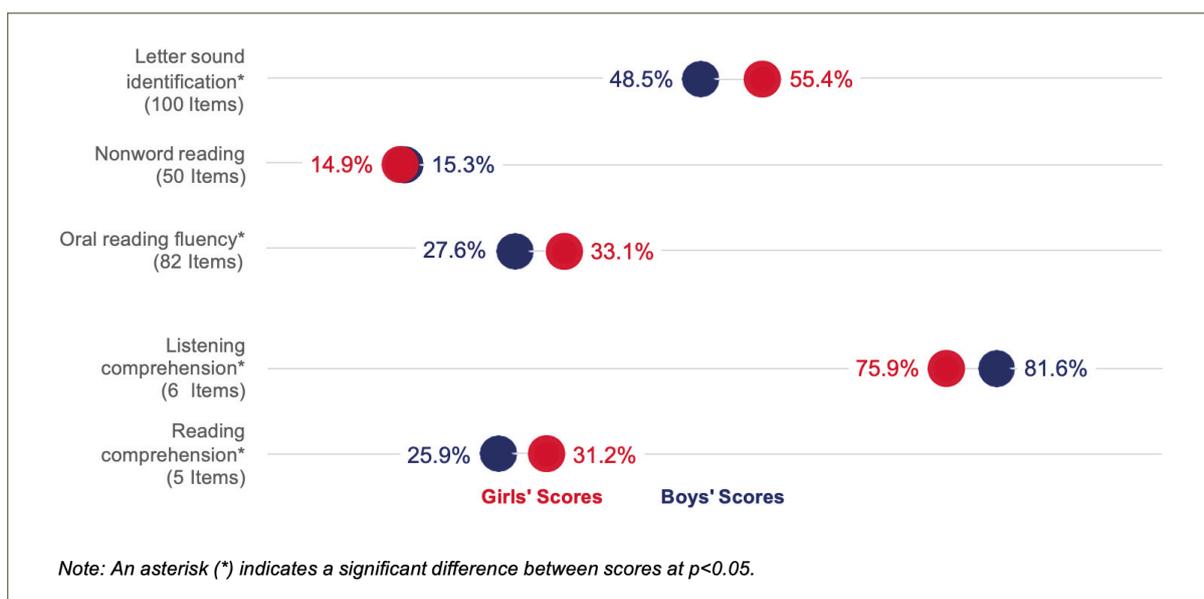
While AO schools have higher scores on some reading subtasks — reading comprehension and listening comprehension — than QE schools, there were no differences within AO schools by gender. However, gender did correlate with differences in reading scores within QE schools (Figure 4). Girls in QE schools had statistically significantly higher ORF accuracy scores than boys in QE schools. The lower ORF accuracy scores were observed amongst boys in QE schools, whilst girls in QE schools performed comparably to girls and boys in AO schools. By contrast, boys in QE schools had statistically significantly higher listening comprehension scores than girls in QE schools.

⁶ On average, AO students answered 34.6% of the 82 ORF items correctly, compared to 30.5% for QE students. AO students also answered an average of 81.9% of the six listening comprehension items correct, compared to 78.6% for QE students.

⁷ Of AO students, 14.6% did not identify a single letter sound, compared to 10.31% of QE students. Similarly, 53.7% of AO students did not read a single nonword, compared to 45.8% of QE students.

⁸ There were two addition subtasks. The first (Addition one) had 20 questions administered to all students. The second (Addition two) had five questions and was only administered to students who answered all of Addition one's first 10 questions correctly.

Figure 4. Reading Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender



On the mathematics assessment, boys in both QE and AO schools had significantly higher scores than their female peers on all subtasks; however, on word problems, a significant difference was only observed between boys and girls in QE schools (Figure 5 and Figure 6).

Figure 5. Mathematics Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender

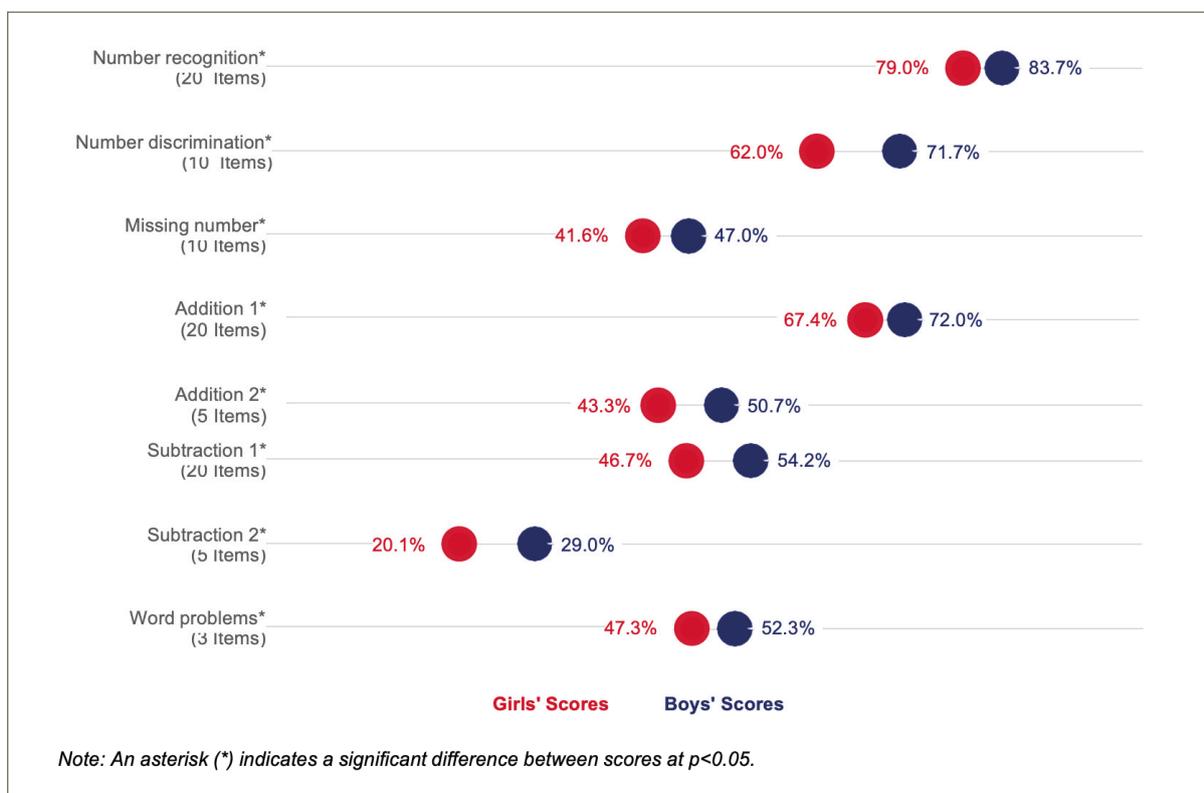
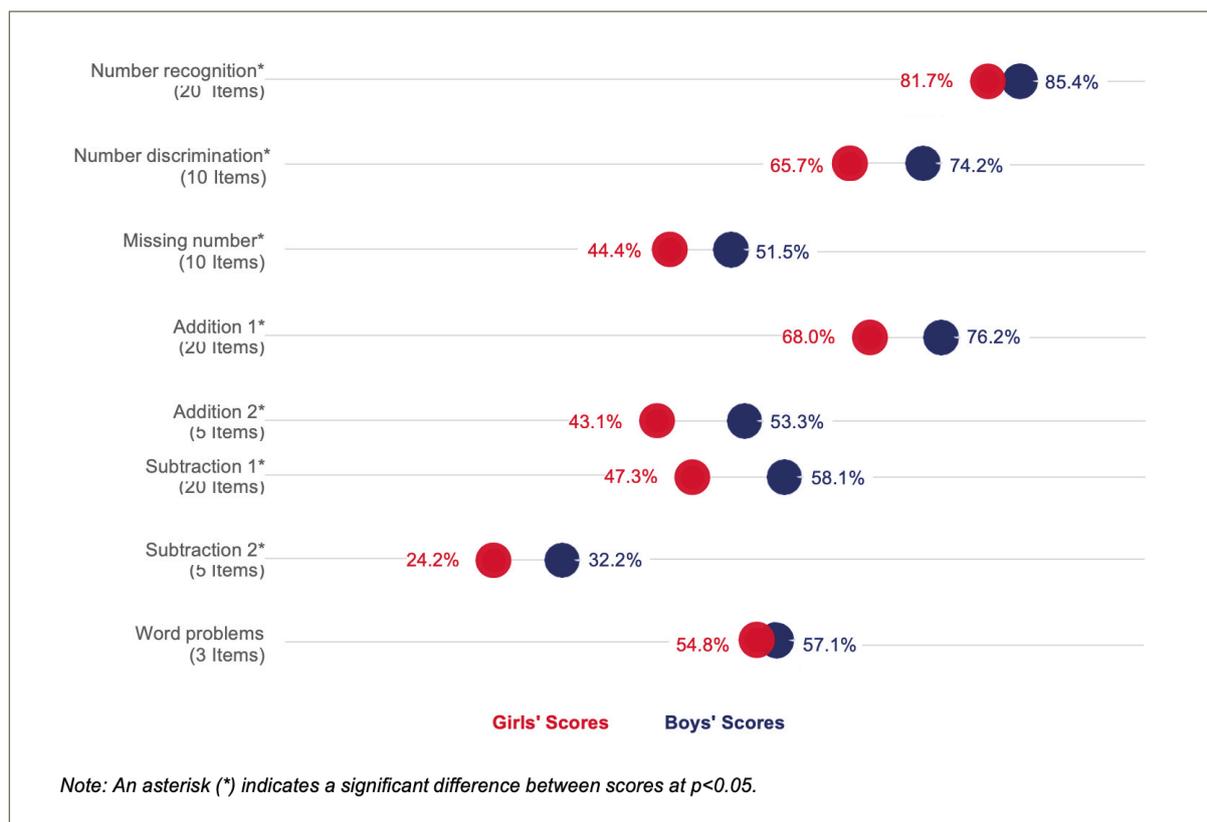


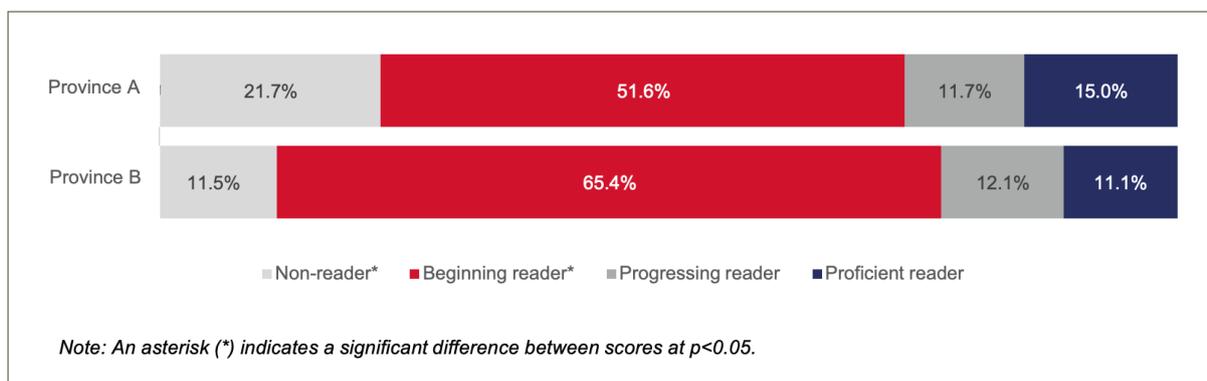
Figure 6. Mathematics Percentage of Items Correct (Accuracy Scores) in AO Schools by Gender



RESEARCH QUESTION 4: OUTCOMES BY PROVINCE

In reading, students' distribution differed significantly by province at the non-reader and beginning reader levels (Figure 7). A higher percentage of students from Province A (21.7%) were non-readers than in Province B (11.5%). Conversely, a higher percentage of Province B students (65.4%) were beginner readers than Province A (51.6%). Proportions were comparable at the progressive and proficient level. The district Manahel works in is densely populated with a high proportion of displaced children and continued high levels of movement. However, there is also considerable displacement into and from Province A. Therefore, provincial differences do not necessarily result in teaching/learning differences in those locations – displacement could be a more significant factor.

Figure 7. Percentage of Students in Province A and Province B by Reading Proficiency Level



A comparison of zero scores supports the trend that students in Province B performed better in reading than their Province A peers. On average, 13.1% of students in Province A did not identify a single letter sound correctly, significantly higher than the 7.7% in Province B. Similarly, 52.1% of students in Province A did not read a single nonword compared to 37.8% of Province B students; 21.7% of Province A students did not read a single word in the ORF subtask, compared to 11.5% of Province B students. Manahel reported that all learners should know the shape and name of the letter by Grade 2; they hypothesise that the results here are because of difficulties reading the modifiers.

Similar to EGRA accuracy scores, students in Province B performed significantly better than students in Province A on three mathematics subtasks – number discrimination, missing number and word problems. The proportion of students with zero scores was significantly different between the provinces on three EGMA subtasks – addition level 1, subtraction level 2 and word problems – with more students in Province A receiving zero scores than in Province B. As noted above, displacement confuses the picture: a learner assessed in Province A might not have received most of their teaching in Province A.

Usually teaching in Province B is slightly better, all other factors being equal. There are fewer schools and more qualified teachers. Province B has a standardised test for teachers, with a heavy focus on Arabic and numeracy. In Province A, there are many schools and fewer teachers.

Results by Gender

When considering gender within each province, some differences emerge. While students in Province A were more likely to be classified as non-readers, boys within Province A tended to have lower accuracy scores than girls in Province A (Figure 8). On three reading subtasks – letter-sound identification, oral reading fluency and reading comprehension – girls had higher accuracy scores than boys in Province A. On one subtask – listening comprehension – boys in Province A had higher accuracy scores than girls. In Province B, girls had higher accuracy scores than boys on only the letter sound identification subtask (Figure 9). No significant differences were observed on the remaining reading subtasks.

Figure 8. Reading Percentage of Items Correct (Accuracy Scores) in Province A by Gender

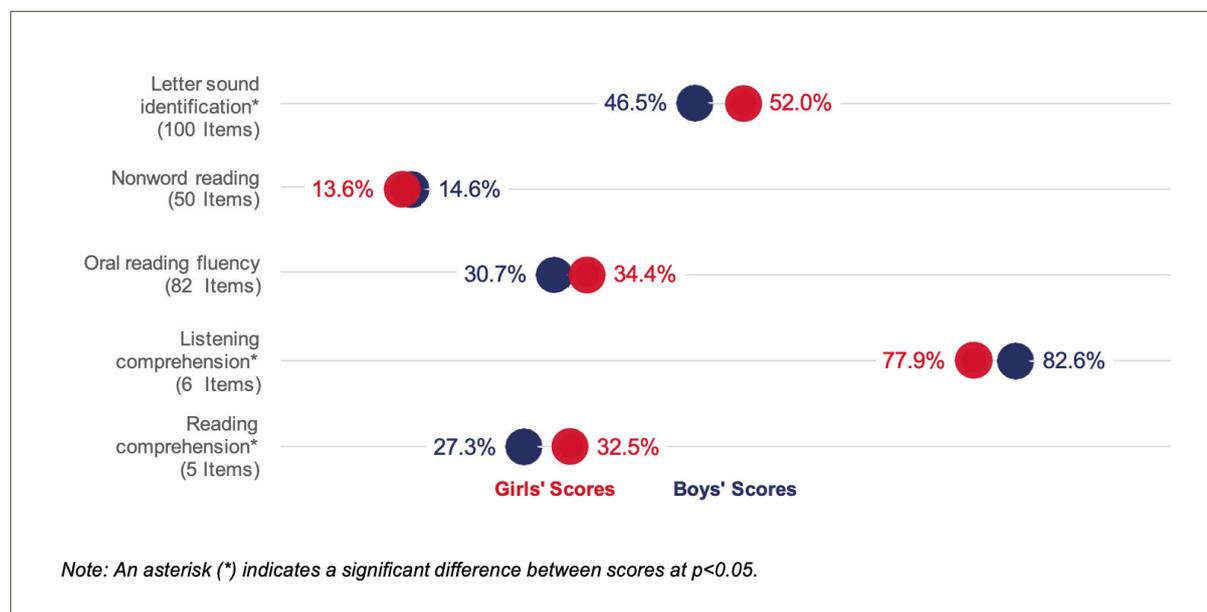
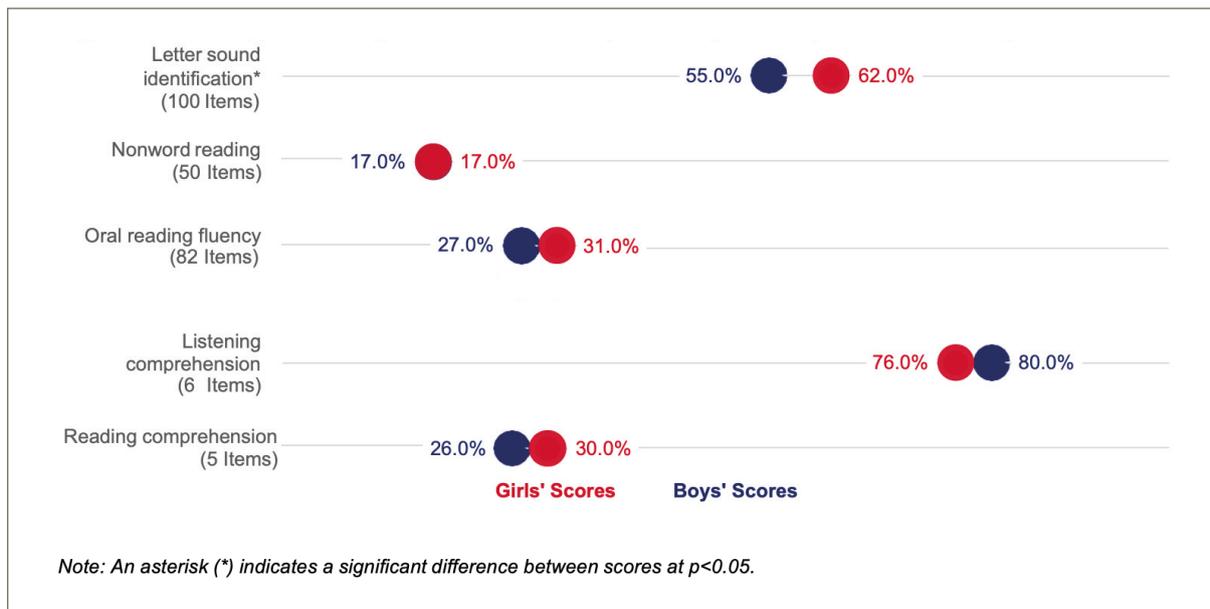


Figure 9. Reading Percentage of Items Correct (Accuracy Scores) in Province B by Gender



In mathematics, boys tended to have higher accuracy scores than girls, but this held true in both Province A and Province B (Figure 10 and Figure 11).

Figure 10. Mathematics Percentage of Items Correct (Accuracy Scores) in Province A by Gender

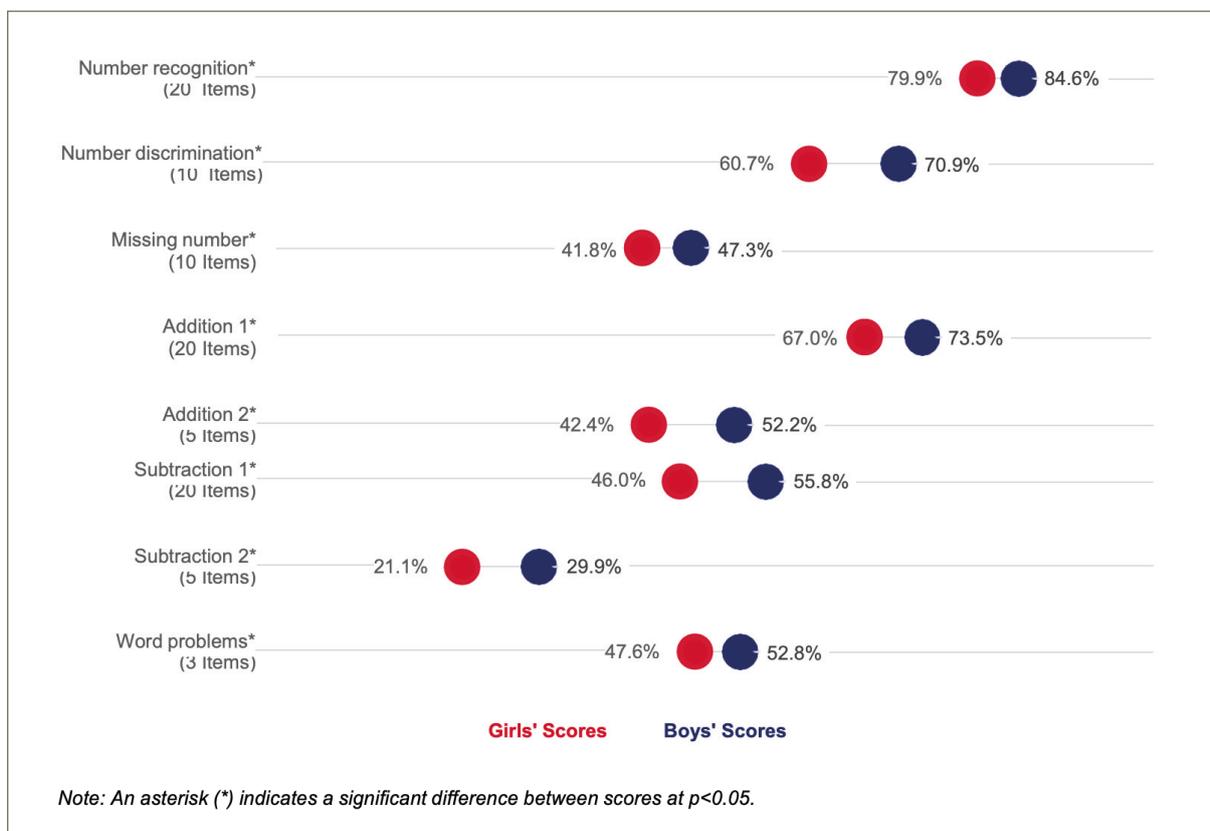
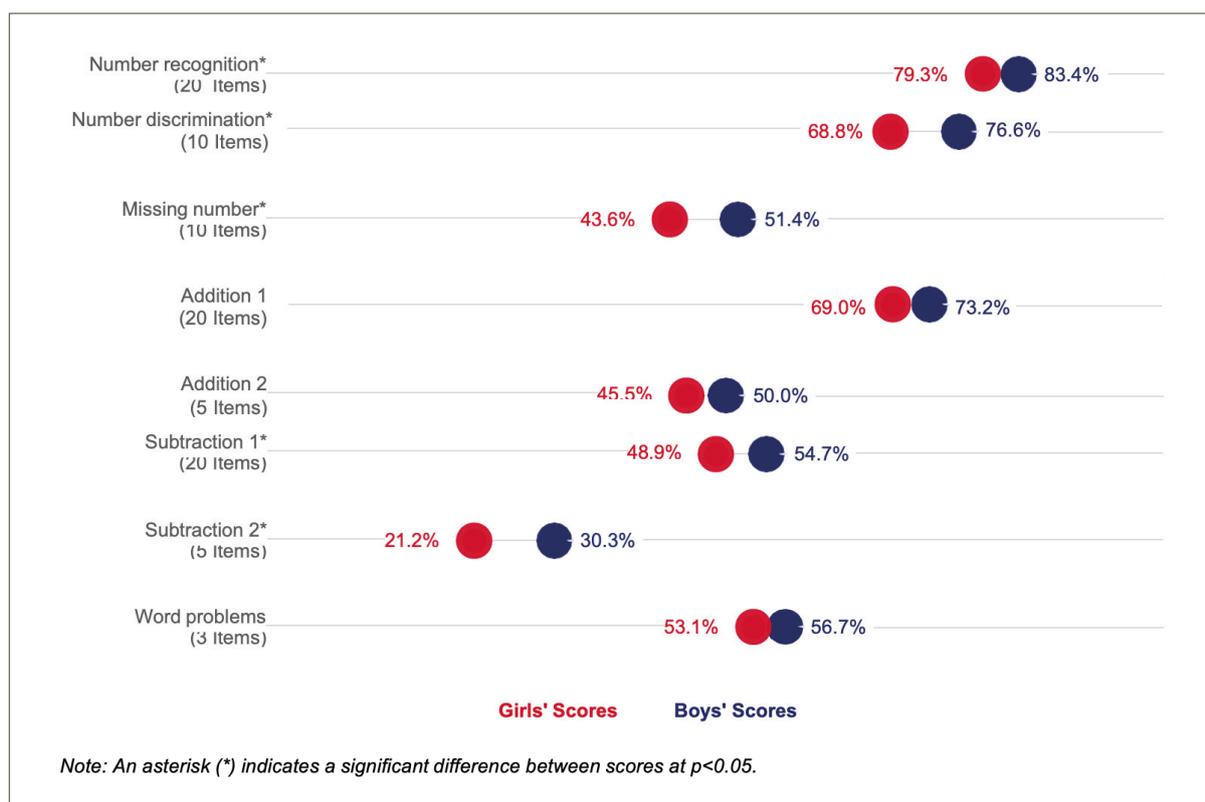


Figure 11. Mathematics Percentage of Items Correct (Accuracy Scores) in Province B by Gender



RESEARCH QUESTION 5: LINKING EGRA PERFORMANCE WITH MANAHEL READING LEVELS

Teachers at QE schools rated the reading proficiency of their students on an eight-level scale developed for use within the Manahel project. Manahel uses these eight reading levels to track progress in students' advancement towards Arabic reading proficiency. Although the reading levels were not designed to be directly compatible with EGRA, limited comparisons help contextualise EGRA results within Manahel's reading levels. The original intention of this research question was to examine the rating assigned by a teacher to a student against the students' EGRA scores. However, the student-level Manahel ratings were not provided to STS for analysis. Instead, to answer the research question, STS conducted a general alignment of Manahel levels and EGRA subtasks. At endline, if the link between Manahel reading levels and EGRA performance is still of interest to the project, those ratings should be compiled and provided for the sampled students.

Using the description of the Manahel levels and the skills assessed in each task, STS mapped each Manahel level to a task, with letter-sound identification as the simplest task up through reading comprehension as the most complex reading task. The alignment of Manahel levels and EGRA subtasks is provided in the body of the report. On average, 21.7% of students mastered Manahel level 1 – learners can recognise letter names – which mapped to the letter sound identification subtask. Fewer than one percent of learners mastered Manahel levels 2 through 4. The proportion of students who mastered Manahel reading level 2 – learners can read letters, words and sentences with short vowels—was 0.8% and the proportion who mastered Manahel reading levels 3 and 4 – learners can read and sound out all letters, words and sentences with the Sokoon modifier and long vowels – was 0.7%. No EGRA subtasks aligned with Manahel

levels 5 and 6. For Manahel levels 7 and 8, 5.5% of students mastered this level — learners read sentences and paragraphs with comprehension.

Results by Gender

When considering gender within each Manahel level mastered, girls were more likely to have mastered level 1 than boys. However, the difference between boys and girls was similar within provinces and school type.

RESEARCH QUESTION 6: STUDENT STRESS AND THE CONFLICT

STS examined the relationship between students' EGRA and EGMA scores with responses to the Student Stressor Survey — including frequency of moving due to war and students' tiredness and hunger at school as proxies for their stress levels. Generally, students who reported moving homes six times or more due to the war had lower reading and mathematics outcomes. Tiredness at school was generally associated with lower reading and mathematics outcomes. Experiencing hunger at school only showed a relationship with students' listening comprehension performance and mathematics outcomes, whereas those who experienced hunger had lower scores than those who did not.

Two-thirds of students reported that they were displaced at least once due to the war — more in Province B than in Province A. Additionally, students in District 3 were the most likely to report being displaced due to the war, followed by students in District 2 and District 4 districts and District 1. Students in District 5 and District 6 were more likely than students in District 1 or District 4 to have been displaced five times or fewer. Of those students who moved, the majority reported being displaced between one and three times. Relatedly, half of all students reported experiencing hunger at school and more than one-third reported tiredness. These rates were higher among students in Province B than in Province A.

Analysis of learning outcomes by displacement rate show that, in general, students who have moved fewer than six times as a result of the war have higher reading and mathematics outcomes compared to students who moved six times or more.⁹ This is in line with findings highlighted in, for example, ODI (2015)¹⁰, which notes that continuity of education is important for learning and wellbeing. Due to weak and statistically insignificant relationships between EGRA and EGMA subtask scores and how frequently a student reported displacement due to the war, students were grouped into two groups based on their displacement response: those who reported being displaced one-to-five times and those displaced six times or more.¹¹ Using this dichotomous grouping of displacement, the proportions of students in reader categories were notably different, suggesting that more than six displacements due to war more is associated with statistically significantly lower learning scores, whilst fewer displacements are not associated with lower scores. Specifically, 28.0% of students who moved one-to-five times were progressing or proficient readers, compared to only 22.7% of students who moved six or more times.

⁹ Comparisons by two groups — students with one to five moves and students with six or more moves — allowed analysts to examine trends by subgroups. Regression analyses of the number of moves on individual subtasks yielded inconsistent and unclear trends; in addition, there were poor correlations with frequency of moves and subtask scores.

¹⁰ Education in emergencies and protracted crises: Toward a strengthened response (Nicolai, Hine and Wales)

¹¹ The correlations ranged from -0.14 between frequency of displacement and reading comprehension scores to -0.03 between frequency of displacement and number recognition.

Further analysis by district reveals that frequency of displacement has a mixed relationship with learning outcomes. Students in District 2 moved, on average, 2.5 times due to the war and tended to have higher scores than students in District 3 who moved an average of 4.1 times. However, students in District 1 moved an average of 2.4 times and had statistically significantly higher scores than students in District 3 on several subtasks.

STS also collected data on whether students reported attending a new school due to war-related displacement. Overall, 40.1% of students reported attending a new school as a result of moving due to the war. On average, students attended 2.3 new schools as a result of the war, and 22.3% of students reported that there were times when they did not have a school to go to. Students were not asked the length of time they were out of school during this study, but this may be a helpful question to add at endline.

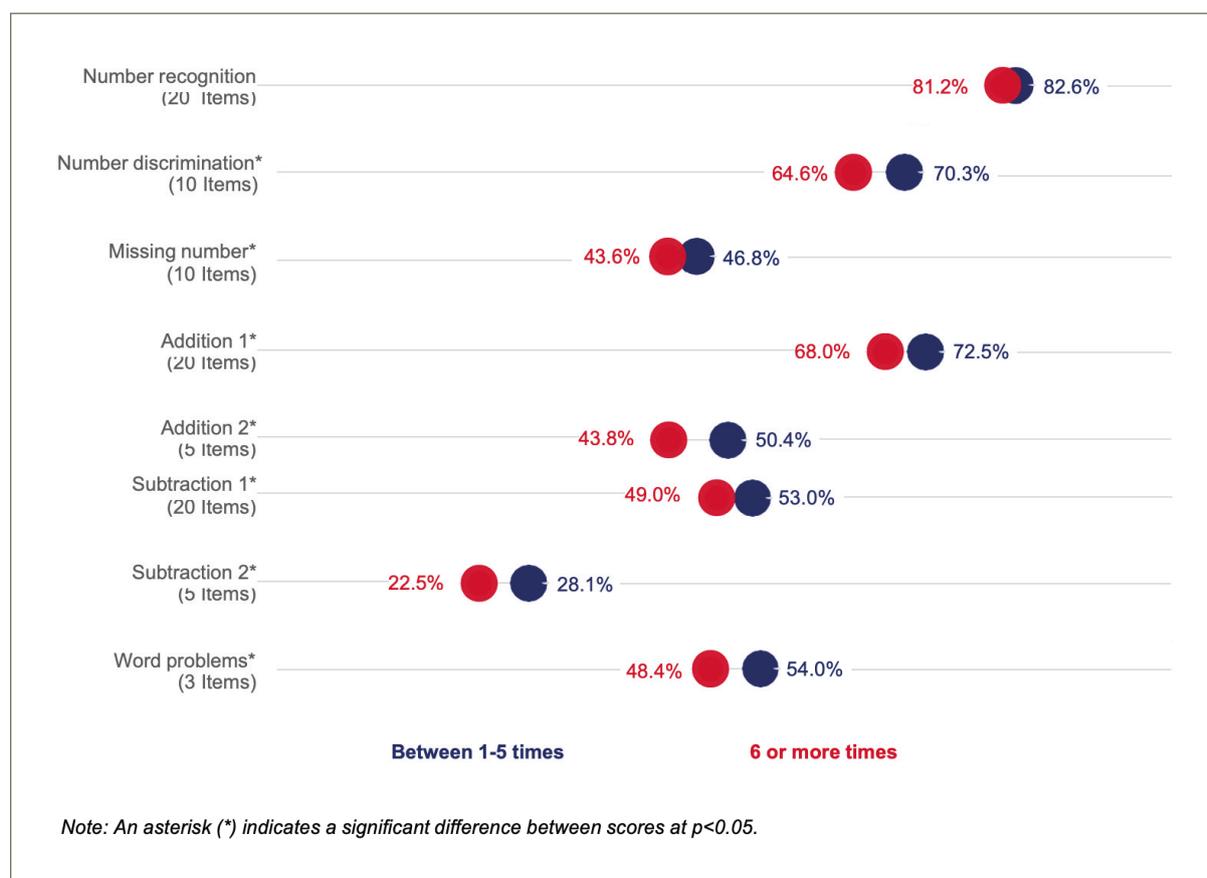
High rates of attending a new school also had a negative relationship with learning outcomes. Students who were non-readers were more likely to have attended four or more new schools. By contrast, students who moved between one and five times and had attended fewer than four schools were more likely to be classified as progressing or proficient readers.

Moving homes and schools due to war appears to be confounded. Students who reported that they did not move to a different school but had moved homes due to the war had comparable scores to students who moved both homes and schools. Amongst students who did not have to move homes, those who attended a different school had higher scores than those who did not attend a different school. This finding suggests that the impact of displacement of homes and schools is intertwined with the impact on communities overall – and that students who did not experience displacement are not necessarily most likely to succeed.

Overall, fewer house moves and fewer schools were associated with higher accuracy scores on all reading and mathematics subtasks. Students who attended between one and three schools had significantly higher accuracy scores on all EGRA subtasks – except letter sound identification – than students who attended four or more new schools. On all EGMA subtasks except number recognition, students who had moved between one and five times as a result of the war had significantly higher accuracy scores than students who moved more than six times (Figure 12).

Determining the impact of displacement on learning amongst students who have experienced fewer than five house moves or four or fewer school moves is not possible given the available data, given the mixed results by district and confounding of moving both home and school. What may be more appropriate is a measure of the learning environment, inclusive of the frequency of home and school moves that the student experiences. To further disaggregate and understand the impact of displacement on student learning and the learning environment in general, additional data on the time between displacements, distance moved, extent of a move's impact on family, change in schooling experience, peers and staff, impact on community, as examples, should be collected.

Figure 12. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Displacement Group



Conclusions and Recommendations

The conclusions of the study point to several recommendations for Manahel to consider as the project carries out its activities. A summary of these recommendations and the key findings upon which they were based appears in Table 2, in order of priority.

Table 2. Summary of Key Conclusions and Recommendations

Key Conclusion	Recommendation
Manahel Program Implementation	
Overall, the majority of Manahel students were beginner readers at the beginning of Grade 3.	<ul style="list-style-type: none"> → The emphasis of Manahel's work with teachers and teaching resources should shift more to the earlier literacy levels. Continue to emphasise reading mechanics and comprehension in Manahel interventions to help move students from beginner readers to progressing readers. → Provide targeted support to those districts where children have the lowest reading performance, including District 3, District 5 and District 1.

<p>A significantly larger proportion of girls than boys were proficient readers, but boys outperformed girls on all mathematics subtasks.</p>	<p>→ Consider gender within teacher training and instruction and provide additional support to boys for reading and girls for mathematics to bridge the performance gaps between genders.</p>
<p>Overall, results from QE schools with mobile and fixed libraries were comparable, but schools with mobile libraries outperformed their peers in oral reading fluency.</p>	<p>→ Alter the way in which AO and QE labels are determined at endline so that labels accurately reflect the interventions received.</p> <p>→ Additional data on usage and access to libraries is needed to identify any differences between the schools that may be associated with higher reading scores, including the availability of age-appropriate materials, frequency of access, and tie-in to instruction.</p>
<p>Student stressors, including displacement, lower attendance and tiredness, had predictable relationships with student reading and mathematics skills.</p>	<p>→ Explore ways for existing supports, such as safeguarding officers, to engage with families to help increase attendance for those students with chronic absenteeism, particularly in more stable areas of Provinces A and B.</p> <p>→ Beyond students with acute needs who must be prioritised, more finely triage the needs of students transitioning between schools to target emotional and academic support and balance the disruption to their schooling.</p> <p>→ Additional data may be useful to understand factors that may preclude students from attending school regularly or may hinder their performance at school.</p>
<p>Students in Province B generally performed better in reading and mathematics foundational skills than students in Province A.</p>	<p>→ Provide targeted and additional support to schools labelled as QE in 2019 in Province A to improve their performance. Understanding the nature of other supports being provided to schools will also be important at endline, particularly if differences in Provinces A and B persist.</p>
<p>Study Design for 2021 Endline</p>	
<p>Nonword reading skills are not strongly related to other foundational reading skills or reading comprehension.</p>	<p>→ While this subtask is a good differentiator between proficient readers and progressing or emerging readers, it has a weaker relationship with reading comprehension than other tasks. Recommend assessing whether the subtask is appropriate for the instructional approach used by Manahel and retain at endline accordingly.</p>
<p>Results from the assessments are student-centred, while interventions are teacher- and school-centred.</p>	<p>→ Include a teacher interview at endline, which asks about teacher displacement and participation in the Manahel program to contextualise learning outcome findings as well as capture participation in other interventions by other organisations. Possible format examples include interviews, interviews using a 'most significant change' survey, or focus groups with teachers, head teachers and/or child safeguarding officers.</p>

	<ul style="list-style-type: none"> → Add follow-up questions within the Student Stressor Survey for more nuanced information on the number of moves and schools attended, length of time at current school, length of time out of school, information on previous school and participation in Manahel interventions. → Conduct an alignment exercise between Manahel reading levels and international reading levels, such as those proposed in the Global Proficiency Framework.¹²
<p>The delineation between the AO and QE school types was not an appropriate binary for analysis.</p>	<ul style="list-style-type: none"> → Adding teacher-provided responses regarding participation in interventions and assign intervention levels based on this information during data analysis.
<p>Comparison between the results of Manahel's predecessor's study in 2017 and the 2019 Manahel results is not a valid comparison.</p>	<ul style="list-style-type: none"> → Update Manahel logframe to remove comparisons to 2017 results due to differences in time of assessment, tools and context. → Revise the research question to focus on the change from 2019 to 2020. → Capture demographic information during the endline to determine the comparability of the 2019 Grade 3 population versus the 2021 Grade 3 population. → Generate an index of conflict, reported at the district level at a minimum, preferably at the school level. The index can be included in analyses to examine the influence of conflict on student learning outcomes, without having to survey individual students directly about their conflict experience.

¹² Global Alliance for Monitoring Learning, "Global Proficiency Framework for Reading and Mathematics - Grades 2 to 6," October 2019, http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/Global-Proficiency-Framework-18Oct2019_KD.pdf



Introduction

Conflict and Education in Syria

The protracted Syrian crisis, which has been marked by ongoing conflict since March 2011, has been characterised by human rights violations, staggering displacement, and significant loss of life. According to the United Nations High Commissioner for Refugees, nearly 5.7 million Syrians have fled the conflict to other countries while 6.2 million more have been displaced within Syria.^{13,14} Together, these groups comprise 55% of the pre-war population of Syria.¹⁵

Nearly nine years of acute crisis has had a devastating impact on the education sector in Syria. Syria held a 93% enrolment rate before the war.¹⁶ Today 2.1 million school-aged children are out of school, and one in three schools have been destroyed.¹⁷ Children in the early grades of primary school have never known a Syria without war. Their school days are punctured by airstrikes and burdened by the emotional and physical toll of personal loss and continued instability. Teachers, bearing the same burdens, choose to go to schools in the face of danger, missing or sporadic pay and the challenge of providing a semblance of normalcy for their students.¹⁸

Manahel Project Background

The three-year Manahel Syria Education Programme is funded by the United Kingdom's Foreign, Commonwealth and Development Office and implemented by Chemonics International. From February 2018 to June 2022, the Syria Education Programme aims

¹³ 'Syria Regional Refugee Response,' Operational Portal, *United Nations High Commissioner for Refugees*, last modified December 1, 2019, data2.unhcr.org/en/situations/syria

¹⁴ 'Internally Displaced People,' Syria, *United National High Commissioner for Refugees*, www.unhcr.org/sy/internally-displaced-people

¹⁵ British Broadcasting Corporation, 'Why is there a war in Syria?'

¹⁶ 'School Enrolment, Primary (% net) - Syrian Arab Republic,' World Bank, data.worldbank.org/indicator/SE.PRM.NENR?locations=SY

¹⁷ 'Whole of Syria Facts and Figures, Education, Mid-Year 2019,' *United Nations Children's Fund*, www.unicef.org/syria/media/4231/file/UNICEF%20Facts%20and%20Figures%20Education%20Q2%202019.pdf

¹⁸ 'Education,' *United Nations Children's Fund*, www.unicef.org/syria/education

to provide access to safe, inclusive and quality learning opportunities for children in Northwest Syria, while strengthening educational actors to manage education effectively.

The Manahel project builds upon a previous iteration of the project, funded by legacy DFID and the European Union between 2014 and 2018. The previous project began as a project to provide financial support to teachers and education staff and build the capacity of the education directorates (ED) supported by Syrian Interim Government (SIG). This consistent support allowed the EDs to maintain a basic level of education within opposition-held areas of Syria.

For the final six months of the project, legacy DFID requested that the previous project add an intensive pilot of activities aimed at improving the quality of education, psychosocial support and inclusive learning for students within Province A, Province B and rural Province C. In preparation for these activities, the project conducted baseline Early Grade Reading Assessments (EGRAs) and Early Grade Mathematics Assessments (EGMAs) with Grade 3 students at the end of the 2016/2017 academic year. The previous project included a complementary suite of Snapshot of School Management and Effectiveness (SSME) tools and a student War Stressors Survey to provide deeper context for the EGRA and EGMA results.¹⁹

Manahel Interventions

Building on the lessons learned from the previous iteration of the project, the Manahel project focuses on four components:

1. Stipend payments to teachers and education staff, along with a teacher payment and attendance monitoring system;
2. evidence-driven quality education initiatives;
3. inclusion interventions and,
4. safeguarding and psycho-social support activities.

The programme's intended impact is on student resilience and learning outcomes. These should be improved through the provision of quality and inclusive formal and informal learning opportunities. The measures of impact are:

- Percentage of students in the top two categories in proficient reader and advanced progressive reader of early grades students as measured by EGRA results.
- Percentage of children with improved resilience as measured by the Strengths and Difficulties Questionnaire (SDQ) scale.

The intended outcome of the project is that children (50% of girls) have access to safe, inclusive, and quality learning opportunities in learning environments that foster psychosocial well-being. This is to be achieved through the outputs in Table 3, below.

¹⁹ The previous project adapted the War Stressors Survey from the 'Exposure to War Stressors Questionnaire' developed by the Children and War Foundation. The survey consisted of 18 questions about a student's exposure to specific aspects of the war – from witnessed acts of violence, impacts on their home and family and direct personal experiences of violence. The original tool can be found on the Children and War Foundation's website at childrenandwar.org/projectsresources/measures.

Schools supported by Manahel are assigned to receive two different levels of intervention — Access Only (AO) or Quality Education (QE). In academic year 2019/2020, Manahel supported 450 schools; 150 receiving AO support and 300 receiving QE interventions. Activities within the two levels of intervention are outlined in Table 3. It should be noted that these descriptions are only the interventions provided by Manahel. Interventions by other nongovernmental organisations are not considered when differentiating between the school types.

Table 3. Manahel Intervention Levels — Schools That are Exclusively Supported by Manahel

Access Only Intervention	Quality Education Intervention
<ul style="list-style-type: none"> • Monthly stipend payments to teachers and education staff • Teacher payment and attendance monitoring 	<ul style="list-style-type: none"> • Monthly stipend payments to teachers and education staff • Teacher payment and attendance monitoring • Psycho-social support workshops and activities • Child safeguarding and child protection activities • Reading and mathematics instruction, including three reading and one mathematics lessons per week • Fixed and mobile library support • Teachers training, supervision and coaching • Identification and adaptations to accommodate for children with mild or moderate disability

Manahel assigned the level of intervention for each school based on consultations with the community and the SIG education authorities. They considered several factors, including the number of students, the presence of a formal school building, the accessibility of the location to children, the availability of space to host a library and the severity of need in the community based on the Syrian Humanitarian Needs Overview and displacement data.²⁰

The Manahel project introduced the reading and mathematics activities within the QE intervention in phases known as ‘waves’, beginning in December 2018 (Table 4). As of September 2019, all QE schools received supplemental reading and mathematics lessons for Grades 1 to 4, psychosocial support training for teachers and school staff, additional reading materials through fixed or mobile libraries and teacher training and coaching to improve instruction. For the purposes of this study, it is important to note that the assessment tested students who were currently in schools in each of these groups. The study did not specifically sample schools with intact student populations that had participated in the interventions from the start.

²⁰ The Humanitarian Needs Overview for Syria is produced annually by the United Nations Office for the Coordination of Humanitarian Affairs. More information can be found at <https://www.unocha.org/syria>.

Table 4. Description of Manahel Intervention Waves

School type	Description	Starting Date	Total students in group	Average number of moves due to war	Average number of schools attended
Wave A	A school that received reading intervention for the full 2019 spring semester and both reading and mathematics for summer 2019, through teacher training and direct implementation (teaching & learning materials)	December 2018	352	3.1	2.2
Wave B	A school that received reading intervention for 1 month in spring 2019 and both reading and mathematics for summer 2019	March 2019	120	2.1	2.3
Wave C	A school that received reading and mathematics during summer 2019 (5 - 8 weeks)	July 2019	51	4.1	2.5
Wave D	A school reading and mathematics for summer 2019 but does not have a library on the site; the school receives a weekly visit from a mobile vehicle that is equipped with storybooks	July 2019	250	3.1	2.4

As part of the formative assessment by teachers at QE schools, Manahel developed internal levels of reading progression for teachers to identify where their students were. The eight levels — beginning at a student’s ability to identify the names of letters and continuing through their ability to read paragraphs with comprehension — allowed teachers to better understand students’ progress and adapt instruction and support as needed.

Study Purpose and Research Questions

The 2019 Manahel Learning Assessment examined the performance of students attending Manahel-supported schools in reading and mathematics. The target audience for this study was Manahel and legacy DFID (FCDO). The study sought to provide both Manahel staff and FCDO with insights on factors influencing student performance to ensure the project met the needs of the schools and students it serves. The results of the study also serve as a point of reference for comparison to the previous project’s 2017 study and a 2021 endline study, as well as provide recommendations for the endline study design.

School-to-School International (STS) partnered with Manahel to conduct this 2019 study. The study draws on data collected from Grade 3 students at the start of the 2019/2020 academic year in October and November 2019. Data collection captured student and school data using three tools: the EGRA, which measures student performance on the foundational skills required for reading fluency; the EGMA, which measures student performance on the foundational skills of mathematics; and a Student Stressors Survey, which examines students' experience and current feelings of stress during the conflict. Head teachers provided school enrolment and attendance data to allow for analysts to appropriately weight the data.

Six research questions guided the study:

1. What proportion of Grade 3 students are classified as 'progressing' and as 'proficient' readers? How do these proportions compare by subgroups?
2. How do Manahel students' learning outcomes at the beginning of Grade 3 in reading and mathematics compare to those for students assessed under the previous project in 2017 at the end of Grade 3?
3. How do beginning of Grade 3 students' learning outcomes in reading and mathematics in QE schools compare with that of AO schools? Within QE schools, how do schools with fixed libraries compare with those with mobile libraries?
4. How do beginning of Grade 3 students' learning outcomes in reading and mathematics compare between Province A and Province B?
5. How do beginning of Grade 3 students' EGRA findings relate with the eight reading levels that Manahel-supported teachers are using to track learners' progress? What is the relationship between Manahel- developed reading levels and EGRA performance?
6. What do we know about children's current levels of stress and the relationship with their learning?



Methodology

This section details the methods used to answer the study's research questions, including sampling, tool development, data collection, data analysis and the study's limitations.

Sampling

For the 2019 study, Manahel and STS chose to assess Manahel students at the beginning of Grade 3. This placed the study in closer alignment with the logical framework of the Manahel project, as well as the international EGRA recommendations that tailor EGRAs to end of Grade 2 students. As a result, the study was no longer comparable to the 2017 study under the previous project wherein students were assessed at the end of Grade 3 and thus had received additional months of schooling than students tested in 2019. However, this decision means that the comparison between the 2019 and 2021 at endline will be appropriate, given the Manahel interventions that begin in Grade 1, as well as the consistent EGRA tool.

The sample size was determined using G-Power and Optimal Design, assuming a two-stage sampling design.²¹ The sample size required to compare means between two groups — assuming an alpha level of 0.05, Minimum Detectable Effect Size (MDES) of 0.25 and a power of 0.80 — is 620 students. Factoring in a design effect of 1.80, increased the required sample size to 1,116 students — or 10 students in each of 112 schools. STS and Manahel increased the desired sample to 150 schools as that allows for 25% attrition rate by endline. Allowing for an attrition was important given that access to schools and students within schools during conflict zones is uncertain.

Selection of schools was undertaken using a two-stage approach. At the first stage, STS randomly selected several schools per province and Manahel school type that was proportional to the full Manahel school list. At the second stage, 10 Grade 3 students — five boys and five girls — were randomly selected to complete the learning assessments within the target classroom. If a school had more than one class of Grade 3, data collectors randomly selected one classroom in which to identify the 10 students.

²¹ G-Power and Optimal Design are software programs used for sampling.

Manahel’s desired sample of 150 schools represented one-third of the 450 schools supported by Manahel in academic year 2019/2020. The target sample size of 150 schools and 1,500 students was proportionally divided by province and school type. Table 5 provides a summary of the target and final sample by province, school type and gender.

Table 5. Target and Final Sample

Province	School type	Total schools	Target sample				Final sample			
			Schools	Students			Schools	Students		
				Total	Boys	Girls		Total	Boys	Girls
A	Quality Education	216	72	720	360	360	77	763	363	400
	Access Only	108	36	360	180	180	30	301	174	127
B	Quality Education	84	28	280	140	140	29	286	137	149
	Access Only	42	14	140	70	70	13	129	75	54
Total		450	150	1,500	750	750	149	1,479	749	730

The sample was drawn to be generalisable at the province level and at the school-type level. Any results at lower, subgroup levels will be associated with lower levels of confidence. While the targeted number of boys and girls to be assessed varied within province and school type, the numbers in the final sample do not deviate significantly. Therefore, results by gender are valid.

REPLACEMENT PROCEDURES

In addition to the sample, a robust list of replacement schools was created to allow for unforeseen challenges. During data collection, enumerators discovered that several Manahel-supported schools within the sample had closed or merged with another whilst others were inaccessible due to the conflict. For each closed, merged or inaccessible school, the study team selected a comparable school from the list of replacement schools to assess. These replacements were documented and tracked throughout data collection to ensure their appropriateness.

FINAL SAMPLE

Enumerators ultimately assessed students at 154 schools. During the data cleaning process, records from schools were dropped from the sample – two schools were following an accelerated learning programme and, therefore, were not comparable, and three schools were not a sample or replacement school and were visited accidentally. Of the remaining 149 schools, 132 schools came from the original sample whilst 17 were replacements. The final sample, outlined in Table 5, is made up of nearly three-quarters of students from Province A and one-quarter from Province B.

Assessment Tools

The Manahel learning assessment builds on previous early grade reading and mathematics research conducted within Syria and the broader region. In 2017, Manahel's predecessor project conducted an EGRA and EGMA, working with the SIG's Ministry of Education and EDs to review existing Arabic EGRA and EGMA tools and reports from neighbouring countries. They determined that the EGRA, EGMA and SSME tools created for the United States Agency for International Development's (USAID) 2012 MAHARAT project in Iraq were most relevant for the Syrian context. The previous project had added a War Stressors Survey to their 2017 study to assess the conflict's effect on learning outcomes. The previous project used these four tools in April and May 2017 in Province A, Province B, and Province C with students at the end of Grade 3.

To maintain consistency with the study conducted by the previous project, the Manahel learning assessment used the same EGRA and EGMA tools. However, Manahel did make changes to the tools to improve their quality and, thus, the accuracy of students' results. This section describes the revisions and finalisation for the 2019 assessment.

TOOL UPDATES

The largest change to the tools was the method of administration. Under the previous project, the EGRA and EGMA were completed with paper assessments and timers. Manahel chose to collect the data electronically on tablets via the Tangerine® data collection software to ensure more accurate scoring and better overall data quality.²² This change required extensive updates to all instructions. STS updated the instructions for the tablet administration in line with the Early Grade Reading Assessment Toolkit, Second Edition²³ and the Early Grade Mathematics Assessment Toolkit.²⁴ The Manahel team translated the instructions from English to Arabic and reviewed the translations for quality. Lastly, STS electronically programmed the Arabic tools into Tangerine®.

Manahel also decided not to administer the SSME surveys and War Stressor Survey from the previous project. The SSME surveys did not directly address any of the research questions, so they were removed from the assessment. Instead, Manahel kept only a few student questions on basic demographic data — such as gender, age, home language and school attendance during the past five days. The War Stressors Survey conducted under the previous project was deemed to be too lengthy and too blunt for the children — and there was concern that it could be upsetting for students. As a result, a Student Stressors Survey of 12 questions was created. It was adapted from Save the Children's internal survey on school safety in Northwest Syria.²⁵ A copy of the Student Stressors Survey appears in Annex B.

Whilst the EGRA and EGMA subtasks remained consistent, evaluators made item-level revisions within the letter sound identification, nonword reading, oral reading fluency (ORF and word problems subtasks).

²² Tangerine® is an open-source software developed by RTI International specifically for the administration of EGRA and EGMA.

²³ RTI International, *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*. Washington, DC: United States Agency for International Development, 2015.

²⁴ RTI International, *Early Grade Mathematics Assessment (EGMA) Toolkit*, 2012.

²⁵ Save the Children, 'A Better Tomorrow: Syria's Children Have Their Say,' 2019.

For letter sound identification, Manahel staff reported that nearly all 100 letters within the previous project’s grid lacked a modifier — this would allow students to give multiple sounds as correct responses. Manahel staff added a mix of modifiers to 70 letters to limit correct responses to one answer. In 19 cases, only the modifier was present in the grid. For those items, Manahel added a letter to the provided modifier to allow the students to identify the corresponding sound. Similarly, within the ORF passage, appropriate modifiers were added to twelve of the words. These modifiers did not change the word or its meaning but only clarified the pronunciation for the students.

Evaluators also made changes to nonword reading items. Manahel staff reported that some words within the grid did not follow the linguistic rules of Arabic, while others were real words. Manahel replaced 34 of the original nonwords with new nonwords that followed the linguistic rules of Arabic. They also added modifiers to some nonwords for clarity. More details of changes made to the EGRA tools can be found in Annex C.

With the EGMA, the only changes made to the tool were within the word problems subtask. Manahel staff simplified the vocabulary within the word problems to a Grade 3 level. This impacted six words but did not change the meaning of the word problems.

Finally, Manahel and STS decided to incorporate accommodations within the administration of the EGRA and EGMA due to the prevalence of disabilities within the wider student population. These accommodations included lengthening the maximum time allowed for timed subtasks from one to two minutes, printing student stimuli in large-print font and allowing for prompts to be given up to three times if requested.

SUMMARY OF TOOLS

The final assessment included three tools administered with the Grade 3 students — EGRA, EGMA and Student Stressors Survey — and one survey administered once per school with the head teachers.

Table 6 presents a summary of both the EGRA and EGMA subtasks.²⁶ Each subtask measures a foundational skill of reading or mathematics to determine where a student may be in their progression towards fluency. Based on results, analysts computed average accuracy scores for all subtasks and fluency rates for timed subtasks. Accuracy indicates whether students are mastering the skill, while fluency indicates whether they can complete the task efficiently enough to support comprehension.

In addition to the EGRA and EGMA subtasks, each student was given a Student Stressors Survey. This survey asked students questions about their experience of the conflict — including questions on family size, displacement, and feelings of stress. Lastly, in order to properly apply sampling weights to the results, head teachers completed a brief survey on student enrolment and attendance.²⁷

²⁶ Final EGRA and EGMA tools are included in Annex B.

²⁷ The headteacher survey is included in Annex B.

²⁸ Additionally, learners who did not correctly answer any items on the addition or subtraction level 1 subtasks were not asked items from the corresponding level 2 subtask.

Table 6. Summary of EGRA and EGMA Subtasks

Tool	Subtask	Thematic Skill	Purpose	Administration	Scoring
EGRA	Letter sound identification	Mechanics of Reading	Alphabet knowledge	Timed – two minutes; autostop after first 10 items	Accuracy (% correct) and fluency (correct letter sounds per minute (CLSPM); 100 items total
	Nonword reading	Mechanics of Reading	Decoding	Timed – two minutes; autostop after first 5 items	Accuracy (% correct) and fluency (correct nonwords per minute (CNWPM); 50 items total
	Oral reading fluency	Mechanics of Reading	Decoding and reading fluency	Timed – two minutes; autostop after first 11 items	Accuracy (% correct) and fluency (correct words per minute (CWPM); 82 items total
	Reading comprehension	Comprehension	Reading comprehension	Untimed; number of questions asked corresponds to how many words read in oral reading fluency passage	Accuracy (% correct); five items total
	Listening comprehension	Understanding	Oral language comprehension and vocabulary	Untimed; all questions asked of all respondents	Accuracy (% correct); six items total
EGMA	Number recognition	Whole numbers	Numerals and numericities identification	Timed –two minutes; no autostop	Accuracy (% correct) and fluency (correct numbers per minute (CNPM); 20 items total
	Quantity discrimination	Whole numbers	Numerical magnitudes comparisons	Untimed; autostop after four consecutive incorrect items	Accuracy (% correct); 10 items total
	Missing numbers	Whole numbers	Number patterns identification	Untimed; autostop after four consecutive incorrect items	Accuracy (% correct); 10 items total
	Addition (level 1)	Operations	Arithmetic skills	Timed – two minutes; no autostop ²⁸	Accuracy (% correct) and fluency (correct addition problems per minute (CAPPM); 20 items total
	Addition (level 2)	Operations	Arithmetic skills	Untimed; no autostop; only administered if respondent correctly answered at least one item correct on Addition level 1	Accuracy (% correct); 5 items total
	Subtraction (level 1)	Operations	Arithmetic skills	Timed – two minutes; no autostop	Accuracy (% correct) and fluency (correct subtraction problems per minute (CSPPM); 20 items total
	Subtraction (level 2)	Operations	Arithmetic skills	Untimed; no autostop; only administered if respondent correctly answered at least one item on Subtraction level 1	Accuracy (% correct); five items total
	Word problems	Real world problems	Conceptual and real-world mathematics understanding	Untimed; autostop after four consecutive incorrect items	Accuracy (% correct); six items total

Data Collection

This section outlines the process used to collect data in the Provinces A and B.

ENUMERATOR TRAINING

For the data collection, twenty enumerators were selected internally – 10 Syria-based Manahel staff members and 10 staff members from a local Manahel partner. An external data collection firm was not used for this study due to child safeguarding, security, accessibility, time and budget constraints. Prior approval from the SIG EDs was required to access schools and the Manahel and Partner Organisation 1 staff were granted approval. For the child safeguarding and security considerations, Manahel and Partner Organisation 1 staff visited these schools regularly and had the training and resources to identify and mitigate risks for both the children and themselves. Lastly, it was more cost and time effective to use project and partner staff.

Unlike the previous project, wherein the enumerator training was conducted remotely via webinar in English with an interpreter, Manahel's enumerator training followed a training-of-trainer (TOT) cascade model. The TOT took place remotely over Skype on 9 and 10 October 2019 for 4.5 hours each day. This was in English and Arabic, with the subsequent training to enumerators in Arabic. The STS team based in the United States trained four Manahel trainers on data collection procedures. Manahel's project leads for the two arms of the Manahel intervention—the QE Lead and the AO Lead. The TOT was given in English with translation provided by the Manahel project leads. The TOT reviewed the enumerator training agenda and materials, the administration protocols for the assessments and Student Stressor Survey, the use of the Tangerine® application on the tablets, the preparation for the data collection and potential challenges.

The Manahel trainers, in turn, trained twenty enumerators in Arabic with materials provided by STS from 14 to 16 October 2019 in Province A. The enumerator training lasted three full days and was conducted by two in-person trainers in Province A with remote support from an additional two = trainers.

The enumerator training included:

- Specific training on:
 - The purpose of the learning assessments
 - The administration of the learning assessments and survey
 - The use and management of tablets
 - Uploading data
 - Respecting child protection and ethical considerations
- A series of hands-on exercises and group discussions to equip participants to conduct learning assessments, navigate challenges during data collection and complete other data collection tasks
- Detailed plans for data collection
- Safeguarding and 'Do No Harm' issues were also discussed

DATA COLLECTION

Between 19 October 2019 and 6 November 2019, enumerators visited 154 Manahel schools. Twenty enumerators were divided into 10 pairs for the school visits. Each pair visited one school per school day and assessed 10 Grade 3 students. Enumerators uploaded data daily from their tablets via wi-fi to a secure, password-protected server maintained by STS staff.

SUPERVISION AND QUALITY CONTROL

Throughout the three and a half weeks of data collection, enumerators were closely supervised to ensure data quality. The Manahel AO Lead provided remote supervision and tracked the progress of the data collection daily. The Manahel trainers in Syria performed site visits to ensure enumerators were following protocols. STS monitored the data daily by checking results uploaded to the server for completeness. STS worked with Manahel to maintain detailed documentation of all issues encountered. More generally, tablets used for electronic data capture improved data quality, consistency and collection efficiency as it streamlined fieldwork and reduced the number of measurement and data entry errors.

An additional measure of data quality control was the use of inter-rater reliability (IRR) measures during data collection.²⁹ Per standard EGRA practice, IRR was conducted with 10% of the sampled students. Each day, both enumerators would assess the first student together — one enumerator acting as the assessor and one as the observer — and their results were compared by STS. STS recorded no instances of IRR less than a 95% rate of accuracy between enumerators.

CHILD PROTECTION AND RESEARCH ETHICS

The study tools were reviewed by the Manahel team prior to the beginning of data collection to ensure that the study adhered to applicable ethical rules and societal norms. All enumerators received training on the project's code of conduct and its child protection policies and procedures.

Affirmative informed consent was obtained from all head teachers and classroom teachers to assess the children in their care. All children provided affirmative assent to be assessed and interviewed and could opt out of the assessment at any time. Assent was acquired a second time from each child before starting the sensitive Student Stressors Survey.³⁰

Students were selected randomly on the day of the data collection. Information on students with disabilities could not be obtained prior to data collection to be incorporated into the sampling plan. Consequently, during the data collection, students with disabilities were not excluded from the random sampling procedure or from participating. As a result, accommodations for students with disabilities — such as extended time limits for the timed subtasks and large print stimuli — were built into the administration protocols of the assessments and were given to all students throughout data collection.

²⁹ Intra-rater reliability is the degree of agreement between two enumerators who are assessing the same student independently. It allows the data collection monitors to identify and resolve problems within the enumerator teams during data collection to improve the quality of the data collected.

³⁰ 55 children opted out of completing the Student Stressors Survey and 88 children opted out of completing the ORF subtask.

Data Analysis

DATA CLEANING AND EDITING

Upon completion of the data collection, STS cleaned the data to remove practice cases and any invalid data. This ensured that the final data set was complete, accurate and internally consistent. STS followed a multistage data cleaning plan to ensure all data values were within the allowable range. STS developed a master codebook and merged EGRA and EGMA data sets.

WEIGHTING

The STS team applied sampling weights to the students' data to produce more representative estimates in the sample. To compute sampling weights, STS used the following information about all the schools in the relevant population:

- Type of school
- Province
- Number of students enrolled in Grade 3
- Number of students in attendance in Grade 3

This data was collected through the school's head teacher survey at the beginning of each school visit. Weights were computed using SPSS version 25.

GENERATION OF FINDINGS

After applying the weighting functions, STS produced descriptive statistics that were disaggregated by variables of interest. Descriptive results were analysed for statistically significant differences by gender, province and school type using chi-square and t-tests. The chi-square test is a statistical test comparing the proportion of students who did not respond correctly to any items on a subtask — known as zero scores — with what was expected. The independent-sample t-tests compare the difference between the means of two independent groups on the same dependent variable. Associations between respondent characteristics and student performance were further analysed using Pearson bi-variate correlations and logistic regression to identify predictors of student performance. All analyses were conducted using SPSS version 25.

Challenges and Limitations

This section outlines the challenges and limitations that should be kept in mind when reviewing results.

LIMITATIONS DUE TO CONFLICT

- **High levels of displacement within the population.** With the population in flux, it is important to remember that the assessment tested individual students, not schools. Selected students may have transitioned between, or into and out of, QE and AO schools. While students reported moving frequently as a result of the war, the survey did not collect data on the length of time students had been in a Manahel-supported school, or which interventions they had received at that school. It also did not examine the displacement levels of teachers. Therefore, the division of school types is not a clean representation of the interventions that students and teachers may have received.
- **Impact of heightened violence on students' performance and participation.** Given the uptick in violence during the weeks of data collection, the war presented significant challenges to enumerators and students alike. Several enumerators reported that children were distracted during the assessment due to active airstrikes nearby. One school closed early due to the security situation, so enumerators could not assess all 10 selected students. Despite the care given to ensuring the tools were conflict-sensitive, some students had emotional reactions to questions that had seemed benign prior to the data collection. Several students refused to participate in some subtasks or the Student Stressor Survey.³¹ This finding was shared with Manahel at the time; they in turn notified legacy DFID (FCDO).
- **Conflict made some areas inaccessible for enumerators.** Schools in these areas were excluded from the sample by necessity.

DESIGN LIMITATIONS

- **Lack of comparability between the results of the assessment conducted by Manahel's predecessor project and Manahel assessments' results.** Despite the original intention to compare the two assessments, several barriers arose during the design and implementation of the study. One of the most critical limitations is the large difference in the academic year progression between the sampled students in the two studies — the previous project assessed Grade 3 students at the end of the school year while Manahel assessed Grade 3 students at the beginning of the school year. Furthermore, the changes made to the EGRA and EGMA tools prevent a direct comparison of results. In addition to item-level changes, the mode of administration changed from paper-based scoring to tablet scoring for the EGRA and EGMA assessments, which may have changed the accuracy of the results. Consequently, the results of the two assessments are not comparable.

³¹ Under the standard EGRA student assent protocol, students may choose not to participate in the full assessment, a subtask of the assessment, or individual items. During this data collection, 88 students chose not to complete the oral reading fluency subtask and 55 students chose not to complete the student stressors survey.

- Differences in training between the previous project’s assessment and Manahel assessments.** In addition to the design and tool differences, there were also differences between the enumerator training methods used under the previous project and Manahel. Under the former, 48 enumerators were trained remotely over Skype by an English-speaking trainer with simultaneous interpretation provided by the previous project’s staff. These enumerators assembled in two locations in Syria and frequently experienced technical difficulties. The remote training approach also limited the trainer’s ability to monitor enumerators’ progress and accuracy during training. To resolve these challenges, Manahel followed a Training-of-Trainers (TOT) approach to eliminate the need for concurrent interpretation and instead had in-person, Arabic-speaking trainers with the 20 enumerators in Syria. This allowed the trainer to more effectively facilitate and gauge the enumerators’ mastery of administration protocols.
- Inability to pilot revised tools. Manahel and STS conducted a thorough technical review of the tools to improve upon the accuracy of the assessments as described above.** Under typical circumstances, the number of changes made to the tools would have prompted a pilot of the revised versions. However, due to constraints within the context and timeline, revisions to the tools were not able to be piloted with students. As a result, the revision process relied on the expertise of Manahel staff members alongside STS to identify and resolve problems with the tools prior to data collection.
- Structural bias in the comparison of scores of students in ‘Access Only’ schools to students in ‘Quality Education’.** For the most effective use of donor funding, the Manahel project targets potentially struggling schools — based on several contextual factors, such as available resources, accessibility and need — with their robust QE intervention. Consequently, the AO and QE group designations were not created for evaluation purposes. However, at legacy DFID’s direction, evaluators compared results for students in these two existing school types to provide a proxy measure of the difference in learning gains between the robust QE interventions and the minimal support AO interventions. The extent to which this comparison does provide a proxy measure is confounded with the selection criteria of which schools receive AO interventions and which receive QE interventions. Upon further discussion with the Manahel team, STS determined that the comparison between AO and QE provides a very limited view into how these two interventions compare. Any differences in the groups’ results cannot be attributed solely to the different interventions, as students and schools are likely also influenced by the varied contextual factors — including those Manahel used to assign the schools’ interventions in the first place.
- Potential bias of using project and partner staff as enumerators.** Members of the Manahel staff and their partners served as enumerators for this assessment. Typically, an outside data collection firm would be used to limit potential bias within the assessment. However, the decision was made to use internal program staff and partners as they already had the requisite permissions needed to access schools as well as familiarity with child safe-guarding practices. This allowed for efficiency within the short timeline and cost restrictions for the assessment. Furthermore, STS monitored the incoming data daily throughout data collection and felt confident in its quality and accuracy.

There were no departures from the study design TOR or original evaluation design.

Findings

This section reports findings by the study's main research questions. Results that are statistically significant at the $p < 0.05$ level are referred to as 'significantly' lower or higher in the text.

Description of the Sample

As previously described, 1,479 students participated in the 2019 assessment, which included an EGRA, EGMA and Student Stressors Survey. The survey was administered to 96.3% of students who took the EGRA and EGMA; 55 students chose not to answer survey questions. The sample was equally balanced between boys and girls (50.2% and 49.8%, respectively), and all students reported speaking Arabic at home. The average age of Grade 3 students was 9.2 years old, and most were between eight and 11 years old. More students in Province A were on-age for Grade 3 than in Province B (52.7% and 36.0%, respectively). Students generally came from large households — 7.3 people, on average.

Two-thirds of students said they were forced to move because of the war—56.5% in Province A and 81.4% in Province B. Additionally, students in District 3 were most likely to report being displaced due to the war, followed by students in District 2 and District 4 and District 1. Students in Districts 5 and 6 were more likely than students in District 1 or District 4 to be displaced five or fewer times. The majority of students moved between one and five times — 51.5% in Province A and 65.9% in Province B. One-third (35.6%) of students attended between one and three new schools and two-thirds (64.4%) attended four or more new schools. Still, students report very regular attendance; only 22.8% in Province A and 16.8% of students in Province B reported missing school in the last five days. Overall, 78.9% of students reported that they had not missed any school.

Half of students reported experiencing hunger at school — 48.3% in Province A and 59.7% in Province B. One-third (37.5%) reported tiredness. There were no notable difference in these rates between boys and girls; however, far more QE students reported hunger or tiredness than AO students (74.0% and 72.0% versus 26.0% and 28.0%, respectively).

Nearly all students reported feeling safe at home, at school, and on the way to and from school. On average, students travel 11.7 minutes to get to school, during which time more girls than boys felt unsafe — 12.6% versus 9.6% and 14.3% versus 7.2% in Province A and Province B respectively.

Results in the following sections are disaggregated by gender and by province or school type, depending on the research question. See Annexes F, G and H for additional results by district. Analyses by disability and age are not presented because of insufficient data on children with disabilities and because students aged eight, nine or 10 did not have statistically significantly different results from each other. Results by wave are not included due to the high displacement rates of students — and presumably teachers — which renders interventions at the school-level mixed due to the mix of individuals in those buildings at any given point. Additionally, waves were used to identify schools based on their intended interventions, but unique circumstances for individual schools may have warranted additional interventions as needed and lastly, non-Manahel interventions are not accounted for in the wave notations.

Trends are noted using statistically significant differences between groups to illustrate the magnitude of the expected difference in the population. At times, insignificant trends are noted, as those are indicative of an interesting or important finding. Otherwise, trends that are not mentioned were not found to be statistically significant.

Research Question 1: Student Performance in 2019

Research Question 1: What proportion of Grade 3 students are classified as 'progressing' and as 'proficient' readers?

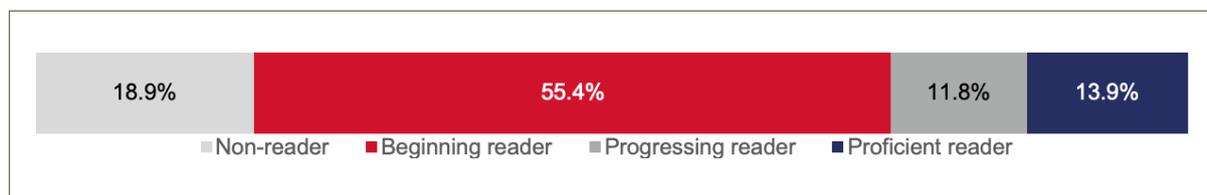
STS classified students' scores into proficiency levels using bands established during the previous project's 2017 reading assessment:

- **Non-readers:** students who did not read a single word of the story reading passage.
- **Beginning readers:** students who read between one and 22 correct words per minute (CWPM) but scored less than 80% on the reading comprehension subtask.
- **Progressing readers:** students who read 23 CWPM or more but scored less than 80% on the reading comprehension subtask. This corresponds to 'advanced progressive reader' within the Manahel logical framework.
- **Proficient readers:** students who scored 80% or more on the reading comprehension subtask.

These categories are useful within the Syrian context and have not been aligned with international frameworks. At the country-level, tracking the proportion of students who move in and out of these categories at endline will be instrumental in understanding the changes in student's learning outcomes between 2019 and 2021. With the availability of the Global Proficiency Framework, which describes categories of learners by reading domain, Manahel may consider engaging in an exercise of aligning the categories reported above with those described in the Global Proficiency Framework for students in Grade 2.

This alignment exercise will reveal the proximity of the previous project's categories to internationally used categories, as well as underlying differences in expectations of students in each domain of reading. Overall, most students showed some reading ability (Figure 14). The majority of students (55.4%) were categorized as beginning readers; 11.8% as progressing readers; and 13.9% as proficient readers. Fewer than one in five students (18.9%) were non-readers. Additional results by subgroup can be found in Annex D.

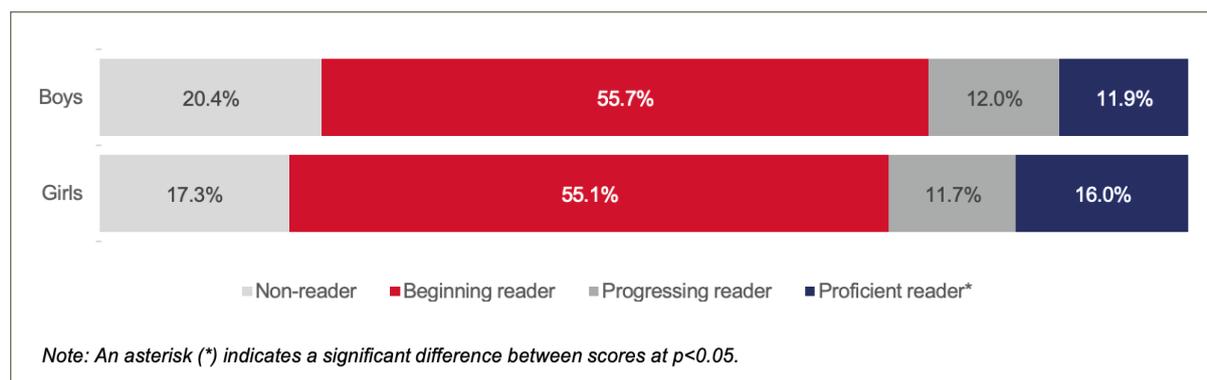
Figure 14. Percentage of Students by Reading Proficiency Level



READING RESULTS BY GENDER

The percentage of girls who are proficient readers (16.0%) was significantly higher than that of boys (11.9%). The percentages of girls and boys were comparable in all other reading proficiency categories (Figure 15).

Figure 15. Percentage of Students by Reading Proficiency Levels by Gender



For the proficiency levels, STS computed fluency rates and accuracy scores for three reading subtasks – letter sound identification, nonword reading and ORF (Figure 16 and Figure 17).

Figure 16. Reading Items Correct Per Minute (Fluency Rates) by Subtask and Gender

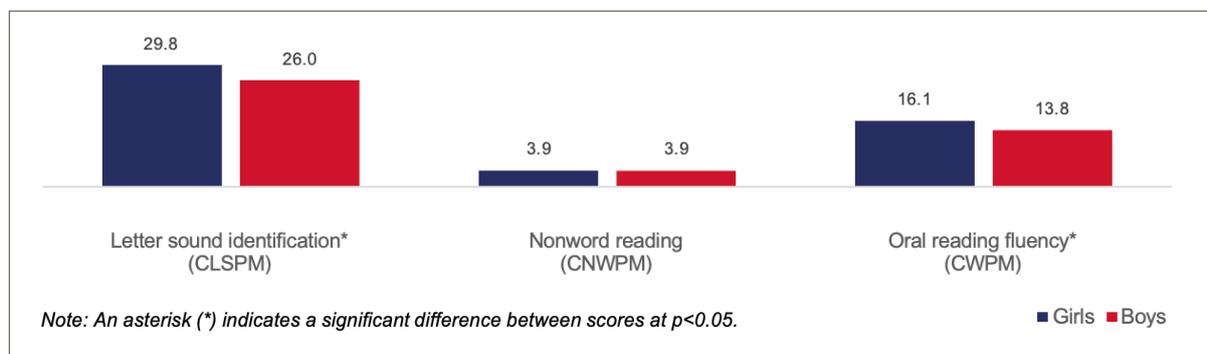
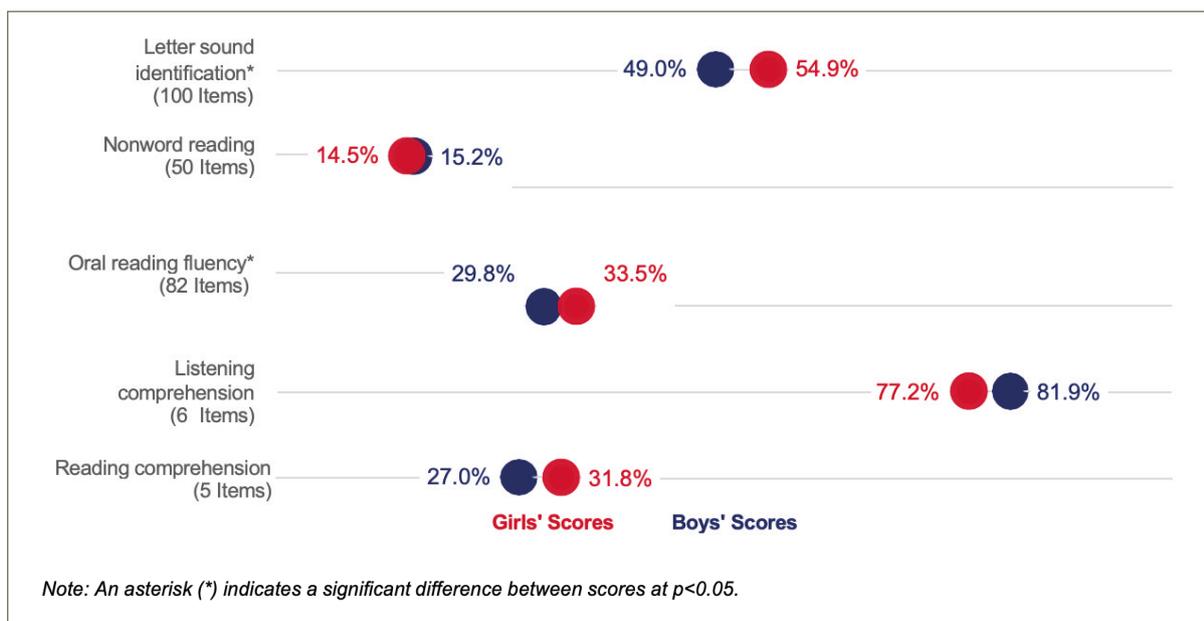


Figure 17. Reading Percentage of Items Correct (Accuracy Scores) by Subtask and Gender



Girls' fluency rates were statistically significantly higher than boys' rates on two subtasks — letter sound identification and ORF. On average, girls could identify 29.8 correct letter sounds per minute (CLSPM) and 16.0 CWPM. By comparison, boys could identify 26.0 CLSPM and 13.8 CWPM. Girls had higher accuracy scores on reading subtasks compared to boys. On average, girls answered 54.9% of the letter sound identification items correctly compared with boys, who answered 49.0% of the items correct; this difference was statistically significant. On the ORF subtask, girls answered statistically significantly more ORF items correctly than boys (33.5% correct compared to 29.8%, respectively). Girls also performed better on the reading comprehension subtask, where they averaged 31.8% of items correct while boys averaged 27.0% correct.

Notably, boys had statistically significantly higher accuracy scores on the listening comprehension subtask than girls. Boys, on average, answered 81.9% of the six listening comprehension questions correctly, compared to girls, who answered 77.2% correct, on average. Boys' higher performance was also reflected in the proportion of zero scores — almost twice as many girls as boys did not answer a single item correctly — 3.2% of girls compared to 1.5% of boys.

MATHEMATICS RESULTS BY GENDER

STS conducted analyses on mathematics outcomes by gender. In contrast to reading, boys outperformed girls in all mathematics subtasks. Boys' higher performance was reflected in their statistically significant higher accuracy scores and lower percentages of zero scores. Figure 19 shows the average accuracy score for boys and girls by mathematics subtask. On every subtask, boys had a statistically significantly higher accuracy score than girls.

Figure 18. Mathematics Items Correct Per Minute (Fluency Rates) by Subtask and Gender

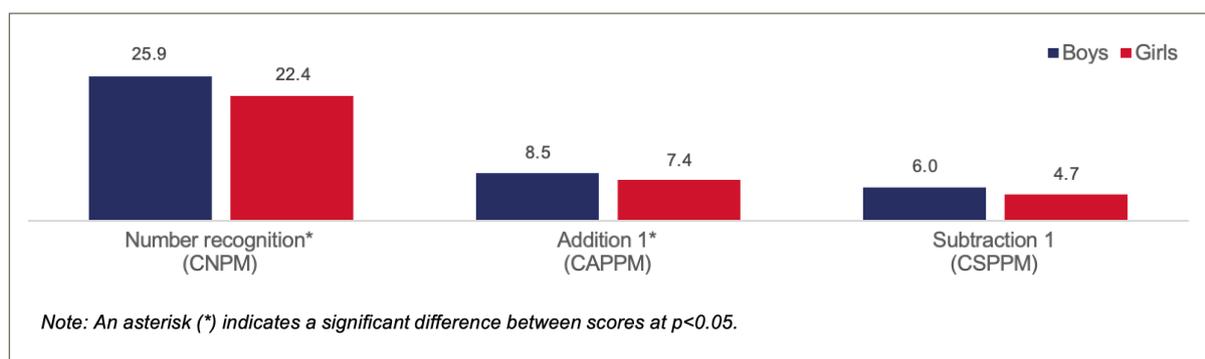
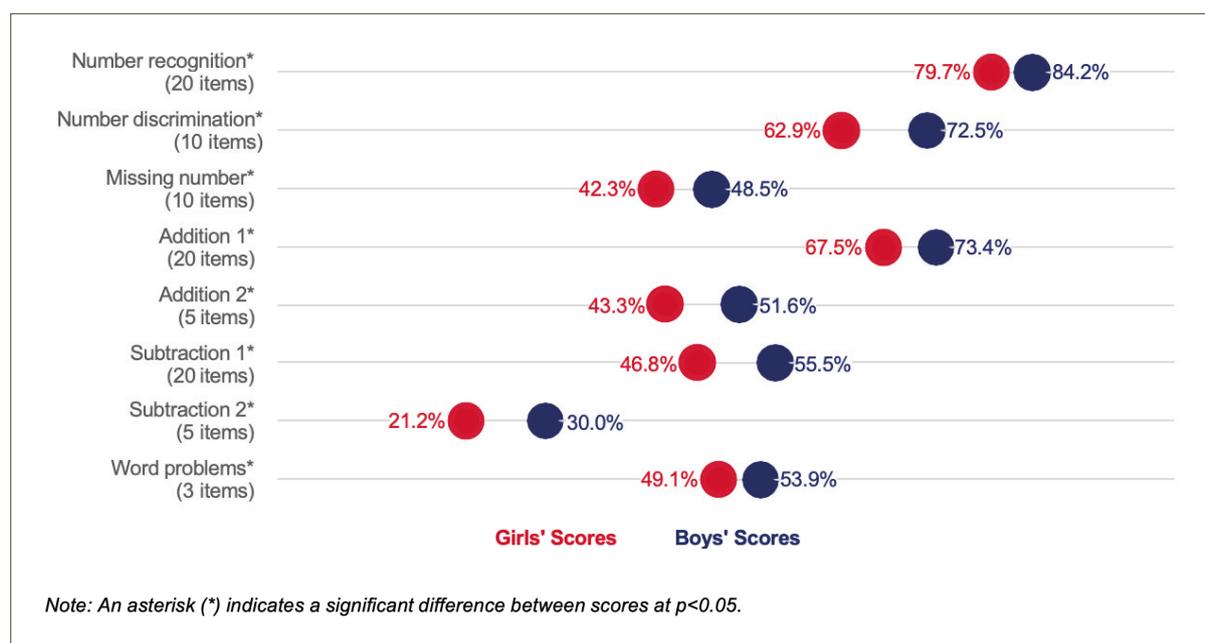


Figure 19. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Gender



Research Question 2: Comparison of 2017 and 2019 Student Performance

Research Question 2: How do Manahel students' learning outcomes at the beginning of Grade 3 in reading and mathematics compare to those for students assessed under the previous project in 2017 at the end of Grade 3?

One of the outcomes in Manahel's logframe is improved learning outcomes, which means that the project's performance will be assessed against a percentage change in EGRA scores, using the 2017 results as a baseline. However, before analysts could report the change in 2019 scores since 2017, the validity of the comparison was first assessed. STS, in consultation with the Manahel team, determined that the comparison between the previous project's 2017 results and Manahel's 2019 results are not comparable. Therefore, results are not presented in the body of this report.

Comparison between the 2017 and 2019 results is tenuous given the difference in timepoints — end of Grade 3 in 2017 and beginning of Grade 3 in 2019 — as well as the changes in the EGRA tools. During the design phase, Manahel and STS chose not to assess students at the beginning of Grade 4, which would have been more comparable to the 2017 assessment. However, assessing Grade 4 students would not be properly aligned with the project’s logical framework. Instead, Manahel chose to align the midline study with the logical framework, with the goal of ensuring a comparable endline study in 2021.³²

The decision to compare students’ performance in 2017 and 2019 was made solely due to the lack of a better alternative. It is recommended that a comparison to the 2017 results be removed from the logframe and the endline study. Instead, endline 2021 results should be compared to the 2019 results — as long as the samples are still comparable.

While the results are not presented here, the comparisons between 2017 and 2019 are provided in Annex G for purposes of the logframe which requires comparison of results to 2017. STS recommends that the logframe be updated to remove comparisons to 2017 for the previous stated reasons.

Research Question 3: Outcomes by School Type

Research Question 3: How do beginning of Grade 3 students’ learning outcomes in reading and mathematics in Quality Education schools compare with that of Access Only schools? Within Quality Education schools, how do schools with fixed libraries compare with those with mobile libraries?

The Manahel program team included a research question examining results by school type in order to see if it was possible to determine the value add of QE interventions. Under the objectives of the program, both AO and QE schools provide children with more stability and psychosocial support, which should have a positive impact on their wellbeing and learning. Beyond this, the Manahel team hypothesized that the additional support to QE schools should result in larger gains in children’s learning.

However, it was uncovered during the validation process that the labels AO and QE do not accurately reflect what has occurred in a school, and in turn, comparison of results by these labels are not reliable. With the reported high levels of displacement within the Student Stressors Survey,³³ students and teachers who have benefited from the Manahel program may have relocated, making it challenging to determine which beneficiaries have received which interventions. Additionally, the role of other NGOs, funders and education organizations in Manahel-supported schools is not tracked and therefore cannot be accounted for in the analyses. Manahel noted that other NGOs would not be supporting the same age group or their teachers. This happens in (very roughly) 10% of schools. There is no overlap that would cause an improvement.

³² This was discussed with legacy DFID and approved by them.

³³ Two-thirds of students (63.4%) reported being forced to move due to the war. The majority of those students reported moving twice. Overall, 40.1% of student reported attending a new school due to the war.

As such, the analyses by these labels—AO and QE—cannot be used to draw conclusions about the comparative impact of QE schools over and above AO schools at this time. At endline, it is recommended that a survey of teachers, head teachers and/or students be considered to capture the types and duration of Manahel interventions that these individuals have experienced. These data can instead be used to generate labels, or classifications, for analysis at endline.

As with the prior research question, results by school type are still included in this report to answer the stated research question. However, results should be interpreted with caution.

READING OUTCOMES BY SCHOOL TYPE (AO AND QE)

Reading results by school type are presented by proficiency level; then trends in accuracy, fluency and zero scores are described. Overall, students’ reading proficiency levels were comparable between students in schools labelled as AO and students in schools labelled as QE but results for individual subtasks show some significant differences. For brevity, results are reported using the terms ‘AO schools’ and ‘QE schools’ but refer to ‘students in schools labelled as AO’ or ‘students in schools labelled as QE.’

Figure 20. Percentage of Students by Reading Proficiency Level and School Type

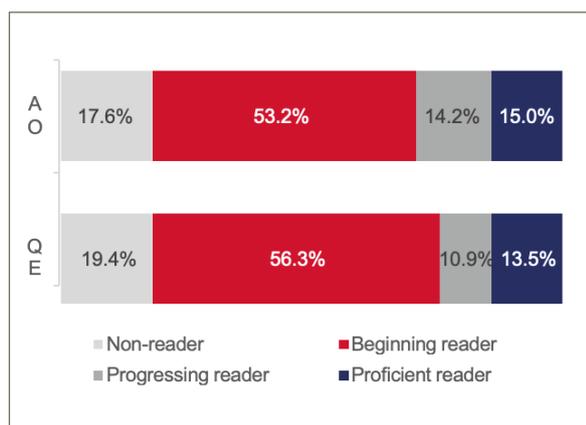
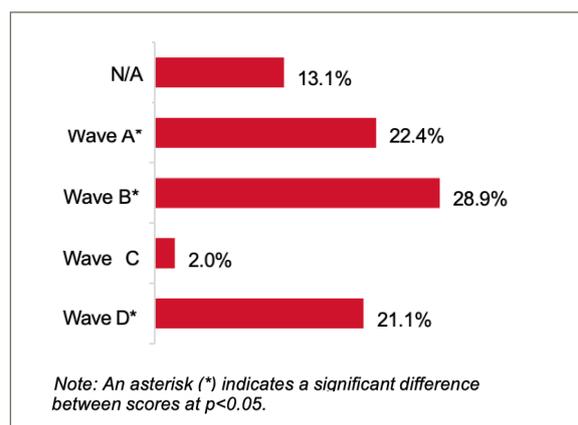


Figure 21. Percentage of Non-Readers by Wave in QE Schools



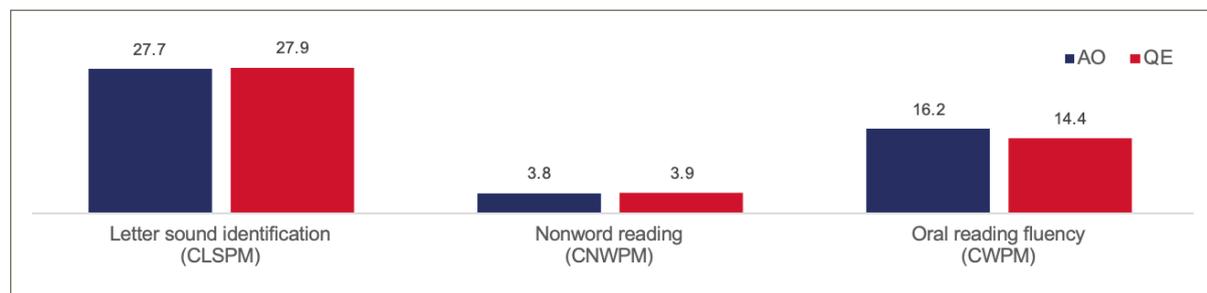
Analysis showed subtask-level differences in accuracy and zero scores by school type — AO schools have higher scores in the reading and listening comprehension subtasks than QE schools but they also had higher proportions of zero scores on letter sound identification and nonword reading subtasks (Annex F). AO students had significantly higher accuracy scores in ORF and listening comprehension than peers in QE schools; however, both groups were comparable ORF rates.³⁴ QE schools had smaller proportions zero scores than AO schools on foundational skills subtasks — letter sound identification and nonword reading.³⁵ On the three timed subtasks — letter sound identification, nonword reading and ORF — there were no significant differences by school type (Figure 22). The lower scores on the nonword reading subtask indicate that the task was difficult for students. However, students who were proficient readers had significantly higher nonword reading scores than students who were beginning and progressing readers.

³⁴ On average, AO students answered 34.2% of the 82 ORF items correctly, compared to 30.5% for QE students. AO students also answered an average of 81.9% of the six listening comprehension items correct, compared to 78.6% for QE students.

³⁵ Of AO students, 14.6% did not identify a single letter sound, compared to 10.3% of QE students. Similarly, 53.6% of AO students did not read a single nonword, compared to 45.8% of QE students.

This suggests that the task of nonword reading is more difficult for students than other tasks, but that it differentiates between students reading abilities.

Figure 22. Reading Items Correct Per Minute (Fluency Rates) by Subtask and School Type

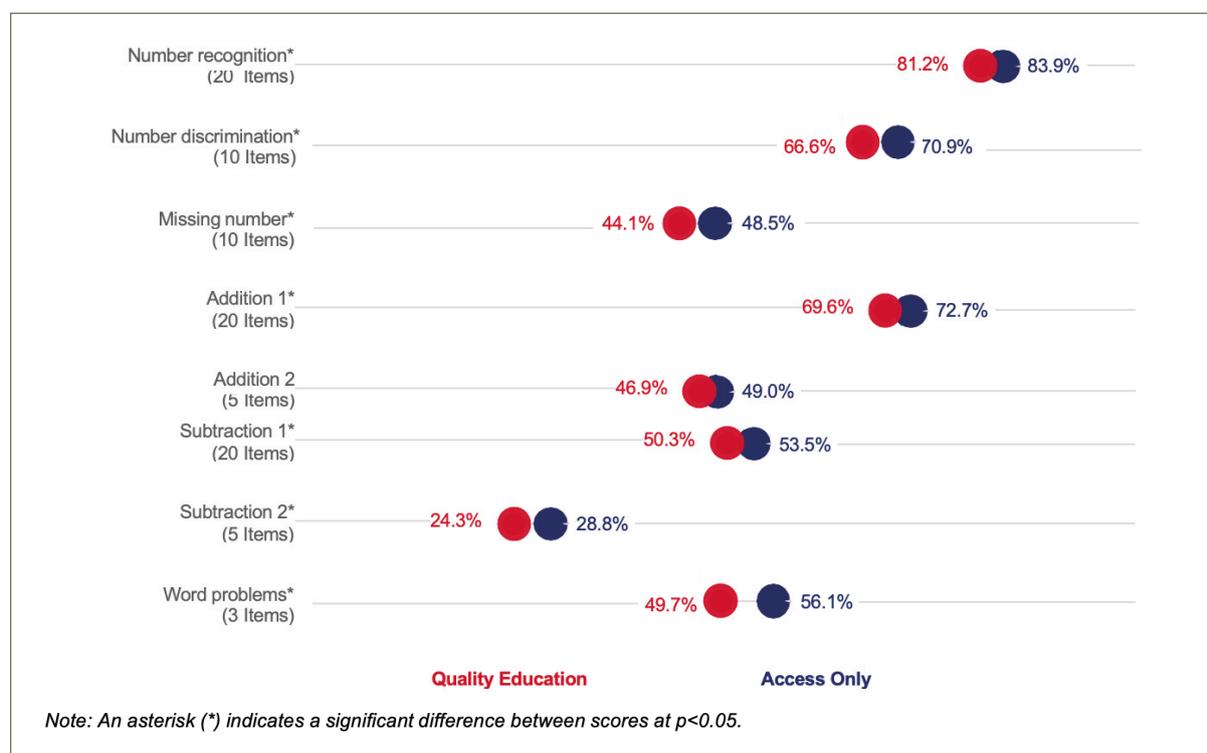


The stronger performance of AO schools is mostly driven by the performance of schools in Province A. In Province A, AO students had higher accuracy scores than QE students in Province A on three EGRA subtasks – ORF, listening comprehension, and reading comprehension (Annex E). Displacement six times or more appears to be associated with lower learning scores across all subgroups. However, students who moved fewer than five times are also more likely to be from certain districts. Therefore, it is not possible to disaggregate the effect of displacement from district.

MATHEMATICS OUTCOMES BY SCHOOL TYPE (AO AND QE)

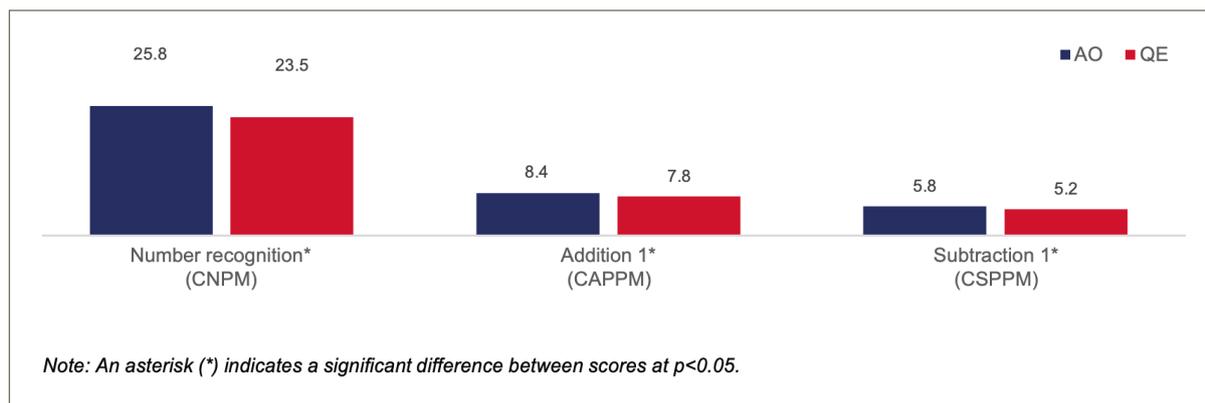
On EGMA subtasks, students in AO schools had higher average accuracy scores than QE peers on all subtasks except addition level 2 (Figure 23).³⁶ AO schools also had higher average fluency rates on all three timed subtasks (Figure 24).

Figure 23. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and School Type



³⁶ There were two additional subtasks. The first (Addition 1) had 20 questions, administered to all students. The second (Addition 2) had five questions and was only administered to students who answered all of Addition 1's first 10 questions correctly.

Figure 24. Mathematics Items Correct Per Minute (Fluency Rates) by Subtask and School



Type AO schools also performed better in terms of zero scores; QE schools had significantly higher proportions of zero scores on three EGMA subtasks – addition level 2 (15.2% versus 10.1%), subtraction level 2 (40.6% versus 32.6%) and word problems (16.5% versus 10.5%).

As with reading performance, the stronger performance of AO schools is mostly driven by the performance of schools in Province A. In Province A, AO students had higher fluency rates than their QE peers on number recognition fluency, addition fluency and subtraction fluency. Province A AO students also had significantly higher accuracy scores on five EGMA subtasks – number recognition, missing number, subtraction 1, subtraction 2 and word problem subtasks (Annex E).

READING AND MATHEMATICS OUTCOMES WITHIN QE SCHOOLS (QE AND QE-MOBILE LIBRARIES)

Overall, students in QE schools with a mobile library had comparable accuracy scores to students in QE schools with a fixed library; scores differed significantly on only two EGRA subtasks – letter sound identification and ORF. On the EGRA, students in fixed-libraries schools had higher letter sound identification fluency rates than students in mobile-libraries schools (28.6 CLSPM and 25.2 CLSPM, respectively). In contrast, students in mobile-library schools had higher ORF scores than fixed-library peers (17.5 CWPM and 13.7 CWPM, respectively).

At the district level, students in fixed-library schools generally performed significantly higher than students in mobile-library schools in 2 districts – District 2 and District 3 – but the opposite was true in one district – District 4 (Annex E). In District 2, students in fixed-library schools had a significantly higher fluency rates on letter sound identification and higher accuracy scores on letter sound identification, ORF, number discrimination and addition level 1. In District 3, students in fixed-library schools had significantly higher accuracy scores on listening comprehension only. In District 4, students in mobile-library schools had higher accuracy scores on four reading subtasks – letter sound identification, nonword reading, ORF and listening comprehension – and three mathematics subtasks – missing number, addition level 1 and subtraction level 2.

As the assessment did not directly ask questions about library use, these results are difficult to contextualize and interpret. The endline in 2020 should include questions about student library use which may further elucidate these findings, but program monitoring data might begin collecting such information to monitor trends in resource usage over time as well.

READING AND MATHEMATICS OUTCOMES IN QE SCHOOLS BY GENDER

By gender within QE schools, there is a higher proportion of girls than boys who are proficient readers. Girls also had higher accuracy scores than boys on three of the five subtasks – letter sound identification, oral reading fluency and reading comprehension. Boys had higher accuracy scores than girls on listening comprehension. No significant differences were observed between boys’ and girls’ accuracy scores on the nonword reading task (Figure 26). In mathematics, boys had higher accuracy scores than girls on all subtasks (Figure 27).

Figure 25. Percentage of Students in QE Schools by Reading Proficiency Level and Gender

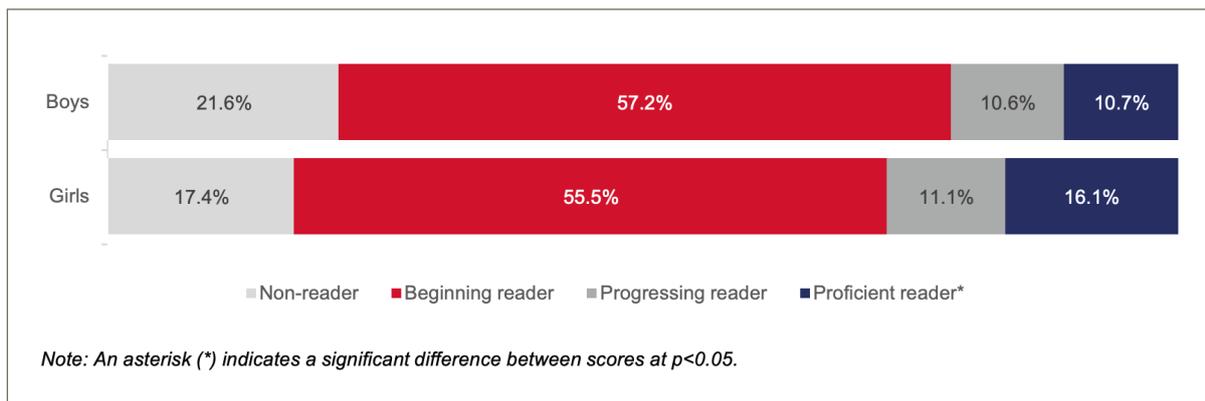


Figure 26. Reading Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender

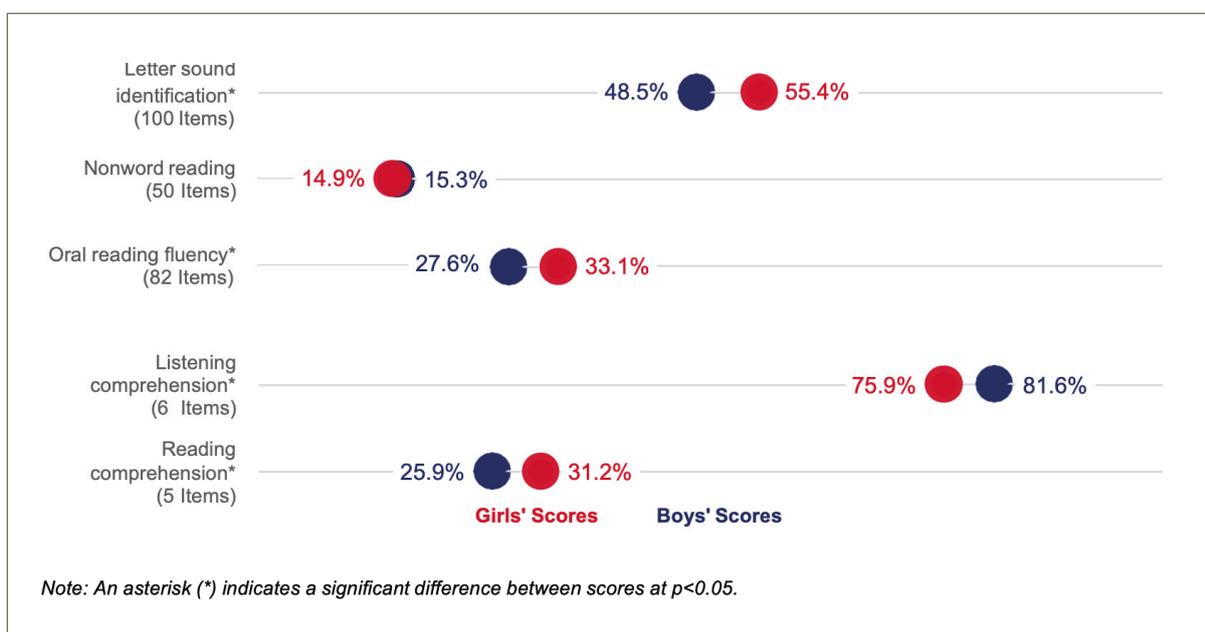
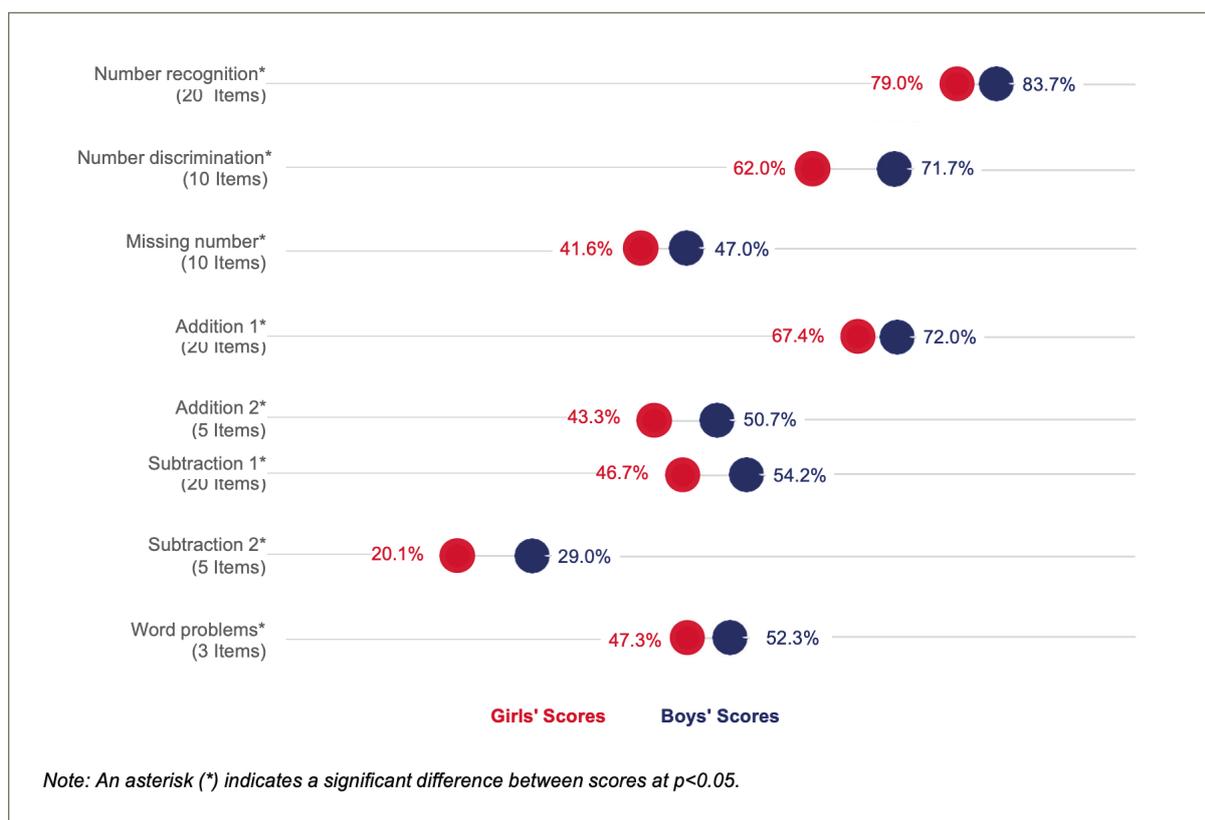
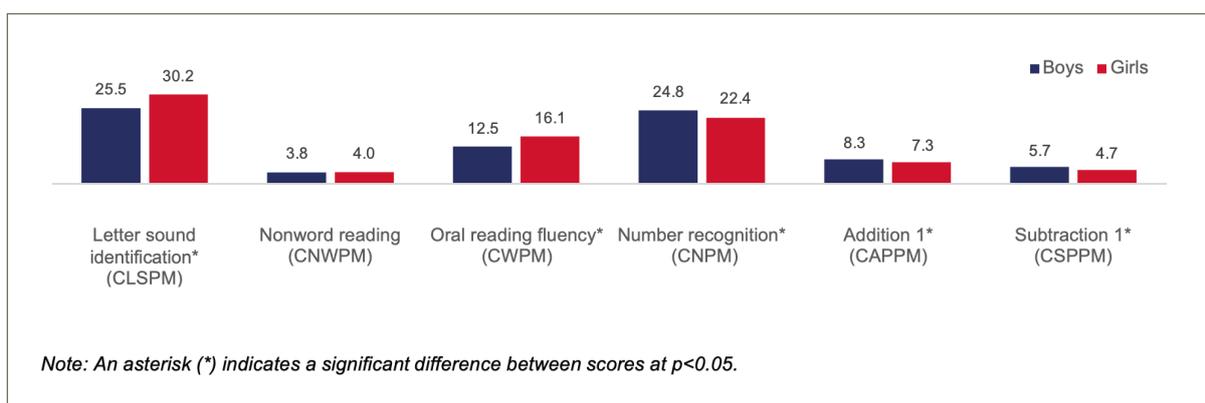


Figure 27. Mathematics Percentage of Items Correct (Accuracy Scores) in QE Schools by Gender



Fluency rates reflect the differences between boys and girls observed in the accuracy scores above. In reading, girls had higher fluency rates than boys in letter sound identification and oral reading fluency subtasks. In mathematics, boys had significantly higher fluency rates than girls on the three timed subtasks – number recognition, addition and subtraction (Figure 28).

Figure 28. Reading and Mathematics Items Correct Per Minute (Fluency Rates) QE Schools by Subtask and Gender



READING AND MATHEMATICS OUTCOMES IN AO SCHOOLS BY GENDER

By gender within AO schools, there is no difference in the proportion of boys and girls by reading proficiency level. Similarly, no differences were observed in reading accuracy scores by gender in AO schools (Figure 30). In mathematics, boys had higher accuracy scores than girls on all subtasks except word problems (Figure 31). These trends are also reflected in fluency rates, where boys had higher fluency rates on the timed mathematics subtasks and no differences were observed between boys and girls on the timed reading subtasks (Figure 32).

Figure 29. Percentage of Students in AO Schools by Reading Proficiency Level and Gender

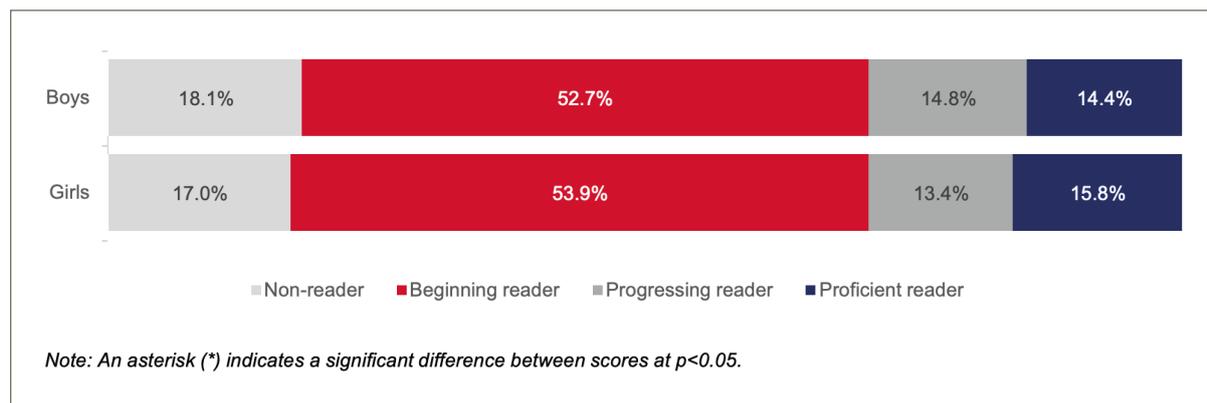


Figure 30. Reading Percentage of Items Correct (Accuracy Scores) in AO Schools by Gender

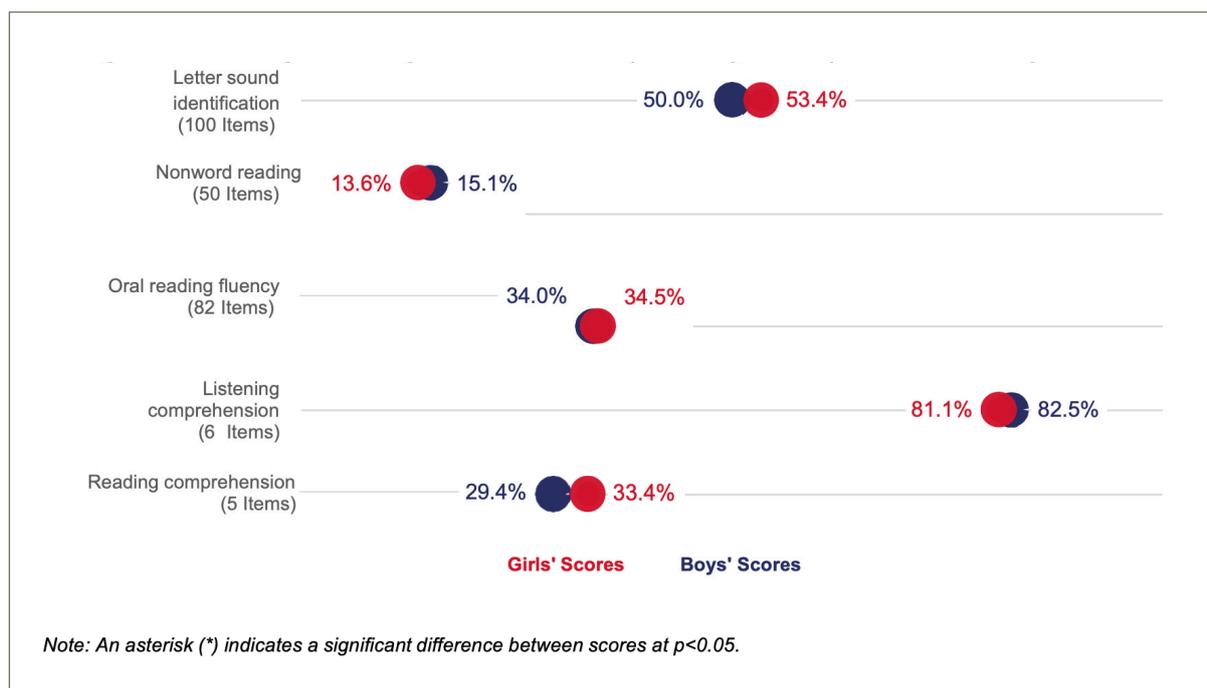


Figure 31. Mathematics Percentage of Items Correct (Accuracy Scores) in AO Schools by Gender

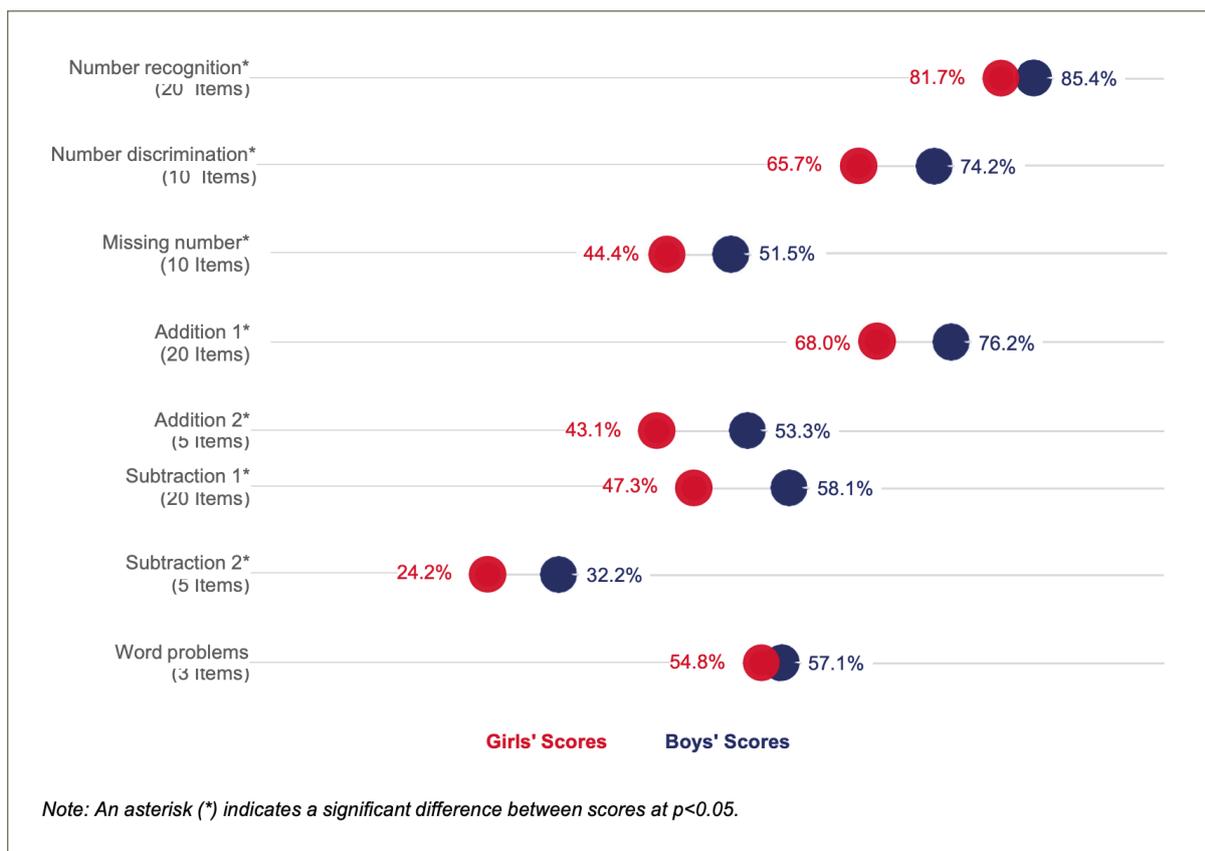
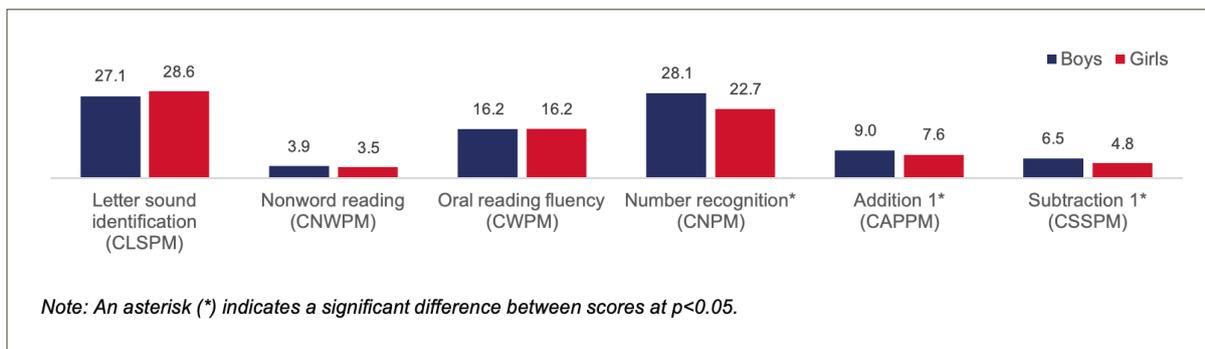


Figure 32. Reading and Mathematics Items Correct Per Minute (Fluency Rates) AO Schools by Subtask and Gender



Research Question 4: Outcomes by Province

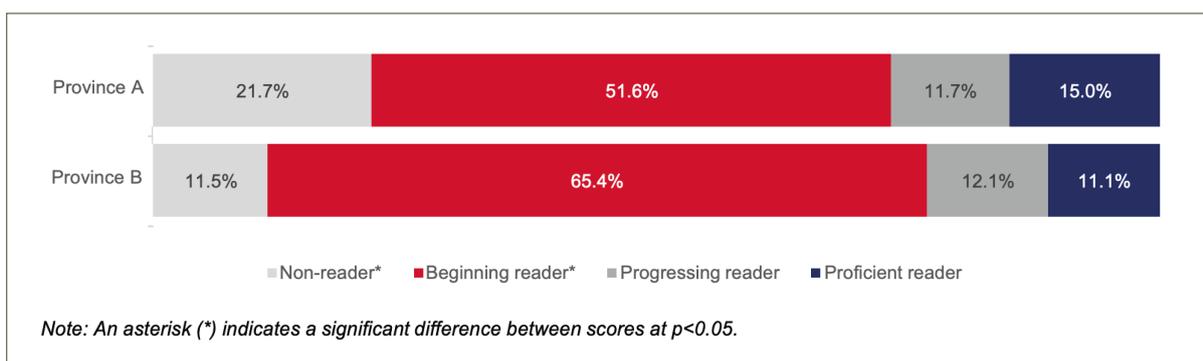
Research Question 4: How do beginning of Grade 3 students' learning outcomes in reading and mathematics compare between Province B and Province A?

This question examining differences in province was included to better understand if there are sizable differences between the two main provinces supported by the Manahel program. The program hypothesised that such differences would indicate equity issues and possibly justify differential levels of support.

READING OUTCOMES BY PROVINCE

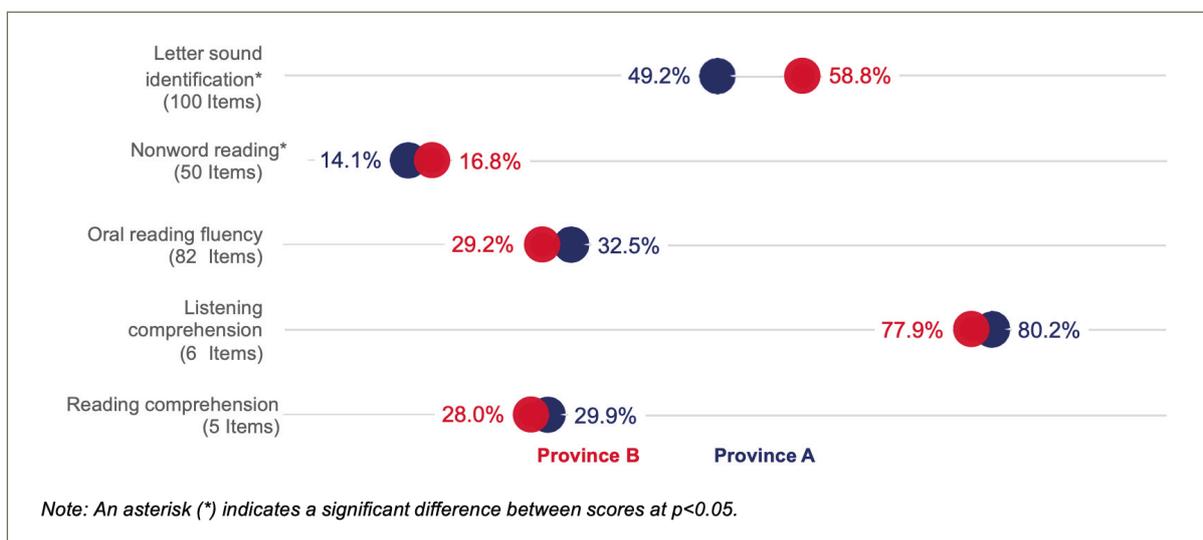
In reading, students' distribution differed significantly by province at the non-reader and beginning reader levels (Figure 33). A higher percentage of students from Province A (21.7%) were non-readers than in Province B (11.5%). Conversely, a higher percentage of Province B students (65.3%) were beginning readers than Province A (51.7%). Proportions were comparable at the progressive and proficient level.

Figure 33. Percentage of Students in Province A and Province B by Reading Proficiency Level



At the subtask level, students in Province B performed significantly better than students in Province A on two reading subtasks – letter sound identification and nonword reading (Figure 34).

Figure 34. Reading Percentage of Items Correct (Accuracy Scores) by Subtask and Province

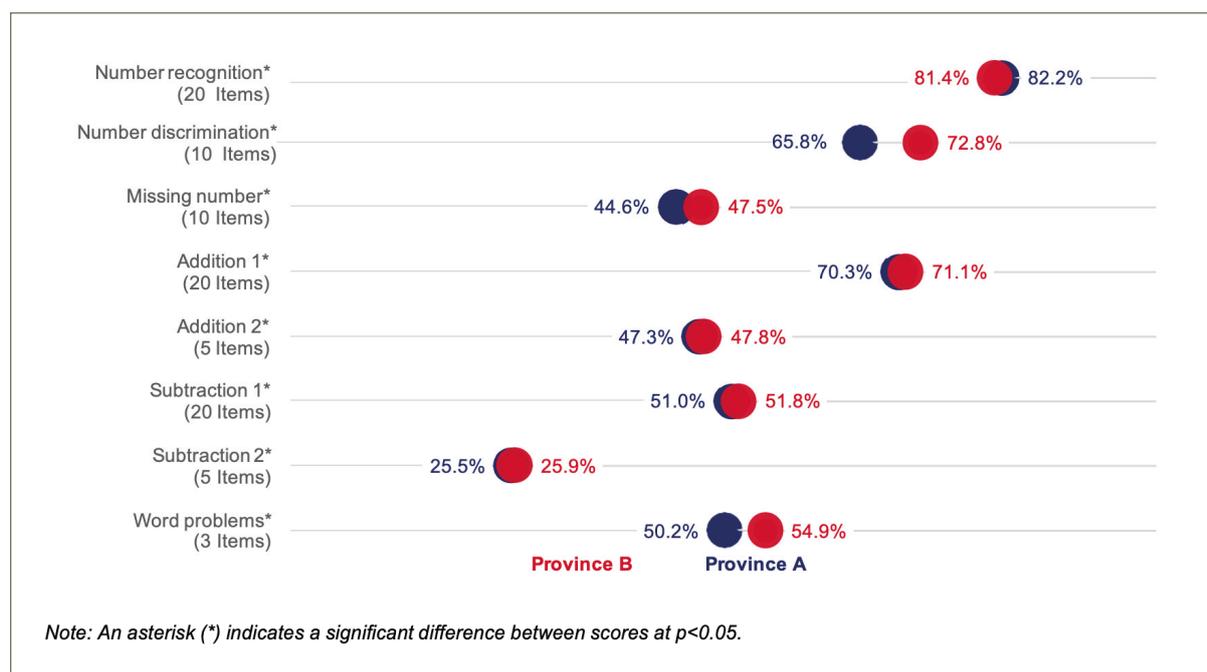


A comparison of zero scores supports the trend that students in Province B performed better in reading than their Province A peers. On average, 13.6% of students in Province A did not identify a single letter sound correctly, significantly higher than the 7.7% in Province B. Similarly, 52.1% of students in Province A did not read a single nonword compared to 37.8% of Province B students; 21.7% of Province A students did not read a single word in the ORF subtask, compared to 11.4% of Province B students (Annex F).

MATHEMATICS OUTCOMES BY PROVINCE

Similar to EGRA accuracy scores, students in Province B performed significantly better than students in Province A on three mathematics subtasks – number discrimination, missing number and word problems (Figure 35).

Figure 35. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Province



The proportion of students with zero scores was significantly different between the provinces on three EGMA subtasks – addition level 1, subtraction level 2 and word problems – with more students in Province A receiving zero scores than in Province B (Annex F).

READING AND MATHEMATICS OUTCOMES IN PROVINCE A BY GENDER

By gender within Province A schools, there is a higher proportion of girls who are proficient readers than boys (Figure 36). Looking at reading accuracy scores, girls have higher scores than boys on two subtasks – letter sound identification and reading comprehension – while boys have higher scores than girls on listening comprehension (Figure 37). No differences by gender were observed in nonword reading and oral reading fluency accuracy scores. In mathematics, boys had higher accuracy scores than girls on all subtasks (Figure 38).

These trends are also reflected in fluency rates, although results for oral reading fluency are different (Figure 39). Using accuracy scores, or number of items answered correctly, there is no difference between boys and girls, but using fluency scores, or number of items answered correctly within one minute, girls have a significantly higher fluency score than boys in Province A.

This suggests that girls may not be answering more items correctly but are reading at a faster pace than boys. The fluency rates for mathematics reflect the same trends observed with the accuracy scores.

Figure 36. Percentage of Province A Students by Reading Proficiency Level and Gender

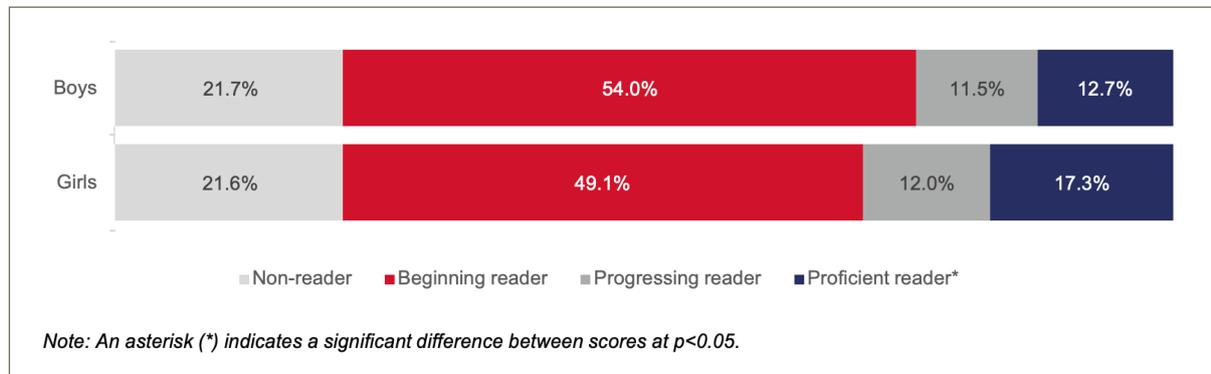


Figure 37. Reading Percentage of Items Correct (Accuracy Scores) in Province A by Reading Proficiency Level and Gender

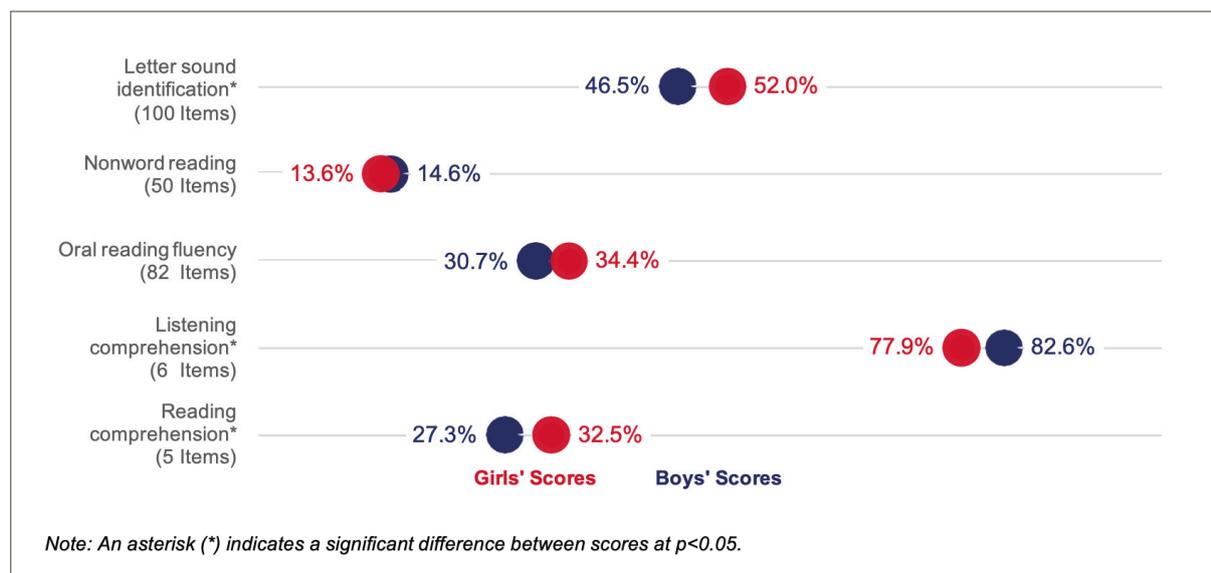


Figure 38. Mathematics Percentage of Items Correct (Accuracy Scores) in Province A by Gender

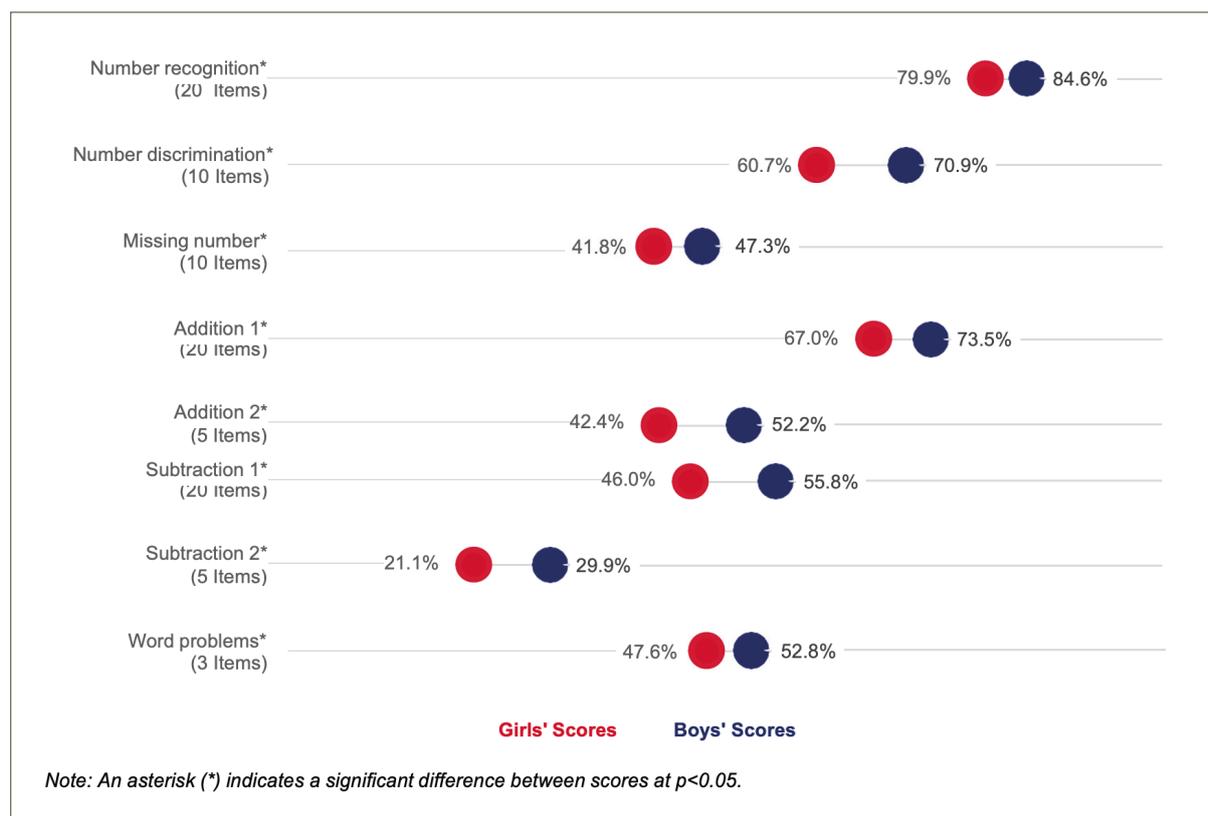
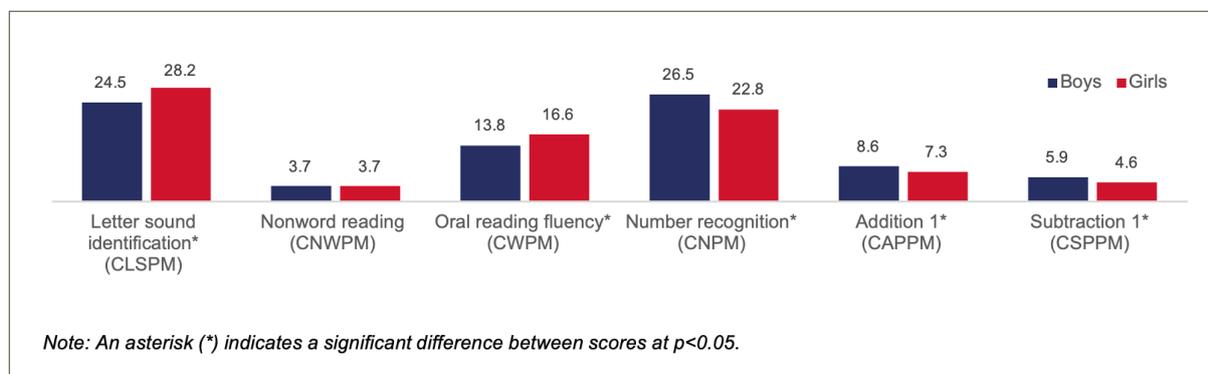


Figure 39. Reading and Mathematics Items Correct Per Minute (Fluency Rates) for Province A by Subtask and Gender



READING AND MATHEMATICS OUTCOMES IN PROVINCE B BY GENDER

By gender within Province B schools, there is a higher proportion of girls who are beginning readers than boys (Figure 40). Looking at reading accuracy scores, however, girls have higher scores than boys on one subtask – letter sound identification (Figure 41). No differences by gender were observed in boys' and girls' accuracy scores for nonword reading, oral reading fluency, listening comprehension or reading comprehension. In mathematics, boys had higher accuracy scores than girls on all subtasks except addition and word problems (Figure 42).

These trends are also reflected in fluency rates, where girls have higher letter sound identification fluency rates than boys, and boys have higher number recognition and subtraction fluency rates (Figure 43).

Figure 40. Percentage of Province B Students by Reading Proficiency Level and Gender

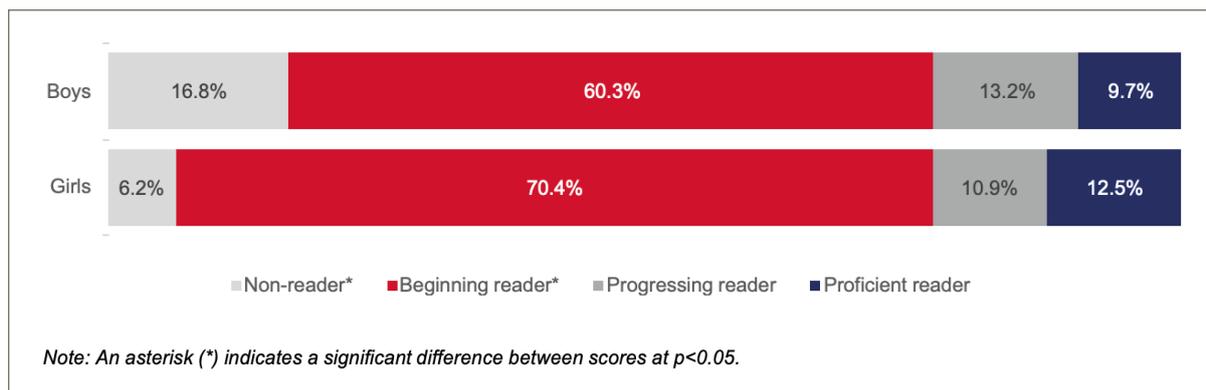


Figure 41. Reading Percentage of Items Correct (Accuracy Scores) in Province B by Gender

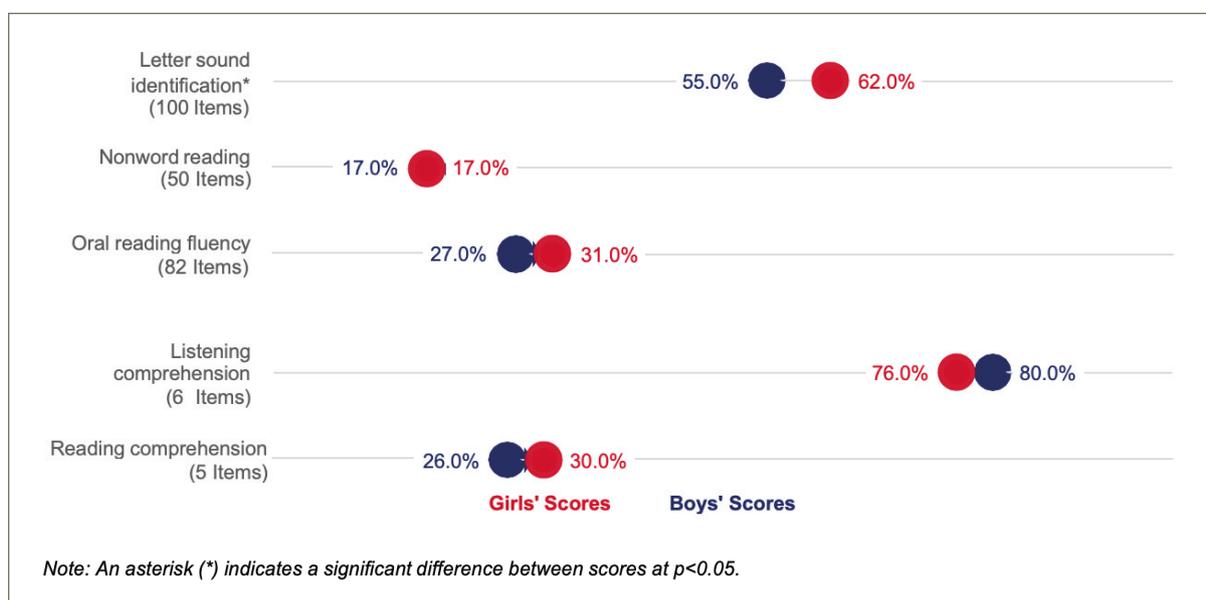


Figure 42. Mathematics Percentage of Items Correct (Accuracy Scores) in Province B by Gender

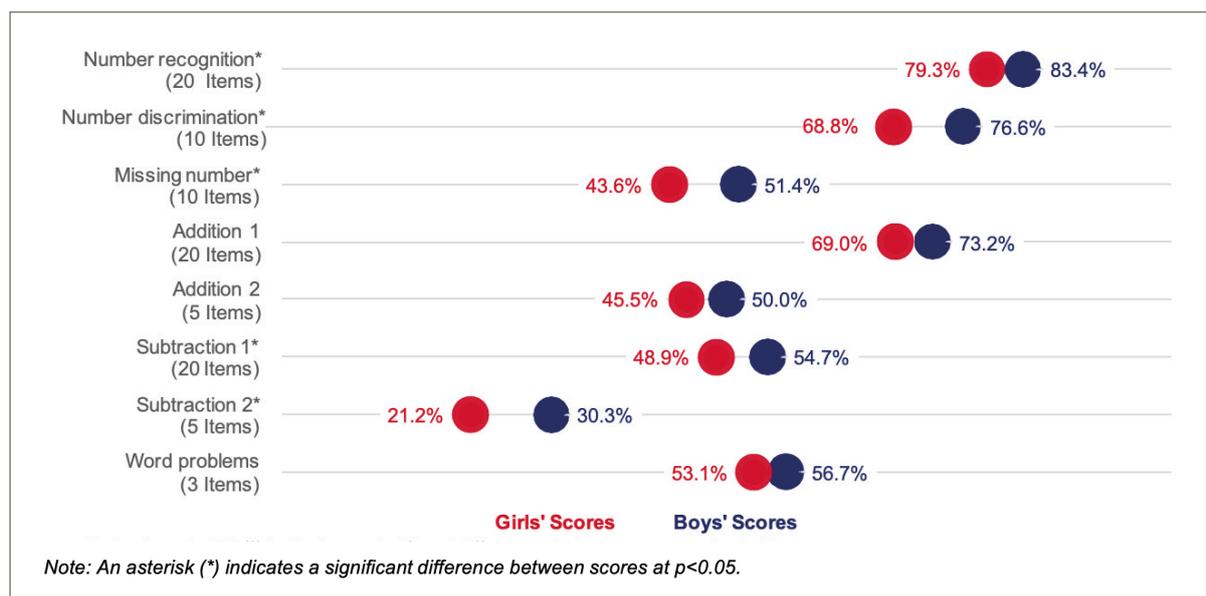
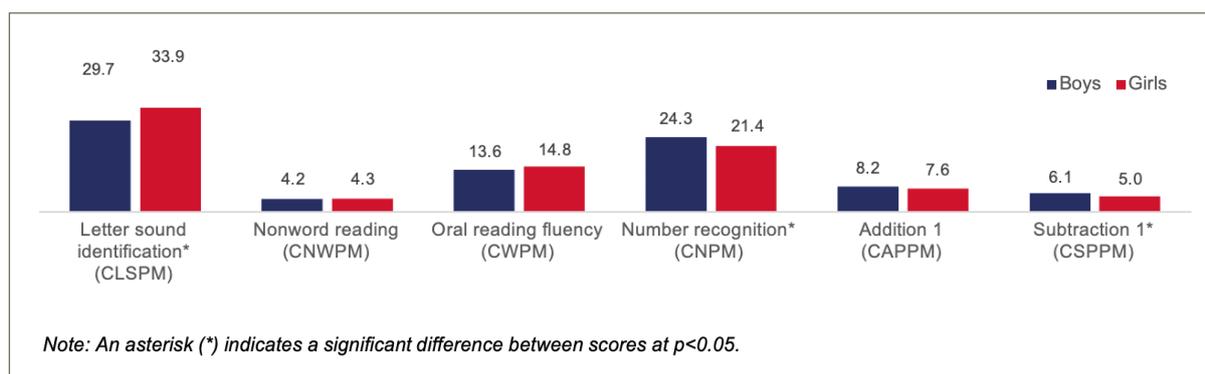


Figure 43. Reading and Mathematics Items Correct Per Minute (Fluency Rates) for Province B by Subtask and Gender

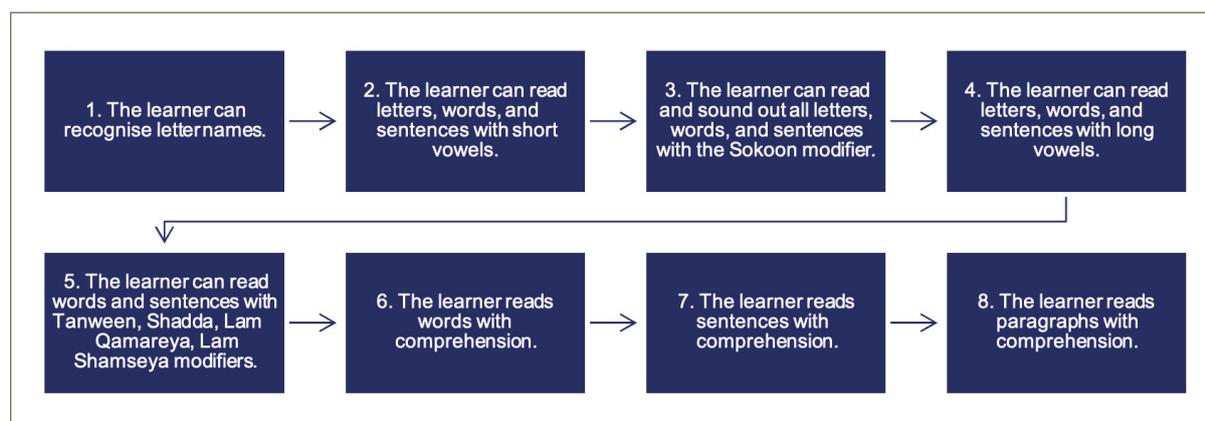


Research Question 5: Linking EGRA Performance with Manahel Reading Levels

Research Question 5: How do beginning of Grade 3 students' EGRA findings relate with the eight reading levels that Manahel-supported teachers are using to track learners' progress? What is the relationship between Manahel-developed reading levels and EGRA performance?

Teachers at QE schools rate the reading performance of their students on an 8-level scale internally developed for use within the Manahel project (Figure 44). Manahel uses these eight reading levels to track progress as movement through the levels reflects advancement towards Arabic reading proficiency.

Figure 44. Manahel Reading Levels



Although the reading levels were not designed to be directly compatible with EGRA, limited comparisons help contextualise EGRA results within Manahel's reading levels. The original intention of this research question was to examine the rating assigned by a teacher to a student against the students' EGRA scores. However, the student-level Manahel ratings were not provided to STS for analysis. Instead, to answer the research question, STS conducted a general alignment of Manahel levels and EGRA subtasks.

Using the description of the Manahel levels and the skills assessed in each task, STS mapped each Manahel level to a task, with letter sound identification as the simplest task until reading comprehension as the most complex reading task.

Where subtasks and Manahel levels align — such as letter names and letter sound identification — direct relationships are ascribed. Where subtasks and Manahel levels do not align — such as levels 2, 3 and 4 — the closest set of subtasks are presented. Where the Manahel level does not align with any EGRA subtask — such as levels 5 and 6 — results are not reported. Table 7 presents the subtasks used to report the proportion of students who have mastered the Manahel reading levels. Students are considered as ‘mastering’ the level if they completed at least 80% of the associated subtask.

Table 7. Manahel Reading Levels Mapped to EGRA Subtasks and Proportion of Students Mastering

Manahel Level		Maps to EGRA Subtask					% of Students Mastering
		Letter Sound Identification	Nonword Reading	Oral Reading Fluency	Listening Comp.	Reading Comp.	
1	The learner can recognize letter names.	☑					21.7%
2	The learner can read letters, words and sentences with short vowels (Al Madd).	☑	☑				0.8%
3	The learner can read and sound out all letters, words and sentences with the Sokoon modifier.	☑	☑	☑			0.7%
4	The learner can read letters, words, and sentences with long vowels (Al Madd).	☑	☑	☑			
5	The learner can read words and sentences with ‘Tanween, Shadda, Lam Qamareya, Lam Shamseya’	Could not be mapped to EGRA subtasks					n/a
6	The learner reads words with comprehension.	Could not be mapped to EGRA subtasks					n/a
7	The learner reads sentences with comprehension.			☑			5.5%
8	The learner reads paragraphs with comprehension.			☑		☑	

MANAHEL LEVEL 1

Learner can recognize letter names (80% accuracy on letter sound identification)

On average, 21.7% of students mastered Manahel level 1, as measured by their performance on the letter sound identification subtask. By province, students mastering level 1 were significantly more likely to be from Province B than from Province A. Specifically, 26.8% of students in Province B mastered Level 2 versus 19.7% in Province A. By district, the highest proportion of students who mastered Level 1 was in District 2 (32.6%) and the lowest was District 5 (4.3%); these differences are statistically significant. There was no significant difference in the proportions by school type — AO versus QE — or within QE schools — fixed versus mobile libraries. Students who mastered Manahel level 1 were likely to be proficient readers (60.1%) or progressing readers (48.6%), to have moved fewer times and to have attended more days of school (Annex H).

MANAHEL LEVEL 2

Learner can read letters, words and sentences with short vowels (Al Madd) (80% accuracy on letter sound identification and on nonword reading)

The proportion of students mastered Manahel reading level 2 was 0.8% (n=12).³⁷ The low proportion of students mastering this level was driven by the nonword subtask. Of the 12 students who mastered Level 2, nine were in Province A (0.8%) and 3 were in Province B (0.6%). Within Province A, four students came from District 1 and five from District 4. Five came from AO schools and seven came from QE schools; all seven students were in QE schools with fixed libraries. Students who mastered Manahel Level 2 were either proficient (n=10) or progressing readers (n=2) and had moved fewer times due to the war.

MANAHEL LEVELS 3 AND 4

Learner can read and sound out all letters, words and sentences with the Sokoon modifier; learner can read letters, words and sentences with long vowels (Al Madd) (80% accuracy on letter sound identification, nonword reading and ORF)

A small proportion of students mastered Manahel reading levels 3 and 4 — 0.7% (n=10). As with Manahel level 2, the constraining subtask was nonword reading. Of the 10 students who mastered levels 3 and 4, eight were in Province A and 2 in Province B. Within Province A, four were in District 1 and four in District 4. By school group, four attended AO schools and six attended QEs with fixed libraries. The majority (9 out of 10) had moved fewer times due to the war and had attended more days of school; nine were proficient readers and one was a progressing reader.

MANAHEL LEVELS 7 AND 8

Learner reads sentences with comprehension; Learner reads paragraphs with comprehension (80% accuracy on ORF and reading comprehension)

Overall, 5.5% of students mastered Manahel reading levels 7 and 8. Proportions ranged from 6.0% in Province A to 4.4% in Province B. There were significantly more students who mastered Manahel levels 7 and 8 in District 4 than in District 3; otherwise, performance was comparable across districts. At AO schools, 6.5% of students mastered this level, while only 5.1% of students attending QE schools did so.

³⁷ Due to the small n sizes, n values are reported instead of percentages.

Within QE schools, significantly more students attended mobile-library schools (8.0%) than in fixed-library schools (4.5%). All students who mastered Manahel levels 7 and 8 were proficient readers. No other correlations were found.

Overall, the comparison of Manahel levels and EGRA performance is not clear, particularly because student-level performance data on Manahel levels was not available. Instead, using a conceptual mapping of Manahel levels to EGRA skills provides a general mapping of the relationship one would expect but does not provide an actual estimate of the relationship. The analysis suggests that Manahel levels and EGRA assess different skills and may not be used as proxy measures of each other. At endline, student-level data for Manahel levels and EGRA scores should be collected so that a correlation can be computed between these scores.

Research Question 6: Student Stress and the Conflict

Research Question 6: What do we know about children's current levels of stress and the relationship with their learning?

To address this research question, STS examined the relationship between students' EGRA and EGMA scores with various stressors reported by students in the Student Stressor Survey — frequency of moves due to war, as well as feelings of tiredness and hunger at school, which serve as proxies for students' stress.

Two-thirds of students reported they were forced to move due to the war — more students in Province B than in Province A. Of those who moved, the majority report having moved between one and three times. Half of all students reported experiencing hunger at school and more than one-third reported tiredness; these rates were higher among students in Province B than in Province A.

Two confounding variables make the analysis by displacement challenging. First, moving homes and schools due to war appears to be confounded. Second, learning outcomes for those who were displaced and those who were not vary by district. Taken together, the impact of displacement is necessarily intertwined with the impact on communities overall. It also indicates that students who did not experience displacement are not necessarily more likely to succeed. As a result, analyses that account for all the variables of interest with learning outcomes — forced to move home, school change, district, school type, student gender — were not possible due to the lack of cases in the dataset at the intersection of all variables. Instead, correlations between pairs of variables — and their relationships with learning outcomes — are used to construct the story in the following sections.

DISPLACEMENT

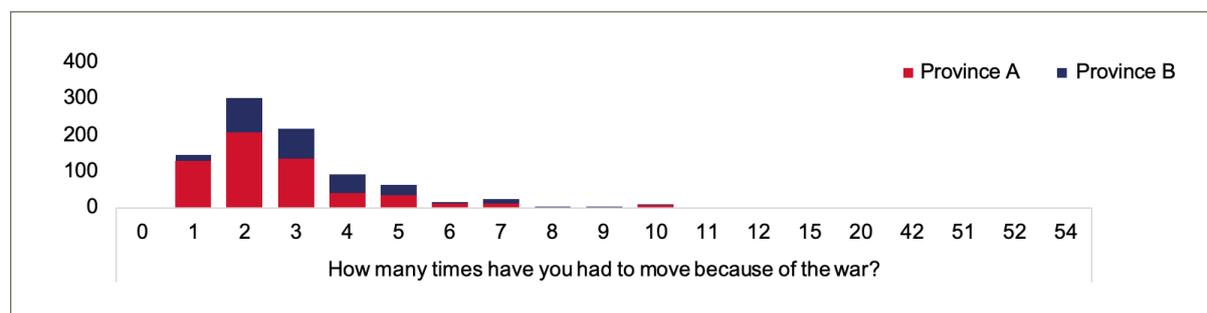
Overall, 63.4% of students reported moving because of the war. The number of reported moves ranged from zero to 54 (Figure 45).³⁸ The sample averaged 3.9 moves, but the majority of students reported moving two times.

³⁸ The highest frequency was 54 times reported by one student and is reflected in the figure.

Analysis of learning outcomes by displacement rate show that, in general, students who have moved fewer than five times as a result of the war have higher reading and mathematics outcomes compared to students who moved six times or more. For example, 28.0% of students who moved between one and five times were progressing or proficient readers, compared to 22.7% of students who moved six or more times.

Furthermore, when analysts examine the results for students who did not report being displaced, the findings underline the need for additional data. Students who reported that they were not forced to move to a different home due to the war had statistically significantly lower scores on reading and mathematics subtasks than students who reported that they were forced to move to a different home. The difference was notable in four districts – District 1, District 3, District 4 and District 6 – and on four to nine different reading and mathematics subtasks. In other words, the data suggest that students who have been forced to move to a different home due to the war have higher scores than students who have not been forced to move. For example, in reading comprehension, students in District 1, District 3 and District 6 who moved had higher scores than students who did not move. This suggests that while displacement can have an impact on learning outcomes, changes in the learning environment and community can also impact learning outcomes. Data on the learning environment was not available in the current dataset and should be considered at endline.

Figure 45. Number of Students by Times They Moved Due to the War



The number of times a student changed schools was also collected. Overall, 40.1% of students reported attending a new school as a result of moving due to the war. On average, students attended 2.3 new schools as a result of the war. In addition, 22.3% of students reported that there were times when they did not have a school to go to. Students were not asked the length of time they were out of school during this study, but this may be a helpful question to add at endline.

As with displacement, high rates of changes in schools also has a negative relationship with learning outcomes (Figure 46 and Figure 47). Students who were classified as non-readers and beginning readers were more likely to have moved six times or more due to the war, with non-readers more likely to have attended four or more new schools. By contrast, students who moved between one and five times and had attended less than four schools were more likely to be classified as progressing or proficient readers. For results by province and district, see Annex D.

Figure 46. Percentage of Students by Reading Level, 1-5 Moves

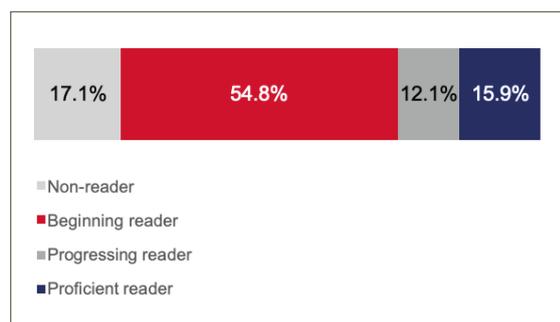
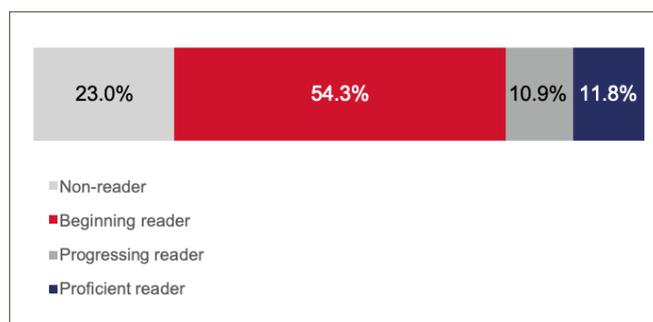
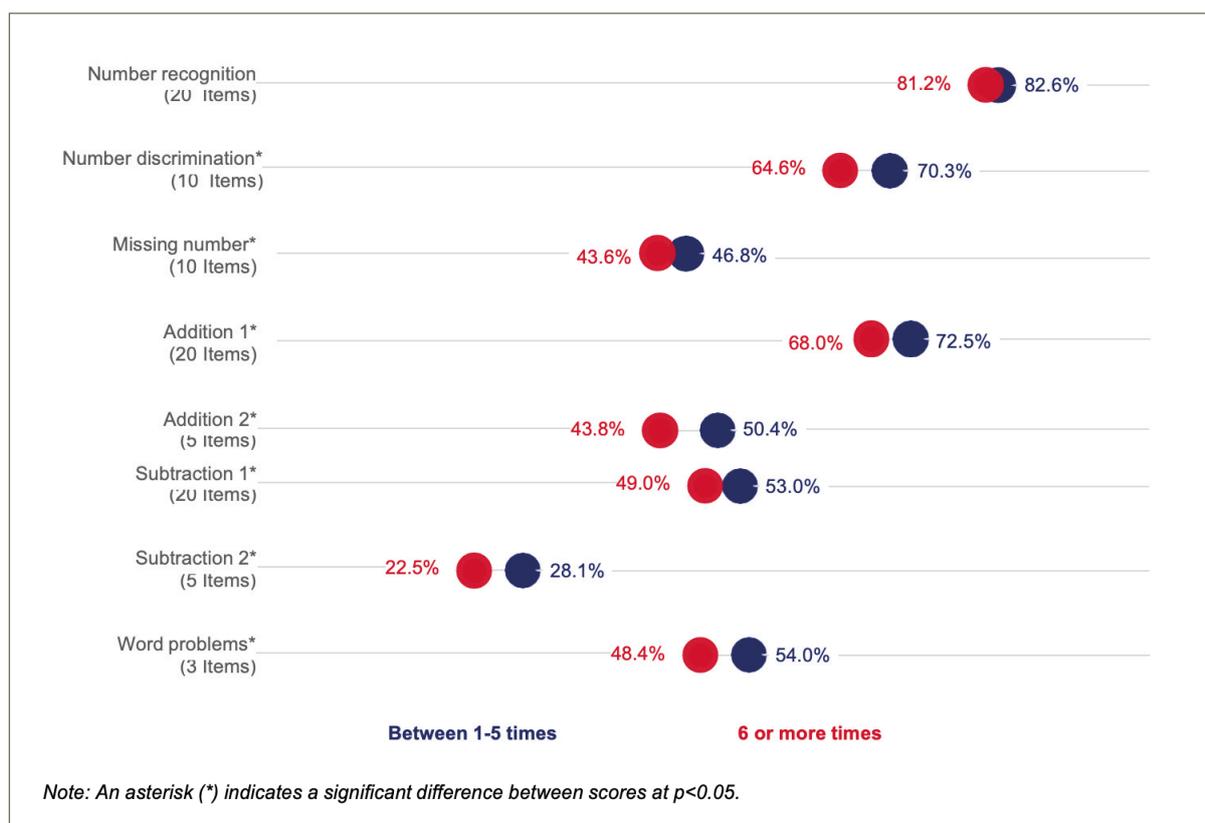


Figure 47. Percentage of Students by Reading Level, 6+ Moves



Similarly, the number of schools attended also correlated with EGRA performance. Fewer moves and fewer schools were associated with accuracy scores on all reading and mathematics subtasks. Students who attended between one and three schools had significantly higher accuracy scores on all EGRA subtasks — except letter sound identification — than students who attended four or more new schools. On all EGMA subtasks except number recognition, students who had moved between one and five times as a result of the war had significantly higher accuracy scores than students who moved more than six times (Figure 48). See Annex E for additional details.

Figure 48. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Displacement Rate



Students who attended fewer new schools outperformed their peers who had attended more new schools on two EGMA subtasks — missing number and subtraction level 2. The prior analyses examine the impact of displacement and school changes on student learning, separately. However, moving homes and schools due to war is related.

Students who reported that they did not move to a different school but had moved homes due to the war had comparable scores to students who moved both homes and schools. Among students who did not move homes, those who attended a different school had higher scores than those who attended the same school. This finding suggests that, as with the interconnectedness of district and displacement, changes in schools with or without a change in homes, is also related.

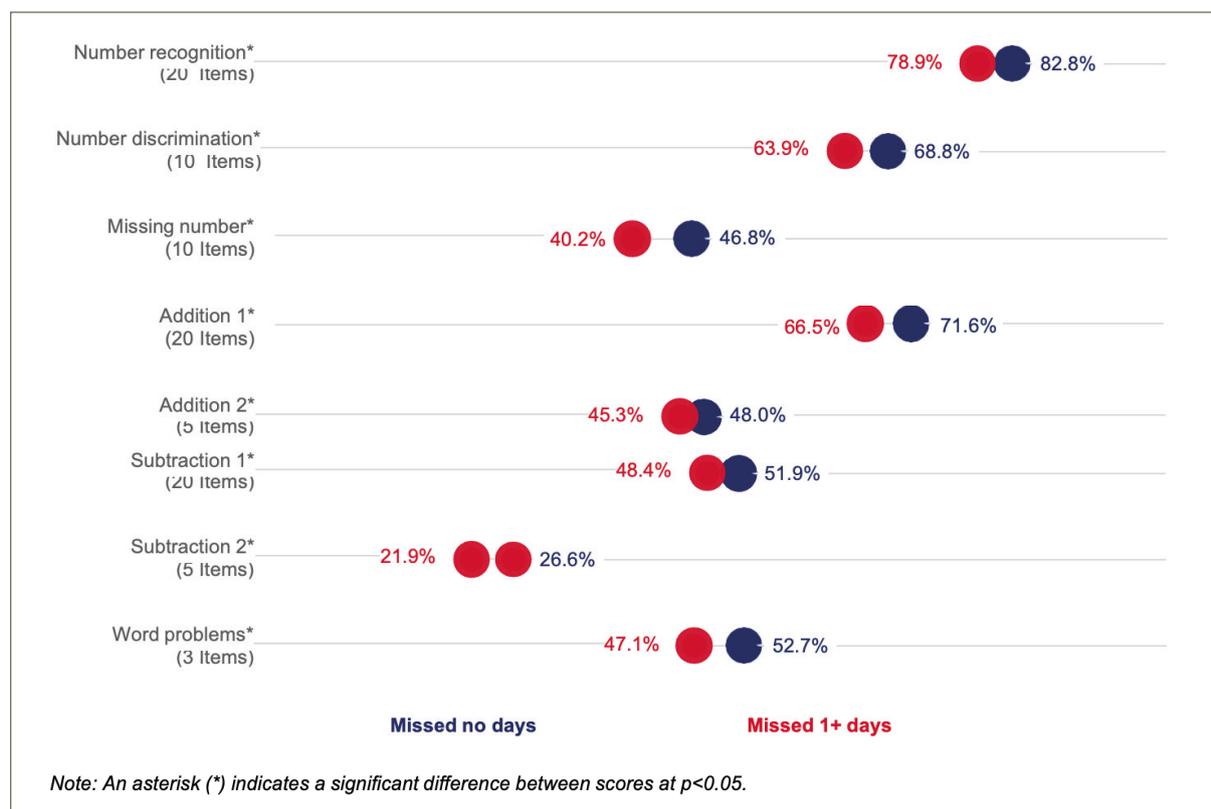
Generally, determining the impact of displacement on learning, and specifically, the frequency of displacement, is not possible given the mixed results by district and confounding of moving home and school. Additionally, what may be more appropriate is a measure of the learning environment, inclusive of the frequency of home and school moves that the student experiences. What is clear is that high displacement frequency, defined as six or more moves due to the war, appears to have a negative effect on reading and mathematics outcomes. To further disaggregate and understand the impact of displacement on student learning and the learning environment in general, additional data on the time between displacements, distance moved, extent of move's impact on family, change in schooling experience, peers and staff, impact on community, for example, should be collected.

SCHOOL ATTENDANCE, TIREDNESS AND HUNGER

Attendance

Students with poorer attendance had poorer learning outcomes, in general. Students who had not missed any school in the past five days were more likely to be progressing or proficient readers than students who had missed school. Collectively, 28.8% of students who had not missed school were progressing or proficient readers, compared to 13.8% of students who had missed at least one day of school. The converse is true for the lower proficiency levels – 86.2% of students who missed school were non-readers or beginning readers, compared to 71.1% of students who had not missed school. See Annex D for results by province. On every EGMA subtask except addition level 2, students who had not missed any days of school outperformed their peers who had missed at least one day of school (Figure 49).

Figure 49. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask and Attendance Rate



Tiredness

Among students surveyed, 63.5% reported that they are never tired at school while 36.5% report that they do experience tiredness at school. Tiredness at school may serve as a proxy for level of general stress. When examined against displacement, there was a very weak but statistically significant correlation of 0.067 between feeling tired and having been forced to move homes due to the war. However, there was no relationship with the frequency of moves and tiredness.

Overall, students who reported tiredness were more likely to be non-readers — 23.3% compared to 16.9% who did not experience tiredness at school. However, students who were never tired at school were more likely to be beginning readers — 58.1% compared to 49.6% of students who did experience tiredness. Tiredness at school did not have a relationship with being a proficient or progressing reader. This trend was seen in each of the provinces as well.

Students who never experience tiredness at school consistently had a higher percentage of items correct in all mathematics subtasks than students who did experience tiredness. On number recognition, students who never experienced tiredness averaged 82.8% correctness, which is statistically significantly more than the 80.7% correct among those who did experience tiredness. Similarly, on addition level 1, students who never experienced tiredness significantly outperformed their peers, averaging 49.3% compared to 45.3% correct.

Hunger

Overall, 49.8% of all students reported experiencing hunger at school at some point. Analysis showed no significant differences in the distributions of reading proficiency levels by hunger, nor any in zero scores or fluency rates. The only significant difference was observed on listening comprehension's accuracy score; students who reported never experiencing hunger at school had a significantly higher average accuracy score (81.7%) compared to those who reported experiencing hunger (78.7%). At the province level, students in Province A who said they experience hunger at school were more likely to be non-readers, while students who were never hungry at school were more likely to be beginning or progressing readers (Annex D).

Hunger did correlate with EGMA performance. Students who never experience hunger at school have higher fluency rates on all subtasks than students who experience hunger. For instance, average number recognition fluency score was 25.1 CNPM for students who are not hungry, compared to 23.2 CNPM among students who do experience hunger. Students who never experience hunger at school had an average addition fluency score of 8.3 CAPP, compared to 7.6 CAPP among students who do experience hunger. Students who never experience hunger averaged a subtraction fluency score of 5.7 CSPM, compared to 5.1 CSPM among students who do experience hunger.



Results and Recommendations by Skills

This section presents students' results by skills in reading and mathematics. EGRA subtasks are organised in three skills categories using the Cognitive Framework for Reading: mechanics of reading, understanding and reading comprehension.³⁹ Similarly, EGMA subtasks are organised in three skill categories using the Global Proficiency Framework: whole numbers, operations and real-world problems.⁴⁰

This section explores the role of each skill in the context of other skills and uses demographic information to describe students with low-, moderate- and high-performance levels. This may allow programmatic interventions to target specific students. Findings for each reading and mathematics skills category describe:

1. the relationship between the skill and other skills to identify the role a specific skill plays in improving students reading and mathematics performance.
2. the gap between proficient readers' and beginning readers' performance on each subtask because the majority of students are beginning readers.⁴¹
3. the demographics of low-, moderate- and high-performers on each subtask.

Unlike the preceding section, which only presents results, this section incorporates findings and recommendations.

³⁹ Wren, Sebastian, 'The Cognitive Foundations of Learning to Read: A Framework,' Southwest Educational Development Laboratory, 2000.

⁴⁰ 'Global Proficiency Framework for Reading and Mathematics,' USAID, 2019

⁴¹ Beginning readers are students who read between one and 22 CWPM but scored less than 80 percent on the comprehension subtask. Proficient readers are students who scored 80% or more on the reading comprehension subtask.

Reading Skills

MECHANICS OF READING

'Mechanics of reading' includes skills that build students' ability to engage with written words and sentences and — when combined with the skill of meaning-making — leads to reading comprehension. Three EGRA subtasks measure the mechanics of reading: letter sound identification, nonword reading and ORF.

The relationship between mechanics of reading and other reading and mathematics skills helps to understand the relative importance of this skill. Students with strong skills in reading mechanics also have strong performance in reading comprehension. However, reading mechanics is weakly related to understanding, which suggests that reading comprehension and understanding skills develop somewhat independently of each other among Manahel learners. Reading mechanics is weakly-moderately correlated with all mathematics skills — whole numbers, operations and real-world problems. This suggests that students with strong reading skills can have weak or strong mathematics skills.

To understand students' ability in mechanics of reading, this section presents accuracy scores and fluency rates. The narrative that follows focuses on comparing beginning readers — which most students are — with proficient readers — which the project hopes all students will become — to examine ways that the gap between these students can be bridged by Manahel interventions. Results for letter sound identification and ORF subtasks are reported together because their trends are similar, and results for nonword reading are separated out.

Figure 50 shows the distribution of letter sound identification accuracy scores by three groups: low scorers (0- 40% correct), moderate scorers (41-80% correct) and high scorers (81-100% correct). Figure 50 also shows the distribution of letter sound identification fluency rates by the same three groups but using fluency rates — low fluency at 0-40 CLSPM, moderate at 41-80 CLSPM and high fluency at 81+ CLSPM. Similar figures are shown for scores on the ORF (Figure 52).

Proficient readers tend to have higher letter sound identification accuracy and fluency. Beginning readers letter sound identification fluency rates range from 0 to 60 CLSPM while proficient readers rated 30 CLSPM and higher (Annex I). Similar trends were observed on the ORF subtask. Beginning readers scored between 0-50% correct while proficient readers tend to have accuracy scores greater than 50% correct (Annex I). Interestingly, proficient students had accuracy scores between 41-80% correct and 81% correct or higher.

On ORF, proficient readers' fluency rates were as low as 22 CWPM to over 90 CWPM, indicating that the pace of reading required for comprehension varies (Figure 52).⁴² Furthermore, while the majority of beginning readers have fluency rates below 10 CWPM, there are beginning readers with fluency rates between 11-20 CWPM. This indicates that these beginning readers are at the cusp of transitioning from the beginning reader category into the progressing reader category (Annex D).⁴³

⁴² WBeginning readers are students who read between one and 22 CWPM but scored less than 80 percent on the comprehension subtask. Proficient readers are students who scored 80% or more on the reading comprehension subtask.

⁴³ Progressing readers are students who read 23 CWPM or more but scored less than 80 percent on the comprehension subtask

These results together indicate that while improving accuracy and fluency of reading a connected text is important, a focus on instruction for comprehension alongside reading mechanics is warranted.

Figure 50. Distribution of Letter Sound Identification Accuracy Scores and Fluency Rates for Beginning and Proficient Readers

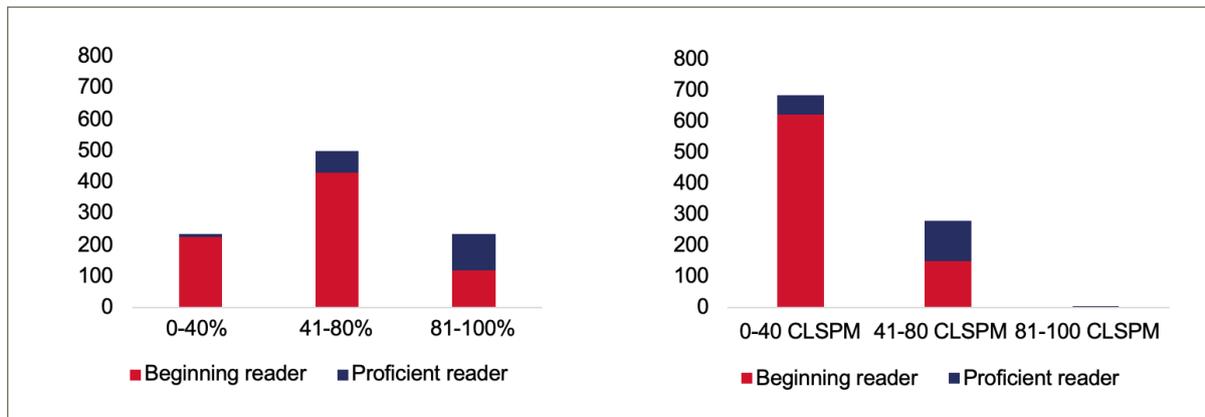


Figure 51. Distribution of Letter Sound Identification Accuracy Scores by Gender

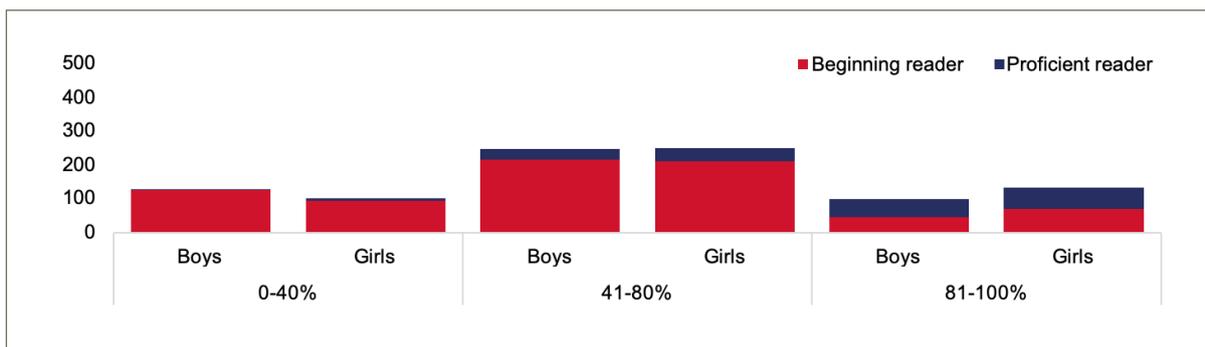


Figure 52. Distribution of Oral Reading Fluency Accuracy Scores and Fluency Rates for Beginning and Proficient Readers

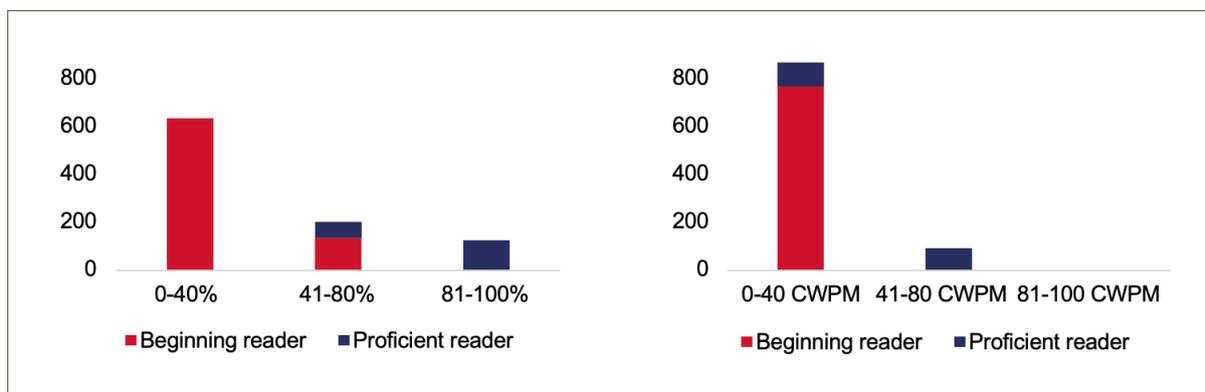
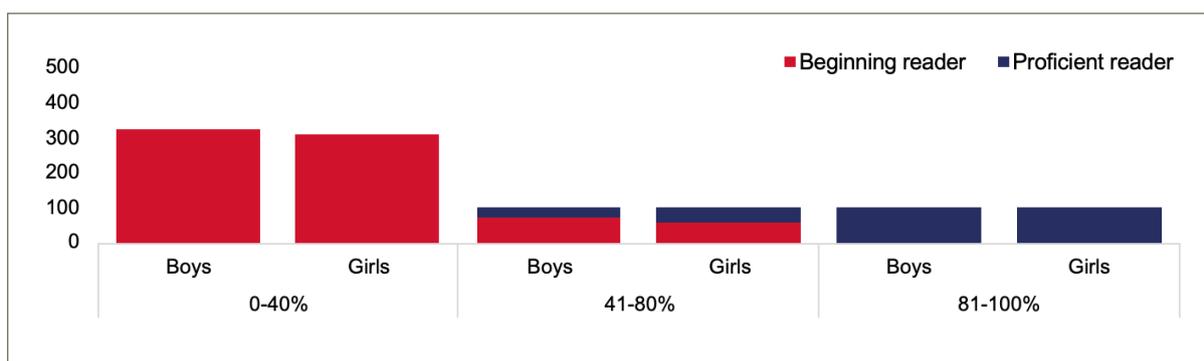


Figure 53. Distribution of Oral Reading Fluency Accuracy Scores by Gender



To better understand students who struggle in letter sound identification and ORF, Table 8 provides demographic descriptions of students with low performance, moderate performance and high performance on the letter sound identification and ORF subtasks.

Table 8. Student Profiles for Mechanics of Reading Skills by Performance Category

	0-40% Items Correct	41-80% Items Correct	81-100% Items Correct
Letter Sound Identification	<ul style="list-style-type: none"> • More likely to be in Province A • Within Province A, more likely to be in District 3 and District 5 • More likely to be QE mobile- library schools • More likely to be Waves B or D • More likely to have missed some school in past five days 	<ul style="list-style-type: none"> • More likely to be from Province B • More likely to be evening shift schools • More likely to be Wave C 	<ul style="list-style-type: none"> • More likely to be from Province B • Within Province A, more likely to be from District 2 • More likely to be in evening shift schools
Oral Reading Fluency	<ul style="list-style-type: none"> • More likely to be QE schools without mobile libraries • More likely to have missed some school in past five days 	<ul style="list-style-type: none"> • Within Province A, more likely from District 4 • More likely to be Wave C More likely to be QE mobile- library schools • More likely to have attended school for the past five days 	<ul style="list-style-type: none"> • Within Province A, more likely from District 4 • More likely to have attended school for the past five days

To bridge the gap between beginning and proficient readers in reading mechanics, instruction needs to focus on improving accuracy and fluency in letter sound identification and ORF. Students in Province A — specifically District 3 and District 5 districts — can benefit from instructional support in letter sound identification, as can students in Wave B and D schools and in mobile-library schools.

Beginning readers at the cusp of ORF need support to cross the threshold into being progressing readers. Within QE schools, students in fixed-library schools are more likely to have low accuracy scores, suggesting that additional supports for reading are necessary.

Because Province B students are more likely to have high accuracy scores, it would be worthwhile to explore how instruction in this district may differ from other districts to see if this is having an influence or if other contextual factors are coming into play. Finally, attendance matters – students who attend more days of school are more likely to have higher scores in reading mechanics. Factors within the project’s scope to address this issue already seem to be in place – specifically, the use of safeguarding officers and psychosocial trainings at the schools. Additional data may be useful in understanding factors that may preclude students from attending school more regularly – including home and family factors.

Though also a reading mechanics skill, nonword reading saw notably different trends in results than letter sound identification and ORF. Scores on nonword reading – both accuracy and fluency – are weakly related to a student’s proficiency in reading. In other words, students who were proficient readers did not necessarily have higher scores on the nonword reading subtask. For that reason, performance in nonwords may not be a critical building block for students’ mastery of the Arabic language. Further examination of the appropriateness of this subtask for Manahel learners is recommended, particularly if Manahel reading materials include this skill. If deemed not appropriate, it is recommended that this subtask be removed from the endline assessment.

UNDERSTANDING

Understanding includes a student’s ability to make meaning from verbal communication, a skill that is independent from reading. In the context of understanding students’ reading abilities, it is beneficial to compare students’ reading skills – mechanics and comprehension – with understanding to identify if they are limited by an inability to understand language. One EGRA subtask measures understanding: listening comprehension.

The understanding skill, measured by listening comprehension, is weakly correlated to mechanics of reading and all mathematics skills; there is a weak-moderate correlation between understanding and reading comprehension. These findings indicate that, although an important skill, the listening comprehension subtask is not necessarily related to reading and mathematics skills development.

For the listening comprehension subtask, students listened to a story read to them out loud and responded to six comprehension questions. Using the number of correct listening comprehension questions, Figure 54 shows the number of beginning and proficient readers grouped into three accuracy categories. Nearly all students in the low performance category are beginning readers (98.6%). Out of the 194 proficient readers, 94.8% scored in the high-performance category.

Despite their nascent reading skills, a majority of beginning readers responded correctly to at least four listening comprehension questions. This supports the conclusion that students who are beginning readers are not necessarily limited by an inability to understand language.

Figure 54. Distribution of Listening Comprehension Accuracy Scores for Beginning and Proficient Readers

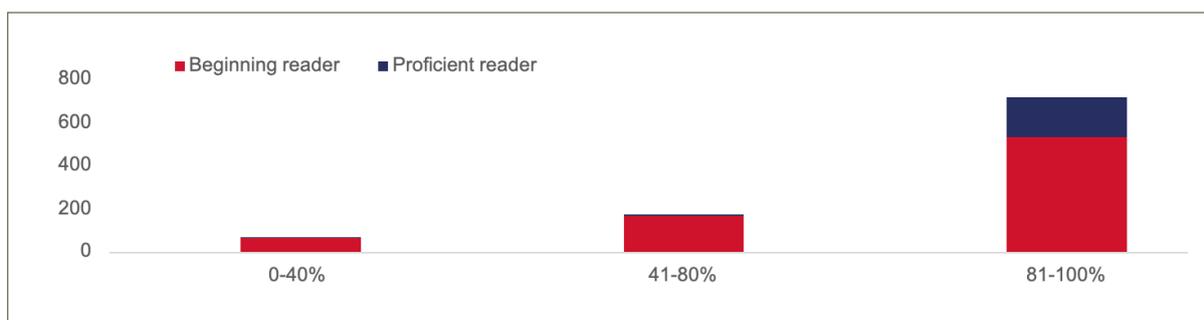


Figure 55. Distribution of Listening Comprehension Accuracy Scores by Gender

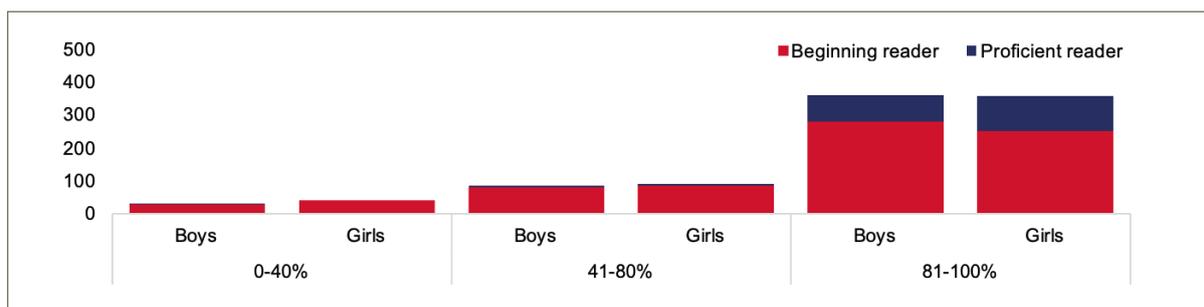


Table 9 provides descriptions of students with low performance, moderate performance and high performance on the listening comprehension subtask.

Table 9. Student Profiles for Understanding Skills by Performance Category

	0-40% Items Correct	41-80% Items Correct	81-100% Items Correct
Listening Comprehension	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 5 • More likely to be full time schools • More likely to have been internally displaced 6 or more times • More likely to have 6 or more people in the household 	<ul style="list-style-type: none"> • More likely to be from Province B • Within Province B, more likely to be from District 6 • More likely to be QE schools • More likely to have missed school in past five days 	<ul style="list-style-type: none"> • More likely to be from Province A • Within Province A, more likely to be from District 2 and District 4 • More likely to be AO schools • More likely to have been internally displaced 1-5 times • More likely to have attended school for the past five days

Because these performance categories for listening comprehension are weakly correlated with reading mechanics, the instructional interventions guided by these findings should focus on verbal communication comprehension.

READING COMPREHENSION

Reading comprehension is the pinnacle skill and one ultimate goal of reading. Comprehension builds on a students' reading mechanics abilities, culminating in students' ability to make meaning out of the text that they read.

One EGRA subtask measures reading comprehension. As expected, reading comprehension has a moderate-strong relationship to reading mechanics; this finding reinforces the importance of building reading mechanics skills in order to achieve reading comprehension. Reading comprehension is weakly correlated to understanding and real word problems. However, reading comprehension is moderately related to whole numbers and operations skills.

For the reading comprehension subtask, students responded to up to five questions — explicitly or inferentially — related to the passage the student read aloud for the ORF subtask. The number of questions asked varied from student to student and depended on how far he or she read in the passage. That is, students were only asked questions about text they read.⁴⁴ Figure 56 shows the number of beginning and proficient readers grouped into three accuracy categories.⁴⁵

Figure 56. Distribution of Reading Comprehension Accuracy Scores for Beginning and Proficient Readers

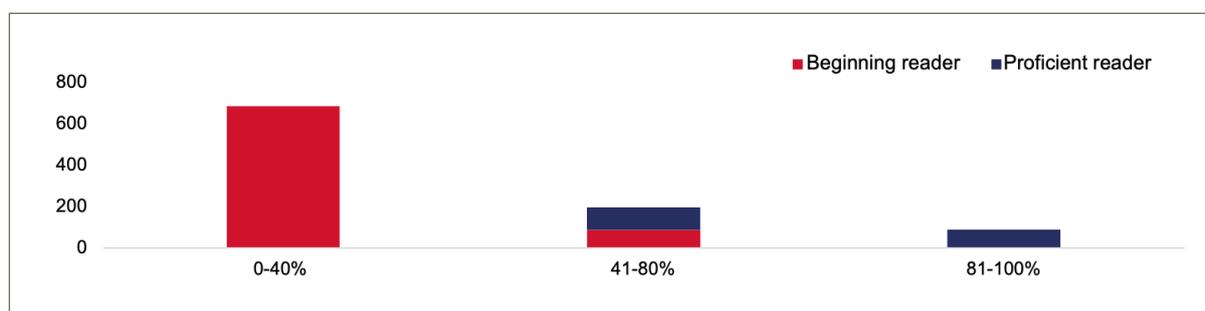
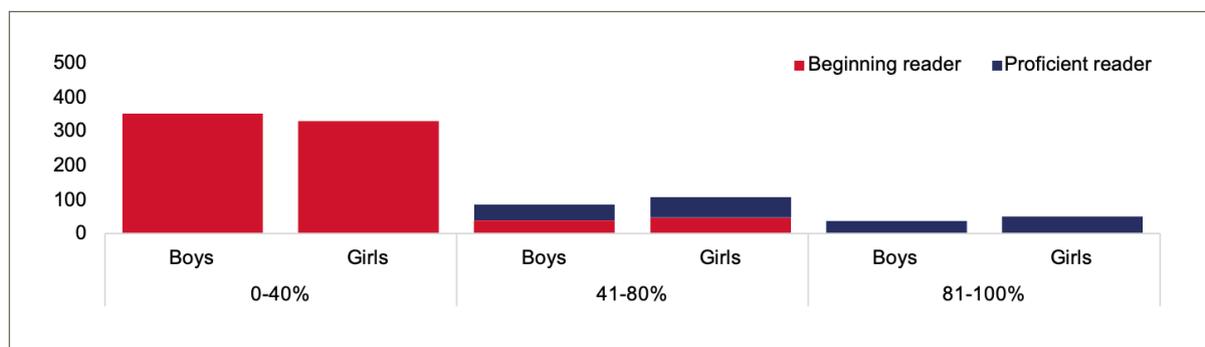


Figure 57. Distribution of Reading Comprehension Accuracy Scores by Gender



Nearly all beginning readers — 9 in 10 — scored in the low performance category.⁴⁶ In contrast, no proficient readers scored lower than 41% accuracy on the reading comprehension subtask; two in five had at least 81% accuracy. Table 10 provides descriptions of students with low performance, moderate performance and high performance on the reading comprehension subtask. Reading comprehension relies on a strong foundation in reading mechanics. Furthermore, students in Province B and in District 3 need additional support to improve their reading comprehension skills.

⁴⁴ More than one-quarter of students did not read far enough to be asked any comprehension questions (28.7%); almost one-quarter read far enough to be asked two questions (23.3%), 19.8% were asked three questions, 11.6% were asked four questions and 10.5% of students read the whole passage and were asked all five questions.

⁴⁵ Student reading performance proficiency levels are based on ORF and reading comprehension results. In other words, the banding of students into beginner and proficient is already based on their reading comprehension accuracy.

⁴⁶ The low performance category includes students who received a zero score on the ORF subtask and were not read a single comprehension question; and students that had the opportunity to respond to at least one comprehension question.

Table 10. Student Profiles for Reading Comprehension Skills by Performance Category

	0–40% Items Correct	41–80% Items Correct	81–100% Items Correct
Reading Comprehension	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 3 • More likely to have missed one or more days of school in past five days 	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 4 • More likely to have attended school for the past five days 	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 4

Mathematics Skills

The next three sections cover skills assessed on the EGMA: whole numbers, operations and real-world problems. The relationship between the three mathematics skills and the three reading skills assessed in the Manahel EGMA and EGMA were weak to moderate. Weaker relationships suggest that better reading skills may be associated with better mathematics skills among Manahel learners, but not always.

Like the reading sections, mathematics skills are reported using beginning and proficient reader categories. Using these reading categories to describe student performance allows consistent analysis of student performance across sections, which allows the reader to focus adjustments to mathematics interventions for subgroups of students alongside adjustments to the reading intervention.

WHOLE NUMBERS

Whole numbers include skills that build students' ability to identify and count whole numbers as well as to identify the relative magnitude of whole numbers.⁴⁷ Three EGMA subtasks measure whole numbers: number recognition, number discrimination and missing number.

Students' performance in whole numbers is moderately related to operations — suggesting that students with stronger skills in whole numbers also have stronger performance in operations. The relationship between reading mechanics and real-world problems was weak, contrary to what one would expect. However, in this study, the relationship is likely weak because students were not required to read the word problems themselves — enumerators read the word problem subtask questions aloud to students.

Figure 58 shows the distribution of number recognition accuracy scores and fluency rates. Figure 60 shows the distribution of number discrimination, and Figure 62 shows missing number accuracy scores, each followed by the respective gender breakdown.⁴⁸

⁴⁷ USAID (2019). Global Proficiency Framework Reading and Mathematics, Grades 2-6. Retrieved from: <https://www.edulinks.org/sites/default/files/media/file/GAML6-REF-16-GLOBAL-PROFICIENCY-FRAMEWORK.pdf> on December 30, 2019. Note that whole numbers also includes the ability to represent whole numbers in equivalent ways, but this skill is not measured by the Manahel EGMA subtasks.

⁴⁸ In each figure, students in the beginning reader proficiency level are shown in red and students in the proficient reader level are shown in blue. These tasks were untimed so only accuracy scores are shown.

Figure 58. Distribution of Number Recognition Accuracy Scores and Fluency Rates for Beginning and Proficient Readers

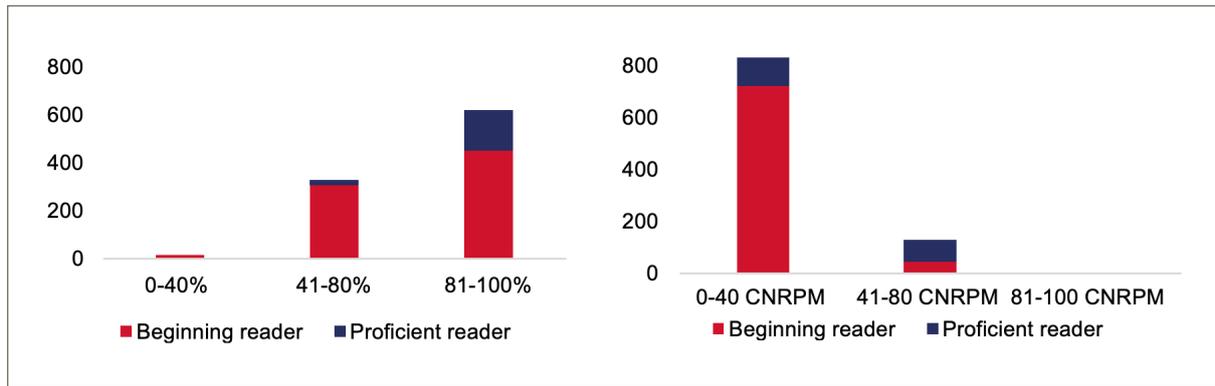


Figure 59. Distribution of Number Recognition Accuracy Scores by Gender

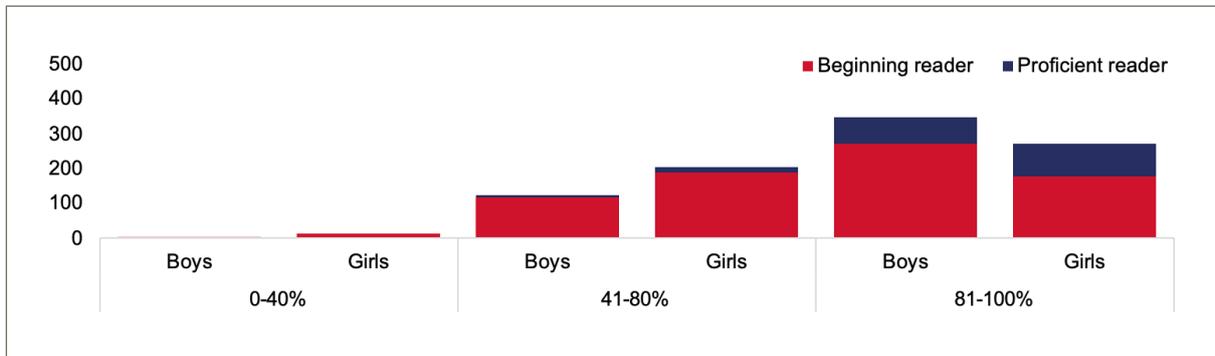


Figure 60. Distribution of Number Discrimination Accuracy Scores for Beginning and Proficient Readers

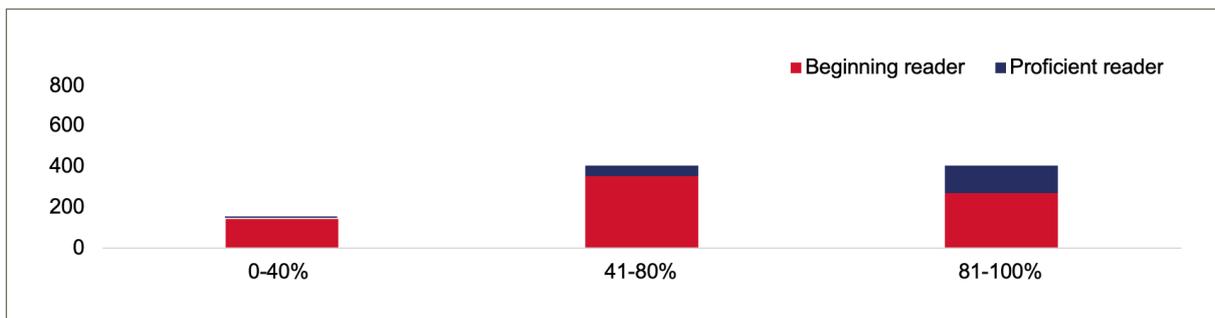


Figure 61. Distribution of Number Discrimination Accuracy Scores by Gender

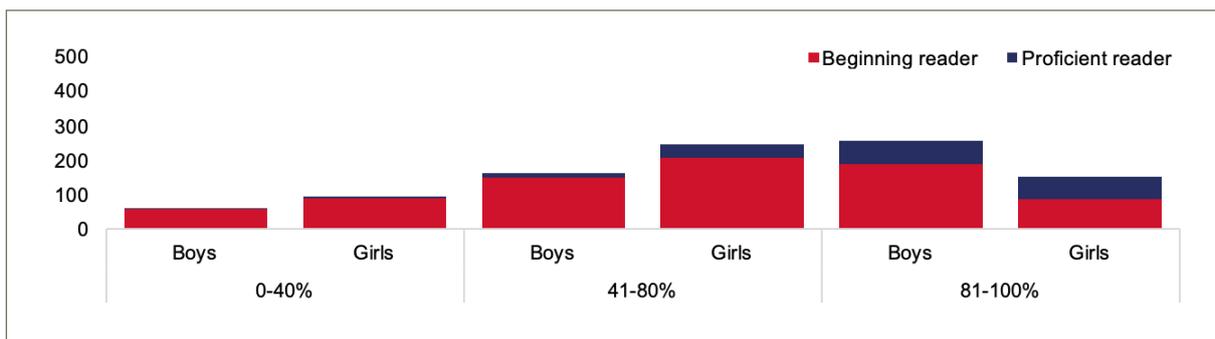


Figure 62. Distribution of Missing Number Accuracy Scores for Beginning and Proficient Readers

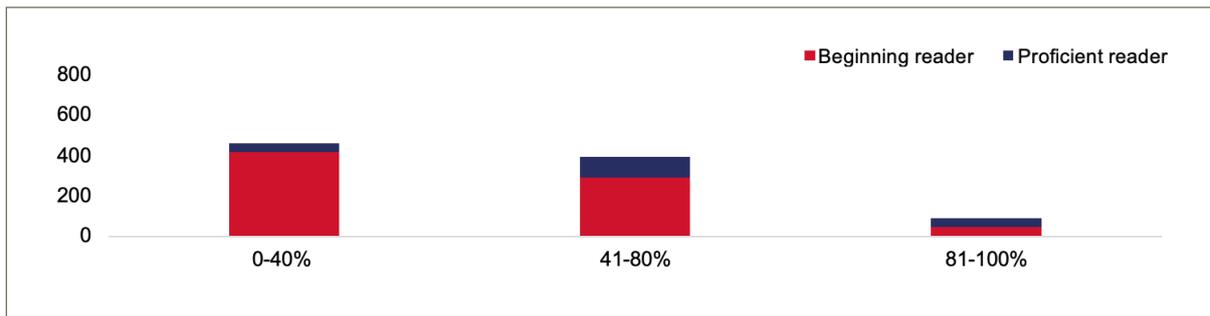
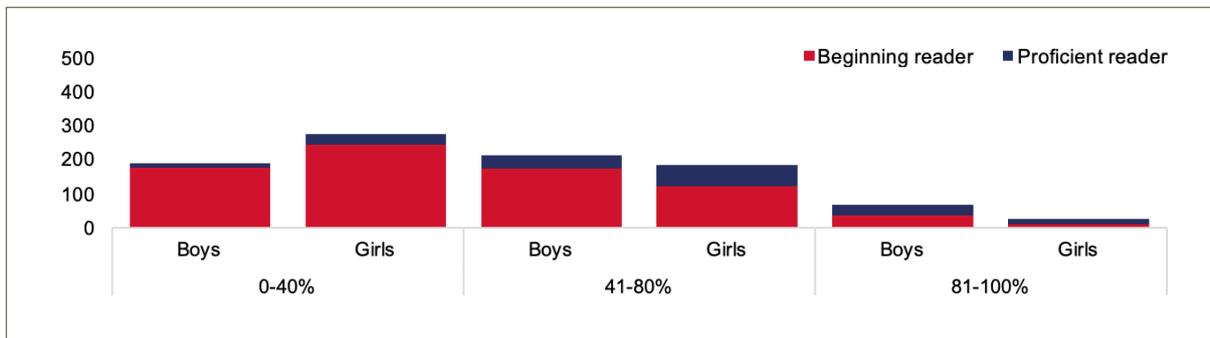


Figure 63. Distribution of Missing Number Accuracy Scores by Gender



Trends for the three whole numbers subtasks are similar. The majority of beginning readers have number recognition accuracy scores of 31% correct and above. Proficient readers tended to have accuracy scores of 50% correct and above. The majority of all students — regardless of reading proficiency — had a fluency rate of 0–40 CNPM for number recognition (Figure 58). Similar trends in accuracy scores were observed when students were asked to discriminate between numbers and identify a missing number in a pattern. Beginning readers tend to have number discrimination accuracy scores of 41–80% correct while proficient readers tend to have accuracy scores that are 41% correct and higher (Figure 60). On the missing number subtask, beginning readers scored 0–40% correct while proficient readers tended to score 41–80% correct.

These results together indicate that students who are beginning readers stand to benefit from mathematics instruction aimed at improving whole number skills such as number discrimination and identifying missing numbers in a pattern. Students who are proficient readers are most likely to benefit from continued support specifically in missing number.

Table 11 provides demographic descriptions of students with low performance, moderate performance and high performance on the number recognition, number discrimination and missing number subtasks.

Table 11. Student Profiles for Whole Number Skills by Performance Category

	0–40% Items Correct	41–80% Items Correct	81–100% Items Correct
Number Identification Subtask	<ul style="list-style-type: none"> • More likely to be Wave B • More likely to have missed school in past 5 days 	<ul style="list-style-type: none"> • More likely to be from District 3 of Province A 	<ul style="list-style-type: none"> • More likely to be from District 4 of Province A • More likely to have attended school for the past 5 days
Number Discrimination Subtask	<ul style="list-style-type: none"> • More likely to be in Province A • Within Province A, more likely to be from District 3 and District 4 • More likely to have moved 6 or more times 	<ul style="list-style-type: none"> • More likely to be from Province B • Within Province A, more likely to be from District 1 or District 2 • More likely to be in QE schools than AO schools 	<ul style="list-style-type: none"> • More likely to be from Province B • Within Province A, more likely to be from District 2 • More likely to be AO schools than QE schools • More likely to have moved 5 times or fewer • More likely to have attended school for the past 5 days
Missing Number Subtask	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 3 or District 5 • More likely to be Wave A • More likely to be in new QE school than AO schools • More likely to have missed school in past 5 days 	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 4 • More likely to have missed school in past 5 days 	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 4 • More likely to be Wave C • More likely to be in AO schools • More likely to have attended school for the past 5 days

To bridge the gap between low and high performance in whole numbers, instruction needs to focus on improving low performers’ accuracy in number discrimination and missing number. Students in Province A — specifically District 3 and District 5 districts — can benefit from instructional support in all three subtasks. Students in Wave A schools could benefit from missing number instruction.

OPERATIONS

Operations include students' ability to add and subtract whole numbers.⁴⁹ Four EGMA subtasks measure students' ability in operations: addition level 1, addition level 2, subtraction level 1 and subtraction level 2.⁵⁰

Students' performance in operations is moderately related to whole numbers and to real-world problems — suggesting that students with stronger skills in operations also have stronger performance in other mathematics areas. The relationship between operations and reading mechanics and understanding were weak, while the relationship with reading comprehension was moderate. This suggests that students who struggle with reading comprehension are also likely to struggle with mathematics operations.

As all students participated in the operations level 1 subtasks, results for addition level 1 and subtraction level 1 are presented together. Figure 64 shows the distribution of addition 1 accuracy scores and fluency rates. Figure 66 shows the distribution of subtraction 1 accuracy scores and fluency rates.⁵¹

Figure 64. Distribution of Addition 1 Accuracy Scores and Fluency Rates for Beginning and Proficient Readers

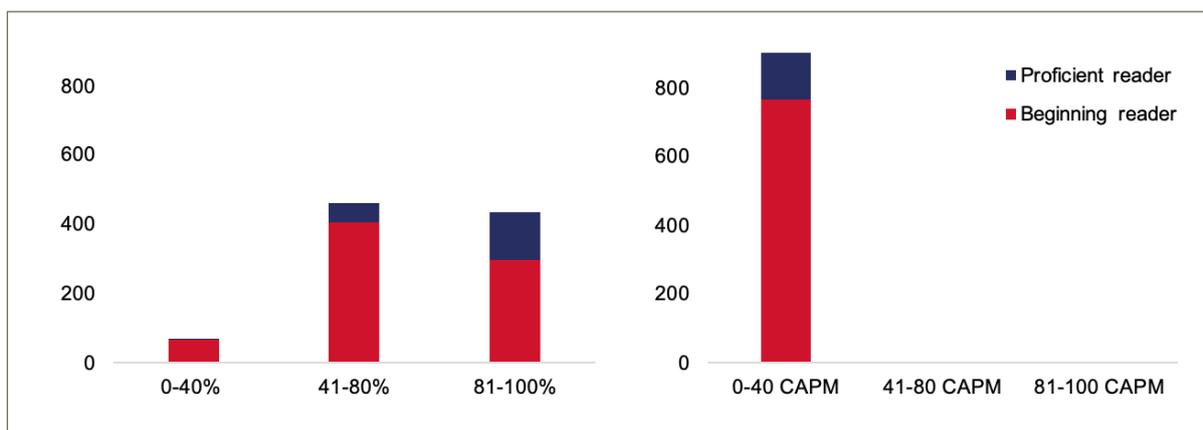
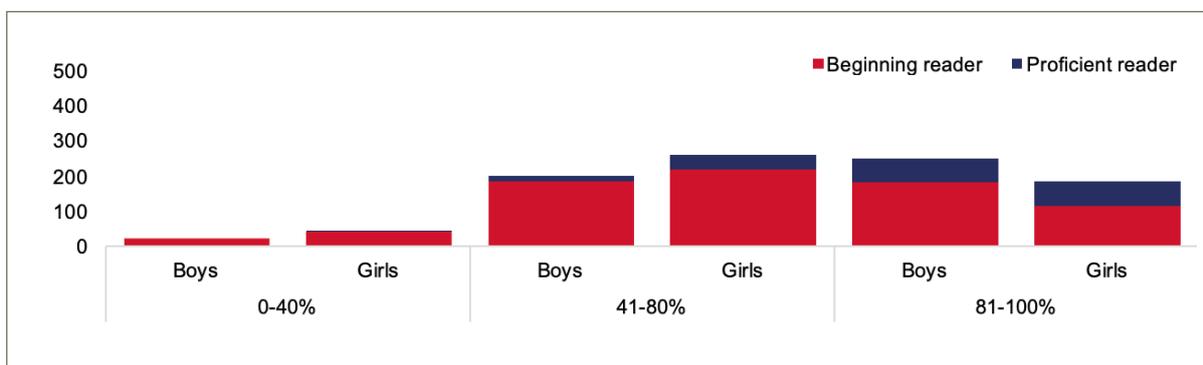


Figure 65. Distribution of Addition 1 Accuracy Scores by Gender



⁴⁹ USAID (2019). Global Proficiency Framework Reading and Mathematics, Grades 2-6 [PDF title]. Retrieved from: <https://www.edulinks.org/sites/default/files/media/file/GAML6-REF-16-GLOBAL-PROFICIENCY-FRAMEWORK.pdf> on December 30, 2019. Note that operations also includes the ability to multiply and divide quantities but this skill is not measured by the Manahel EGMA subtasks.

⁵⁰ Students who answered at least one item on a Level 1 subtask were eligible to take Level 2 subtasks.

⁵¹ In each figure, students in the beginning reader proficiency level are shown in red and students in the proficient reader level are shown in blue.

Figure 66. Distribution of Subtraction 1 Accuracy Scores and Fluency Rates for Beginning and Proficient Readers

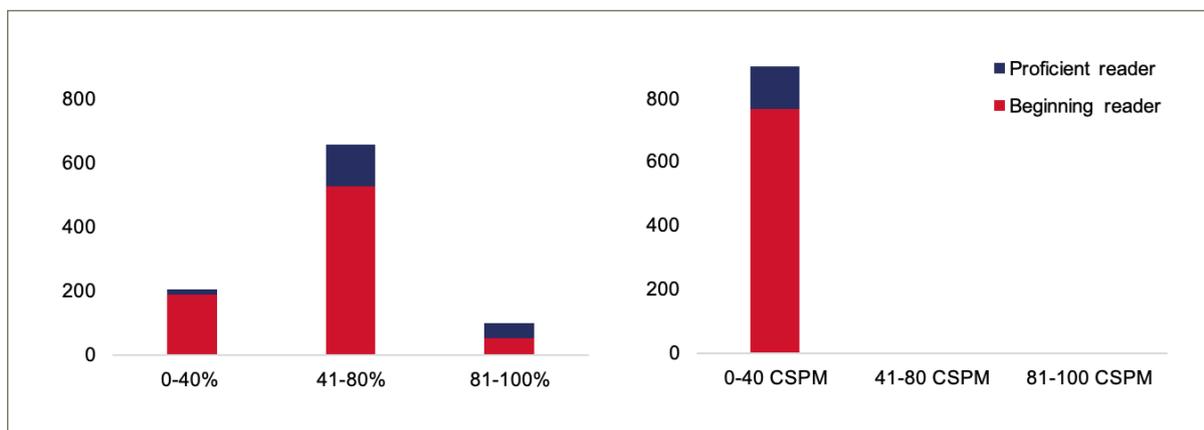
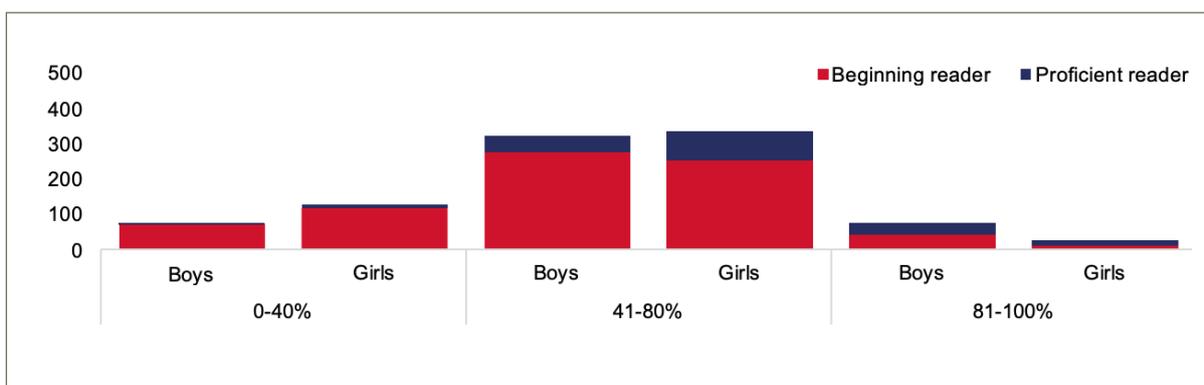


Figure 67. Distribution of Subtraction 1 Accuracy Scores by Gender



The majority of beginning readers have addition 1 accuracy scores of 61% correct and above but subtraction 1 accuracy scores between 31-70% correct. Addition and subtraction 1 fluency rates were fewer than 10 CAPPm and fewer than 10 CSPPM — for beginning and proficient readers alike. This suggests that either all students need support in addition and subtraction or that, in the absence of benchmarks for mathematics, fluency rates below 40 CAPPm and 40 CSPPM are adequate.

Level 2 addition and subtraction subtasks were completed only by students who answered at least one item correctly on the corresponding level 1 subtask.⁵² Figure 68 and Figure 70 present addition and subtraction level 2 accuracy scores, respectively, followed by their corresponding gender breakdown.⁵³ The majority of beginning readers tend to have addition and subtraction level 2 accuracy scores between 0-40% correct. However, proficient readers' accuracy scores were varied and did not trend into low, moderate or high- performance categories. These results indicate that students who are beginning readers may benefit from instructional support on multidigit addition and subtraction problems.

⁵² There were 34 students who did not get any addition level 1 items correct and therefore did not take the addition level 2 subtask. There were 55 students who did not get any subtraction level 1 items correct and therefore did not take the subtraction level 2 subtask.

⁵³ Fluency rates are presented in Annex E.

Figure 68. Distribution of Addition Level 2 Accuracy Scores for Beginning and Proficient Readers

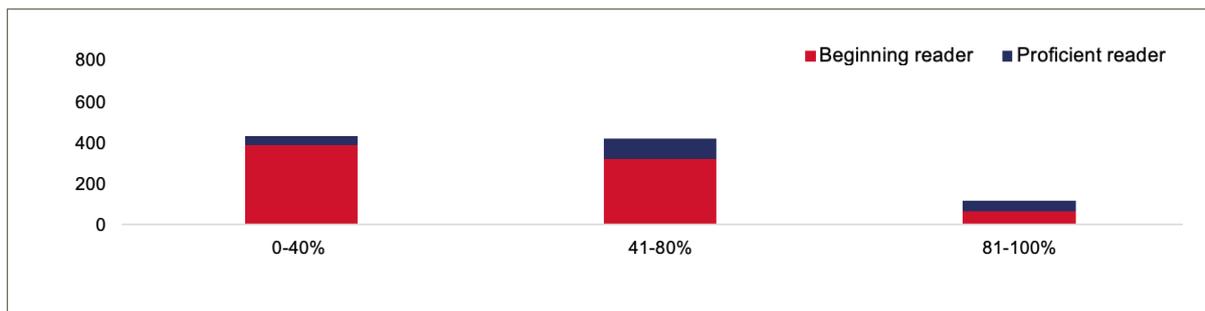


Figure 69. Distribution of Addition 2 Accuracy Scores by Gender

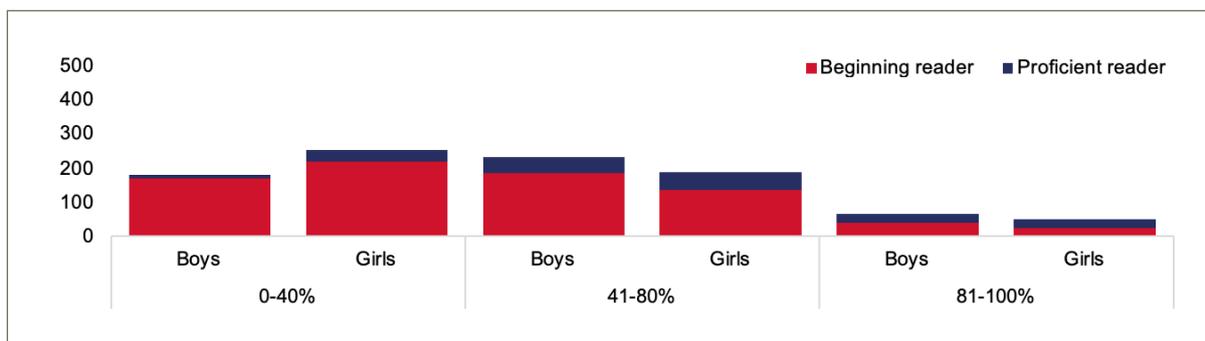


Figure 70. Distribution of Subtraction Level 2 Accuracy Scores for Beginning and Proficient Readers

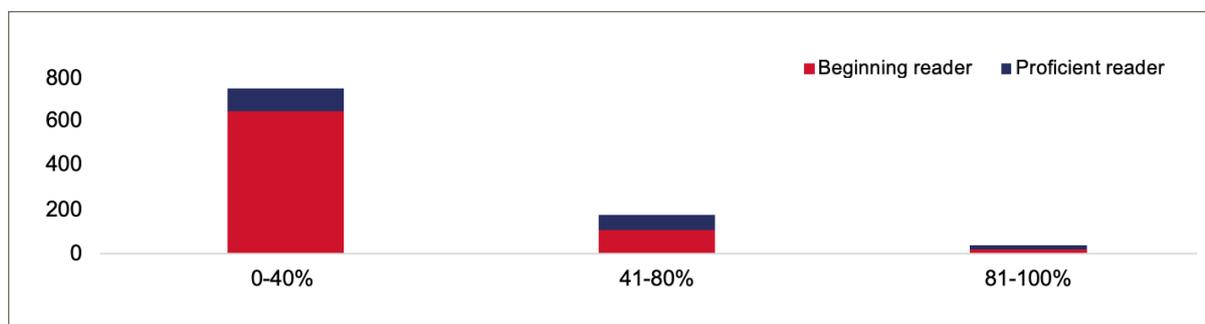


Figure 71. Distribution of Subtraction 2 Accuracy Scores by Gender

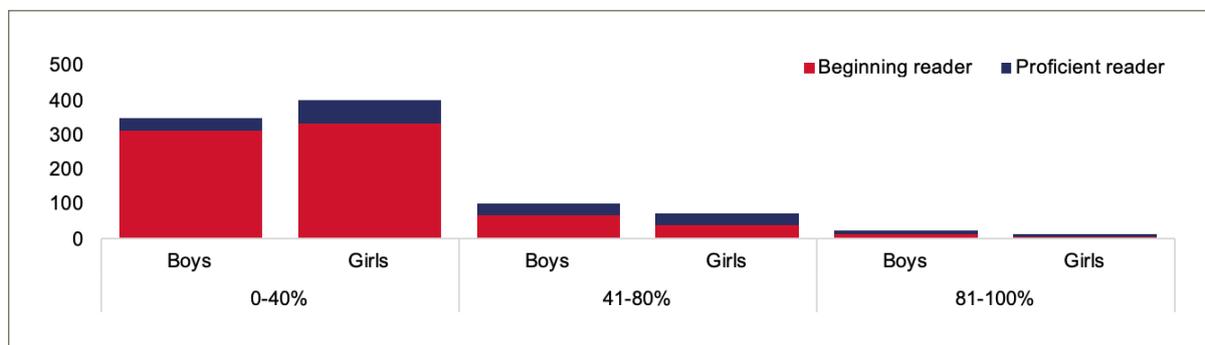


Table 12 provides demographic descriptions of students with low performance, moderate performance and high performance on the addition level 1, addition level 2, subtraction level 1 and subtraction level 2 subtasks.

Table 12. Student Profiles for Operations Skills by Performance Category

	0-40% Items Correct	41-80% Items Correct	81-100% Items Correct
Addition Level 1	<ul style="list-style-type: none"> • More likely to be in Province A • Within Province A, more likely to be from District 3 • More likely to have moved 6 or more times • More likely to have missed school in past 5 days 	<ul style="list-style-type: none"> • More likely to be from Province B • Within Province A, more likely to be from District 1 or District 3 • More likely to be in evening shift schools • More likely to have moved 6 or more times • More likely to have missed at school in past 5 days 	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 4 • More likely to be in evening shift schools • More likely to have attended school for the past 5 days
Addition Level 2	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 3 • More likely to have moved 6 or more times 	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 1 • More likely to have moved 5 times or fewer 	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 1
Subtraction Level 1	<ul style="list-style-type: none"> • Within Province A, more likely to be from District 3 or District 5 	<ul style="list-style-type: none"> • More likely to be Wave C • More likely to be in QE schools 	<ul style="list-style-type: none"> • More likely to be Wave C or non-wave schools • More likely to be in AO schools
Subtraction Level 2	<ul style="list-style-type: none"> • In Province A, more likely to be from District 3 • More likely to have moved 6 or more times • More likely to have missed school in past 5 days 	<ul style="list-style-type: none"> • In Province A, more likely to be from District 1 • More likely to have moved 5 times or fewer 	<ul style="list-style-type: none"> • More likely to have attended school for the past 5 days

Using the reading benchmarks to compare mathematics performance of beginning and proficient readers, the data suggest that mathematics instruction needs to focus on improving low performers' accuracy in level-2 difficulty addition and subtraction problems. Students in District 3 and District 5 districts need particular attention and exploring how mathematics instruction in District 4 district is supporting high performers can lead to further insights. However, to bridge the gap between low and high performing students most effectively in operations, benchmarks for mathematics proficiency need to be developed to truly compare beginning and proficient students' performance.

REAL-WORLD PROBLEMS

The real-world problems skill includes students’ ability to solve real-world problems that involve operations on quantities.⁵⁴ One EGMA subtask measure students’ ability in real-world problems: word problems. This subtask only included three items, which the enumerator read aloud to the student.

Students’ performance in word problems is weak-moderately related to whole numbers and to operations problems – suggesting that students with stronger skills in word problems also have stronger performance in whole numbers and operations. The relationship between real-world problems and reading skills were weak. If students were required to read the word problem for themselves, a stronger relationship between reading skills and real-world problems may have been observed.

Figure 72 shows the distribution of word problem accuracy scores, followed by the gender breakdown in Figure 73.⁵⁵ The majority of beginning readers have accuracy scores of 31-40% correct or 61-70% correct, which translates to one or two of the three items correct. Proficient readers have accuracy scores of 61-70% or 91+% correct, which relates to two or three out of three items correct.

Figure 72. Distribution of Word Problems Accuracy Scores for Beginning and Proficient Readers

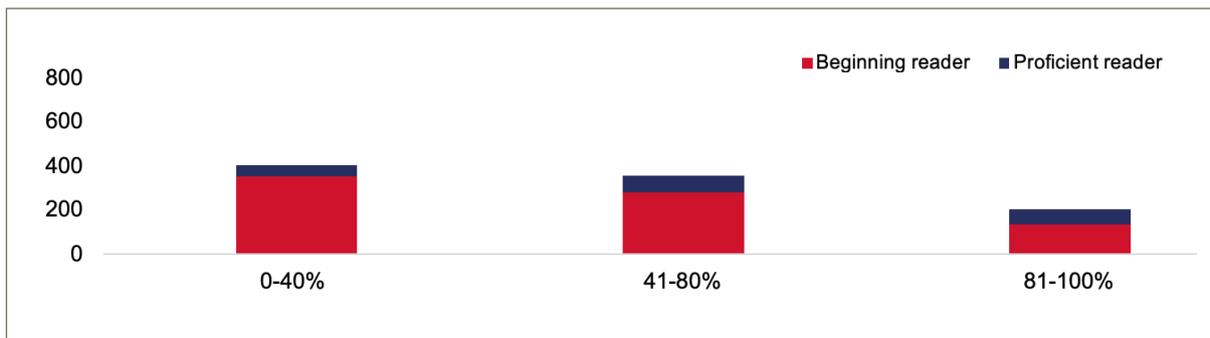
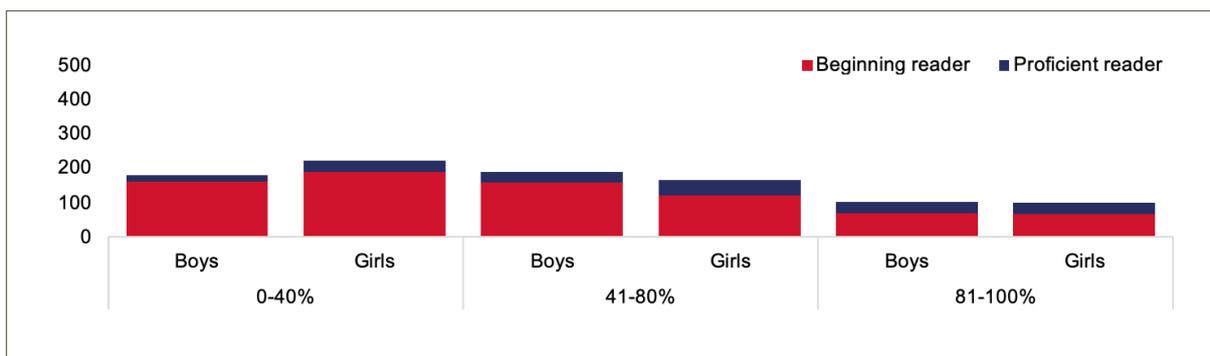


Figure 73. Distribution of Word Problems Accuracy Scores by Gender



⁵⁴ USAID (2019). Global Proficiency Framework Reading and Mathematics, Grades 2-6 [PDF title]. Retrieved from: <https://www.edulinks.org/sites/default/files/media/file/GAML6-REF-16-GLOBAL-PROFICIENCY-FRAMEWORK.pdf> on December 30, 2019. Note that operations also include the ability to multiply and divide quantities, but this skill is not measured by the Manahel EGMA subtasks.

⁵⁵ In each figure, students in the beginning reader proficiency level are shown in red and students in the proficient reader level are shown in blue.

Table 13 provides demographic descriptions of students with low performance, moderate performance and high performance on the word problems subtask.

Table 13. Student Profiles for Real-World Problems Skills by Performance Category

	0-40% Items Correct	41-80% Items Correct	81-100% Items Correct
Word Problems Subtask	<ul style="list-style-type: none"> • More likely to be from Province A • More likely to be in QE schools • More likely to have moved 6 or more times • More likely to have missed school in past 5 days 	<ul style="list-style-type: none"> • More likely to be from Province A • More likely to be in AO schools 	<ul style="list-style-type: none"> • More likely to be from Province B • More likely to be in morning shift schools • More likely to be AO schools

Given the weak relationship between reading skills and real-world problems — and given the small number of items on the word problems subtask — it is not possible to determine specific areas of support on real-world problems with confidence for beginning readers.



Conclusions and Recommendations

Conclusions

RESEARCH QUESTION 1: STUDENT PERFORMANCE IN 2019

- **The majority of Manahel students at the beginning of Grade 3 were beginning readers.** Students' EGRA performance classified about one in every two students as a beginning reader, meaning that they could read between one and 22 CWPM on a story and answered fewer than 80% of reading comprehension questions correctly. Only one-quarter of students classified as progressing or proficient readers. This indicates an ongoing need for foundational reading skills development for students across the Manahel project.
- **A significantly larger proportion of girls than boys were proficient readers, but boys outperformed girls on all mathematics subtasks.** Although the proportions of girls and boys in the non-reader, beginning and progressing levels were comparable, the proportion of girls in the proficient reader level was significantly higher than the proportion of boys. In contrast, boys had significantly higher accuracy scores and lower proportions of zero scores than girls on all EGMA subtasks. Gender appears to be an important factor in reading and mathematics performance for Manahel students.
- **Nonword reading skills are not strongly related to other foundational reading skills or reading comprehension.** Although there are relationships between the letter sound identification, ORF and reading comprehension subtasks, nonword reading skills did not appear to be related to students' ability to read passages and comprehend what they read. This may indicate that the decoding skills measured through the nonword reading subtask may not be a useful measure of reading skills development for Manahel students. Conversely, it is a good differentiator between proficient readers and progressing or emerging readers; indeed, it may help illuminate a skill that most developing readers tend to struggle with. However, the coverage of decoding skills in Manahel materials should be reviewed to inform instruction and determine whether the subtask should be retained at endline.

- **Students' reading skills are not limited by their ability to understand language, as evidenced by their stronger performance on listening comprehension.**

There was not a strong relationship between the mechanics of reading and reading comprehension skills with the understanding skill. Similarly, there was not a strong relationship between reading performance and listening comprehension performance. Additionally, a majority of beginning readers — those who read between one and 22 CWPM but achieved less than 80% accuracy on reading comprehension — responded correctly to at least four out of six listening comprehension questions.

RESEARCH QUESTION 2: COMPARISON OF 2017 AND 2019 STUDENT OUTCOMES

- **2017 and 2019 student learning outcomes are not comparable.** The number of changes to the instrument, the shifting context and the transient nature of students due to high levels of displacement in 2019 make comparisons problematic. Further details can be found in the Limitations section of this report, as well as Annex G.

RESEARCH QUESTION 3: OUTCOMES BY SCHOOL TYPE

- **Comparison by the AO and QE labels are not reliable,** as the labels may not accurately reflect the interventions experienced by the students and teachers who were assessed in each school in 2019.
- **On average, student reading proficiency levels did not differ across AO and QE schools.** Nevertheless, AO students performed better on specific reading and mathematics subtasks. AO students had higher ORF and listening comprehension accuracy than students in QE schools, whilst QE students had lower proportions of zero scores on letter sound identification and nonword reading. Additionally, AO students had significantly higher accuracy scores on all EGMA subtasks than QE students, except on addition level 2 subtask where they performed comparably. Although these differences are not captured in reading proficiency levels, AO students appear to be outperforming their peers in QE schools. However, as mentioned earlier, it is important to keep in mind that the designations of school type do not necessarily mean evaluators can attribute results to the project's intervention, especially due to the high levels of student — and, anecdotally, teacher — displacement within the provinces. Without further data regarding the movement of teachers and students between schools, current attendance may not reflect the interventions received by teachers and students in early 2019.
- **Overall, results of students from QE schools with mobile libraries and those with fixed libraries are comparable.** Only on the ORF subtask did students in schools with mobile libraries do better than their peers. While this finding is promising for the potential impact of mobile libraries, additional data on usage and access to libraries is needed to identify specific aspects of the mobile libraries that may be associated with higher reading scores. Aspects could include the availability of age-appropriate materials, frequency of access, and tie-in to instruction as they may differ between QE schools with fixed libraries versus mobile libraries.

RESEARCH QUESTION 4: OUTCOMES BY PROVINCE

- **Students in Province B generally performed better in the foundational skills of reading and mathematics than students in Province A.** There were proportionally more beginning readers in Province B and more non-readers in Province A, although comparable proportions of progressing and proficient readers. Province B students outperformed Province A students on letter sound identification and nonword reading; however, students across provinces performed comparably on ORF and reading comprehension. This trend was similar on mathematics skills – students in Province B outperformed students in Province A on number discrimination, missing number and word problems, whilst all students performed comparably on both levels of the addition and subtraction subtasks.

RESEARCH QUESTION 5: LINKING EGRA PERFORMANCE TO MANAHEL READING LEVELS

- **Manahel reading levels for each student in the sample were not provided for analysis.**
- **EGRA is not an appropriate tool to evaluate Manahel reading levels.** The reading levels were not designed to be directly compatible with EGRA, thus only limited comparisons were possible. Future efforts should especially focus on adding or tailoring subtasks that measure Manahel levels 5 and 6, as no comparisons were possible through the EGRA.

RESEARCH QUESTION 6: STUDENT STRESS AND THE CONFLICT

- **Student stressors – including displacement, attendance and tiredness – had predictable relationships with students’ reading and mathematics skills.** Students who suffered less displacement, who missed fewer days of school and who did not report being tired during school generally outperformed their peers. Hunger and household size did not appear to affect reading or mathematics skills.
- **Student levels of displacement and number of new schools attended show that the amount of time spent by students in the 2019 sample in Manahel intervention schools, is unknown.** The study did not ask students how long they have been enrolled in their current school so with the volatility in the region, it is possible that students may be new to their current school.

Recommendations

Based upon the findings and conclusions, STS proposes the following recommendations to be considered throughout Manahel’s programming, as well as the 2021 endline study. Recommendations are listed in order of priority, beginning with implementation decisions and followed by endline study design.

Table 14. Summary of Key Conclusions and Recommendations

Key Conclusion	Recommendation
Manahel Program Implementation	
Overall, the majority of Manahel students were beginner readers at the beginning of grade 3.	<ul style="list-style-type: none"> → The emphasis of Manahel’s work with teachers and teaching resources should shift more to the earlier literacy levels. Continue to emphasise reading mechanics and comprehension in Manahel interventions to help move students from beginner readers to progressing readers. → Provide targeted support to those districts where children have the lowest reading performance, including District 3, District 5 and District 1.
A significantly larger proportion of girls than boys were proficient readers, but boys outperformed girls on all mathematics subtasks.	<ul style="list-style-type: none"> → Consider gender within teacher training and instruction and provide additional support to boys for reading and girls for mathematics to bridge the performance gaps between genders.
Overall, results from QE schools with mobile and fixed libraries were comparable, but schools with mobile libraries outperformed their peers in oral reading fluency.	<ul style="list-style-type: none"> → Alter the way in which AO and QE labels are determined at endline so that labels accurately reflect the interventions received. → Additional data on usage and access to libraries is needed to identify any differences between the schools that may be associated with higher reading scores, including the availability of age-appropriate materials, frequency of access, and tie-in to instruction.
Student stressors, including displacement, lower attendance and tiredness, had predictable relationships with student reading and mathematics skills.	<ul style="list-style-type: none"> → Explore ways for existing supports, such as safeguarding officers, to engage with families to help increase attendance for those students with chronic absenteeism, particularly in more stable areas of Provinces A and B. → Beyond students with acute needs who must be prioritised, more finely triage the needs of students transitioning between schools to target emotional and academic support and balance the disruption to their schooling. → Additional data may be useful to understand factors that may preclude students from attending school regularly or may hinder their performance at school. This data may also help project leaders understand better the confounding relationship between displacement, school changes, learning and underlying changes in community and learning environment.

<p>Students in Province B generally performed better in reading and mathematics foundational skills than students in Province A.</p>	<p>→ Provide targeted and additional support to schools labelled as QE in 2019 in Province A to improve their performance. Understanding the nature of other supports being provided to schools will also be important at endline, particularly if differences in Provinces A and B persist.</p>
<p>Study Design for 2021 Endline</p>	
<p>Nonword reading skills are not strongly related to other foundational reading skills or reading comprehension.</p>	<p>→ While this subtask is a good differentiator between proficient readers and progressing or emerging readers, it has a weaker relationship with reading comprehension than other tasks. Recommend assessing whether the subtask is appropriate for the instructional approach used by Manahel and retain at endline accordingly.</p>
<p>Results from the assessments are student-centred, while interventions are teacher- and school-centred.</p>	<p>→ Include a teacher interview at endline, which asks about teacher displacement and participation in the Manahel program to contextualise learning outcome findings as well as capture participation in other interventions by other organisations. Possible format examples include interviews, interviews using a 'most significant change' survey, or focus groups with teachers, head teachers and/or child safeguarding officers.</p> <p>→ Add follow-up questions within the Student Stressor Survey for more nuanced information on the number of moves and schools attended, length of time at current school, length of time out of school, information on previous school and participation in Manahel interventions.</p> <p>→ Conduct an alignment exercise between Manahel reading levels and international reading levels, such as those proposed in the Global Proficiency Framework.⁵⁶</p>
<p>The delineation between the AO and QE school types was not an appropriate binary for analysis.</p>	<p>→ Adding teacher-provided responses regarding participation in interventions and assign intervention levels based on this information during data analysis.</p>
<p>Comparison between the results of Manahel's predecessor's study in 2017 and the 2019 Manahel results is not a valid comparison.</p>	<p>→ Consider revising the research question to focus on the change in the 2017-2019 gap and 2019-2021 gap.</p> <p>→ Capture demographic information during the endline to determine the comparability of the 2019 Grade 3 population versus the 2021 Grade 3 population.</p> <p>→ Generate an index of conflict, reported at the district level at a minimum, preferably at the school level. The index can be included in analyses to examine the influence of conflict on student learning outcomes, without having to survey individual students directly about their conflict experience.</p>

⁵⁶ Global Alliance for Monitoring Learning, "Global Proficiency Framework for Reading and Mathematics - Grades 2 to 6," October 2019, http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/Global-Proficiency-Framework-18Oct2019_KD.pdf

Lessons Learned

The 2019 study of student reading and mathematics outcomes at the beginning of the Manahel project yielded several lessons learned.

First, the definitions of AO and QE schools should have been used at the time of sampling to determine if any additional data could be provided to the sampling exercise. Additionally, having these definitions at the start of the study would have illuminated the challenge in distinction using these labels. Such knowledge, perhaps, could have allowed an earlier shift in the research questions.

Second, any comparison to the 2017 data was known to be problematic because of the difference in timing of the test and the revisions of the tools. However, because a time-based comparison is required in the logframe, the analysis for 2017 data versus 2019 data was completed. Further discussions with Manahel and legacy DFID earlier in the process may have prevented much back-and-forth and ultimately disregarding of results.

Third, the keen interest in developing student profiles to inform project interventions was a driving factor in the structure of the report and analyses. STS engaged with Manahel to understand how these profiles would influence and inform their decisions. Similar discussions with legacy DFID may have also been useful. The focus of the legacy DFID and Manahel teams on using these data to inform their interventions is extremely empowering and gives all the more reason to an evaluator to present findings in a useful format.

Fourth, the confounding of variables along with insufficient cases at the intersection of each of the variables, did not allow for complex analytical models that could help isolate the effects of one factor of interest from another. For high-interest factors — such as displacement — the survey could have included more items to explore their nature and extent. This type of granular data, instead of a single question, could provide more data that differentiates between subgroups and may have allowed for richer analyses.

Annex A:

Evaluation Framework

In 2019, School-to-School International will conduct a study to measure student learning outcomes in reading and mathematics in a sample of Manahel's intervention schools to answer the research questions outlined below. The results of the 2019 study will also serve as a point of reference for comparison to future student learning outcomes in reading and mathematics.

Research Questions:

1. What proportion of G3 students in 2019 and in 2020 are classified as 'progressing' and as 'proficient' readers using the 2017 definitions for these categories? How do these proportions compare by subgroups?
2. How do Manahel students learning outcomes at the beginning of Grade 3 in reading and mathematics compare to those for students assessed under the previous project in 2017 at the end of Grade 3?
3. How do beginning of grade 3 students' learning outcomes in reading and mathematics in Quality Education schools compare with that of Access Only schools? Within QE schools only, how do QE with fixed libraries compare with those with mobile libraries (as possible)?
4. How do beginning of grade 3 students' learning outcomes in reading and mathematics compare between Province B and Province A?
5. How do beginning of grade 3 students' EGRA findings relate with the eight reading levels that Manahel-supported teachers are using to track learners' progress? What is the relationship between Manahel-developed reading levels and EGRA performance?
6. What do we know about children's current levels of stress and the relationship with their learning?

In October 2019, Manahel will assess Grade 3 Syrian students' reading and mathematics levels. The assessment will use the Arabic-language Early Grade Reading and Early Grade Mathematics Assessment (EGRA and EGMA) instruments previously administered with Grade 3 students under the previous project. The study design assumes that an endline at the end of the Autumn semester in 2020, Grade 3 students will again be assessed. Table 1 below illustrates this proposed study design and Table 2 its alignment with Manahel's required indicator reporting.

Importantly, the evaluation design and the associated research questions focus on comparing differences among Manahel project schools rather than comparing outcomes for students in Manahel schools to outcomes for students in non-Manahel schools. As such, no comparison group is proposed. If the research questions change and focus on comparing outcomes of Manahel students to those of students in non-Manahel schools or the impact of Manahel interventions on students learning outcomes, then a comparison group would be required.

Table 15. Proposed Evaluation Years and Grade

Project intervention grades and years	2017 June (end of school year)	2018	2019 October (beginning of school year)	2020 October (beginning of school year)
Grade 1				
Grade 2				
Grade 3	Previous project's EGRA		Manahel EGRA and EGMA	Manahel EGRA and EGMA

The previous project assessed Grade 3 learners in April 2017. They had received three years and seven months of schooling by that point. We will repeat the assessment of Grade 3 learners with Manahel. However, as we are assessing in October 2019, it should be noted that Grade 4 children would be a closer comparison:

- G3 Previous project = 3y 7m instruction by Apr 2017
- G4 Manahel = 4y 1m instruction by Oct 2019
- G3 Manahel = 3y 1m instruction by Oct 2019

Neither assessment of G3 or G4 students would be directly comparable given the difference in time when the assessment is being carried out. Manahel has asked STS to work with G3 students as this seems closer to the spirit of the logframe wording. Manahel will explore with DFID the challenges related to timing and how that affects comparability.

For the 2019 data collection, a 2-stage sampling approach will be used, beginning with a random sample of schools drawn from a full list of intervention schools (sampling frame) and then a random sample of 10 students selected from those schools on the day of data collection.

A sampling frame that includes these data on all Manahel intervention schools will be required to determine a) if all three strata can be accomplished in the sampling strategy and b) to determine the appropriate sample size for the study.

- Representative sample from the population to allow results to be generalized to the population of Manahel intervention schools
- Selection of schools from Quality Education schools and Access Only schools (strata 1)
- Selection of schools from Province B, Province A, and schools supported by mobile libraries (strata 2)
- Selection of schools that overlap with schools that received intervention through Manahel's predecessor project + Manahel interventions and Manahel only (strata 3)

The sample for the study will be determined based on the parameters above, as well as a consideration of the level to which results will be generalized (at the project level or at a sub-level, such as Quality Education versus Access schools). Additionally, logistic details and challenges to data administration in a conflict-affected context may also affect the sample.

To mitigate some of the expected challenges in a conflict-affected context such as Northwest Syria, the sampling approach will include a thorough replacement strategy for both schools and student-level sampling.

The final sample of schools, classes, and students will be determined based on consultation with the Manahel team. STS and Manahel will train data collectors on the use of appropriate sampling strategies to reach the desired number of respondents according to the selected sampling plan.

Annex B: Assessment and Survey Tools

EGRA Assessment Tools

LETTER SOUND IDENTIFICATION - STUDENT STIMULUS

أمثلة:

	أ	ب	ج	د	هـ	و	ز	ح	ط	ك
١	هـ	نـ	قـ	فـ	بـ	أ	تـ	أ	بـ	طـ
٢	حـ	ـ	أ	مـ	هـ	ـ	سـ	ذـ	صـ	ـ
٣	تـ	ضـ	قـ	عـ	أ	إـ	نـ	و	أ	كـ
٤	و	يـ	نـ	سـ	أ	شـ	عـ	أ	مـ	إـ
٥	ظـ	خـ	قـ	ظـ	جـ	و	صـ	ذـ	تـ	صـ
٦	قـ	بـ	لـ	عـ	نـ	مـ	دـ	قـ	هـ	نـ
٧	يـ	ظـ	دـ	غـ	ثـ	أ	إـ	رـ	ظـ	مـ
٨	أ	و	قـ	هـ	زـ	و	فـ	سـ	كـ	فـ
٩	زـ	يـ	حـ	جـ	خـ	هـ	كـ	عـ	أ	دـ
١٠	كـ	تـ	شـ	حـ	مـ	أ	ظـ	عـ	كـ	طـ

NONWORD READING - STUDENT STIMULUS

أمثلة: ناسب ناسب
 الفلأط ميء
 لا

	٥	٤	٣	٢	١
٥	سَدَع رُ	قَمَلِي	رَرُوك	مَمَع جُج	أَفَ اِل
٥١	نَقَّس	أُ طِيخ	أَ نَد رَأُ	مَجُكُنَا	كُ نُتُو
٥١	بِلَو حِب	أُ شِرَّر قُ	القُ سِين	سِيفُ لُ	كَدُ تُ
٥٢	نُقَمَاخ	يَاشْتُو	هَوُ نَ	قَهَ ص	زَعَاك
٥٢	هَصَد هَهُ	أَ طَلَق	الزُ فُلَقَة	عَعْدُ رُ	أَ وُدِ ه شُ
٥٣	لَعَادِيد	خِفَ جَ	شِعِي	لِيَ بُ	قُمُ نَ
٥٣	زُنَيْد بُ	أَ ضِي	مَرَب	عَلْبُ ه مُ	النَّ دِيَق
٥٤	لِجَا	الكُ رُف	سُدَلِي	عِيرَا	بِ يُنَا
٥٤	كَلَم	قَشَد مَ	مَ نَجَ مُ	سُدَلَد نِي	عِيرِ ج
٥٥	كَبِع	كَأ دَ	مَدُ مُج	رِفَ طُ	رِي نَهَر هُ

ORAL READING FLUENCY AND READING COMPREHENSION

Reading Passage Student Stimulus

سارهٌ نذ مؤدبةٌ بعيدة أبو وهي تُد بالنتظافة والرتيب. وتُحاف
تُعيش تَ معها، مع
لي دروسها تساء دوالدتها في الم طَب.

وفي شاهدت سارهٌ بائعاً متجراً يبيع المأكولات لي من الأمدرك
و ولا قبة من
م

فَفَكَرَتْ أَنْ تَأْكُلَ مِنْهُ.

رَجَعَتْ لِي الْأَدِيَّةَ تَوَحِينِ ذِي تِلْإِلَى طَبِصَاحَتْ وَوَضَتْ دَهَا عَلَى
أ الم

طِينَهَا، أَسْرَعَتْ تَهَا أُمَهَا لِي الْكِرَالِ صَدِي تَسَارَةُ الطَّبِيبِ
لَمُرُّ د وَأَذ
و

أَكَلَتْ مِنَ الْإِبَاءِ الْمَتَّجِرِ.

أَجْرَى الطَّبِيبُ فُحُوصَاتٍ بِهَوِ فَدَلَّهَا الدَّوَاءَ وَنَصَّحَهَا بِمِ الْأَطْعَمَةِ
تَتَاوَد

المكشوا لأنها ملوثةٌ يُيرُ صِحَّةً، تسارةٌ: نَأُ رَهَذَا الْإِبَاءِ
و غ قَالَ لَ ر فِعْ بَدَأَ
ل

Reading Comprehension Questions – No Stimulus

Questions		Answers
1	حول من تدور القصة؟	[ساره]
2	أين تحاول ساره مساعدة والدتها؟	[في المطبخ]
3	ماذا يعمل البائع المتجول على مقربة من المدرسة؟	[يبيع المأكولات]
4	ماذا أخبرت ساره الطبيب عندما ذهبت إليه؟	[بما أكلت من البائع المتجول]
5	لماذا نمتنع عن تناول الطعام من الباعة المتجولين؟	[لأنها ملوثة وغير صحية]

LISTENING COMPREHENSION

Listening Passage - No Stimulus

على صبحٍ، تَدَحُّ بِأُمِّي وَتُضَلِّدُ الفطورِ مِمَّا لَدَيْ البقرةِ يَسْتَدِينُ
بِقَرَّتِهَا، حَذَاتِ مُمْتَدِي الإسجدِ بَقَطَّتْ لَط

فَلَمَّ أَرَفَ فطوراً لِي الطاولَةِ أُمِّي فَوَدَّ خَرَجَتْ عِنْدَ الجيرانِ، وَفِي قَلْبِ، وَفِي
كَانَ لِلأبْحَثِ تَالَعَتْهَا رَبُّ د

الدُّنْيَا. غَيْرَ أَنَّهَا لَمْ تَعُدُّوْا عَلَيَّهَا فَبِخُنَانٍ لَأَنَّهَا لَمْ تَعُدُّوْا. فَبَكَتْ دُونَاً لَأَنَّ
بَكَتْ رَعَايَ نَهَا هَا بُهَّا كَا

كَثِيراً فَلَمَّا لَهتْ سَدِمَتْ ضَجِيجاً وَبَدَأَتْ تَأْكُلُ خِيراً مِنْ سَلَّةِ
ي الْم بِبَقَرَةٍ خَضَارِ

Listening Comprehension Questions - No stimulus

	Questions	Answers
1	ماذا تفعل الأم كل صباح؟	[تحلب الأم بقرتها وتحضر لنا الفطور]
2	ماذا حدث للأم في الاسطبل؟	[لم تجد البقرة]
3	لماذا لم يجد الطفل فطوراً على الطاولة؟	[لأن الأم لم تحلب البقرة]
4	أين بحثت الأم عن البقرة؟	[بحثت عنها عند الجيران وفي الحقل وقرب النهر]
5	لم بكت الأم؟	[أنها تحب بقرتها لم تجد البقرة]
6	ما الذي أحدث الضجيج في المطبخ؟	[البقرة]

EGMA ASSESSMENT TOOLS

Number Identification - Student Stimulus

A

٣٠	١٢	٠	٩	٢
٤٨	٢٣	٣٩	٤٥	٢٢
٦٥	٨٧	٧٤	٣٣	٩١
٩٨٩	٧٣١	٥٨٠	٢٤٥	١٠٨

Number Discrimination - Student Stimulus Example Sheet

B1

٨	٤
---	---

٢٢	٢١
----	----

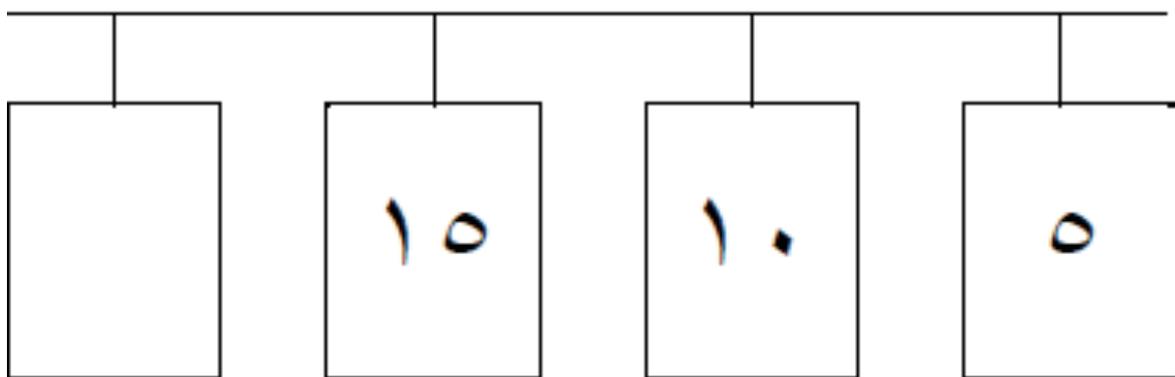
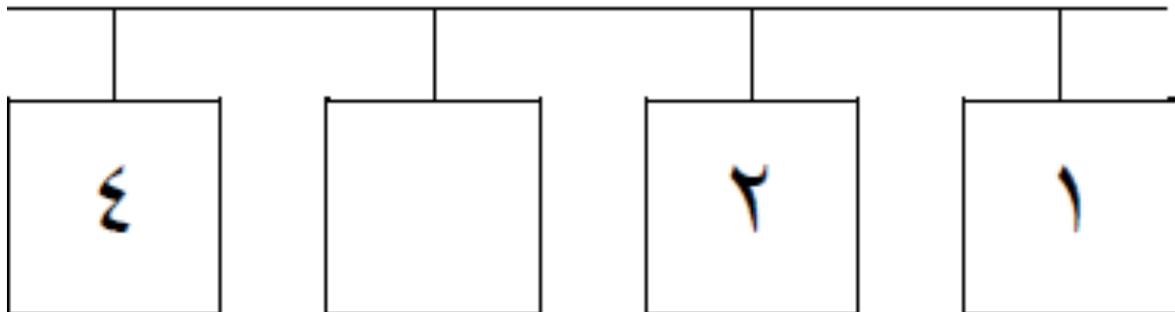
B2

۷	۵
۱۲	۲۵
۳۴	۲۹
۵۸	۴۸
۶۵	۶۷

B3

٩٤	٧٨
١٤٦	١٥٣
٢٨٧	٥٣٧
٦٥٠	٦٠٥
٩٦٥	٩٦٧

C1



C2

	٧	٦	٥
١٧		١٥	١٤
٦.	٥.		٣.
٥...	٤...	٣...	
	٦	٤	٢

C3

٣٥١		٣٤٩	٣٤٨
٢٢	٢٤		٢٨
٤٥		٣٥	٣٠
	٥٣٠	٥٤٠	٥٥٠
١٨		٨	٣

D1

$\square = 3 + 1$
$\square = 2 + 3$
$\square = 2 + 6$
$\square = 0 + 4$
$\square = 3 + 3$
$\square = 8 + 0$
$\square = 3 + 7$
$\square = 7 + 2$
$\square = 0 + 0$
$\square = 8 + 2$

D2

$\square = 3 + 11$
$\square = 4 + 13$
$\square = 3 + 16$
$\square = 6 + 8$
$\square = 8 + 7$
$\square = 7 + 9$
$\square = 8 + 8$
$\square = 11 + 2$
$\square = 2 + 10$
$\square = 10 + 8$

D3

$$\square = 3 + 16$$

$$\square = 7 + 18$$

$$\square = 12 + 24$$

$$\square = 37 + 22$$

$$\square = 26 + 38$$

E1

$\square = 1 - 4$
$\square = 2 - 0$
$\square = 2 - 1$
$\square = 0 - 9$
$\square = 3 - 6$
$\square = 0 - 1$
$\square = 3 - 10$
$\square = 7 - 9$
$\square = 0 - 10$
$\square = 1 - 10$

E2

$\square = 3 - 14$
$\square = 4 - 17$
$\square = 3 - 19$
$\square = 6 - 14$
$\square = 7 - 10$
$\square = 9 - 16$
$\square = 8 - 16$
$\square = 11 - 13$
$\square = 2 - 12$
$\square = 10 - 18$

E3

$$\square = 3 - 19$$

$$\square = 7 - 25$$

$$\square = 12 - 36$$

$$\square = 37 - 59$$

$$\square = 26 - 64$$

Word Problems – No Student Stimulus

Question	Answer
1	(٤)
2	(٧)
3	(٣)

STUDENT STRESSORS SURVEY

Consent

رح اسألك كم سؤال عن الحرب. اذا مانك مرتاح تجاوب عالسؤال قلبي و منزوح عسؤال ثاني. اذا مانك مرتاح تجاوب عليهن كان كمان مافي مشكلة معلية اسألك كم سؤال عن الحرب؟

Number	Prompt in Arabic	Prompt in English
1	كم واحد ساكن بالبيت معك ؟ /كم واحد انتو بالبيت؟	How many people live with you at your home?
2	اضطريت تترك البيت شي مرة بسبب الحرب و القصف؟	Have you ever been forced to move to a different home because of the war?
3	كم مرة صرت مغير البيت؟	How many times have you had to move because of the war?
4a	رحت عمدارس ثانية غير اللي انت فيها هلق؟	Have you ever attended a different school, other than this one?
4b	كم مدرسة صرت رايح عليهن مع مدرستك هي؟	How many different schools have you gone to, including this one?
5	في شي مرة كنت مو مضطر تروح عالمدرسة؟	Have there been any times when you haven't had a school to go to?
6	حس انك جوعان بالمدرسة؟ (أبداً، أحياناً، مراراً، دائماً)	How often do you feel hungry at school? (Always, Often, Sometimes, Never)
7	بتحس انك تعبان بالمدرسة؟ (أبداً، أحياناً، مراراً، دائماً)	How often do you feel tired at school? (Always, Often, Sometimes, Never)
8a	قديه بتحس انك بأمان بالمدرسة؟ (أمن جداً، أمن، غير أمن، خطير جداً)	How safe do you feel at this school? (Very safe, Safe, Unsafe, Very Unsafe)
8b	ليش؟	Why?
9	كيف بتروح عالمدرسة بالعادة؟	How do you normally travel to school?
10	كم دقيقة بديك لتوصل عالمدرسة؟	How many minutes does it take you to travel to school?
11a	قديه بتحس بالأمان عطريق المدرسة و انت رايح و راجع؟ (أمن جداً، أمن، غير أمن، خطير جداً)	How safe do you feel on the way to and from school? (Very safe, Safe, Unsafe, Very Unsafe)
11b	ليش؟	Why?
12a	قديه بتحس انك بأمان بالبيت؟ (أمن جداً، أمن، غير أمن، خطير جداً)	How safe do you feel at home? (Very safe, Safe, Unsafe, Very Unsafe)
12b	ليش؟	Why?

Note: Each question had a 'Refused to Answer' option so that students could opt out of each question, as well as the survey as a whole.

HEADTEACHER SURVEY

Consent

تم اختيار مدرستك للمشاركة في دراسة يقوم بها مشروع مناهل. مشاركتك اليوم مهمة جداً لنا، لكن يمكنك عدم المشاركة بحال كنت تريد ذلك. ستكون الدراسة حول ديموغرافية مدرستك و حجم الصفوف الموجودة. بحال كنت لا تريد الإجابة عن أي من الأسئلة، يمكننا تخطي السؤال. مجدداً، لا تتوجب عليك المشاركة في هذه الدراسة

هل توافق على المشاركة في هذه المقابلة؟

Number	Prompt in Arabic	Prompt in English
1	ما هو عدد الأقسام في الصف الثالث؟	How many sections are there for Grade 3?
2	ما هو القسم المختار للقيام بتقييم EGRA ؟	Which section was selected for the EGRA administration?
3	ما هو اسم المعلم؟	What is the teacher's name?
4	ما هو عدد الأطفال الذكور EGRA خلال تحديد الطلاب؟ الحاضرين في الصف المختار للقيام بتقييم	How many boys are present in the classroom sampled for EGRA at the time of assessment?
5	ما هو عدد الأطفال الإناث EGRA خلال تحديد الطلاب؟ الحاضرات في الصف المختار للقيام بتقييم	How many girls are present in the classroom sampled for EGRA at the time of assessment?
6	ما هو الأطفال الذكور المسجلين في الصف الثالث؟	How many boys are enrolled in Grade 3?
7	ما هو عدد الأطفال الإناث المسجلين في الصف الثالث؟	How many girls are enrolled in Grade 3?
8	ما هو عدد الأطفال الذكور المسجلين في قسم الصف EGRA؟ الثالث الذي سيخضع لتقييم	How many boys are enrolled in Grade 3 in the section selected for the EGRA?
9	ما هو عدد الأطفال الإناث المسجلين في قسم الصف EGRA؟ الثالث الذي سيخضع لتقييم	How many girls are enrolled in Grade 3 in the section selected for the EGRA?

Annex C: EGRA Tool Changes

Table C.1. Summary of EGRA Item-Level Changes

	Letter Sound Identification	Nonword Reading	Oral Reading Fluency
Unchanged	7	9	70
Modifier Added	75	7	12
New Word	18	34	0
Total	100	50	82

Table C.2. Details of Nonword Reading Subtask Item-Level Changes

Nonword Reading		
Item	Change	Reason
1	Modifier Added	
2	New	Replaced, not appropriate pronunciation
3	Modifier Added	
4	New	Revised - added letters to align with Arabic linguistic rules
5	New	Revised - added letters to align with Arabic linguistic rules
6	New	Did not align with Arabic linguistic rules
7	New	Revised - added letters to align with Arabic linguistic rules
8	Modifier Added	
9	Modifier Added	
10	New	Revised - added letters to align with Arabic linguistic rules
11		
12	New	Revised - added letters to align with Arabic linguistic rules
13	Modifier Added	
14	New	Inappropriate meaning
15	New	Revised - added letters to align with Arabic linguistic rules
16	New	Replaced, not appropriate pronunciation
17	New	Similar sound to item 10 and does not align with Arabic linguistic rules
18	New	Did not align with Arabic linguistic rules
19	New	Revised - added letters to align with Arabic linguistic rules
20	New	Did not align with Arabic linguistic rules
21		
22	New	Did not align with Arabic linguistic rules
23	New	Did not align with Arabic linguistic rules
24	New	Did not align with Arabic linguistic rules
25	New	Did not align with Arabic linguistic rules
26	New	Did not align with Arabic linguistic rules
27	New	Did not align with Arabic linguistic rules
28	New	Did not align with Arabic linguistic rules
29	New	Did not align with Arabic linguistic rules
30	New	Did not align with Arabic linguistic rules
31		
32	New	Revised - added letters to align with Arabic linguistic rules

Nonword Reading		
Item	Change	Reason
33		
34		
35		
36	New	Revised - added letters to align with Arabic linguistic rules
37	Modifier	
38		
39		
40	Modifier	
41	New	Did not align with Arabic linguistic rules
42	New	Did not align with Arabic linguistic rules
43	New	Did not align with Arabic linguistic rules
44	New	Did not align with Arabic linguistic rules
45		
46	New	Did not align with Arabic linguistic rules
47	New	Did not align with Arabic linguistic rules
48	New	Did not align with Arabic linguistic rules
49	New	Did not align with Arabic linguistic rules
50	New	Did not align with Arabic linguistic rules

Annex D: Proficiency Band Results

Table D.1. Proportion of Students Per Proficiency Band, by Gender, Province, and School Type

Proficiency Band	Total		Gender				Province				School Type			
			Boys		Girls		Province A		Province B		Access Only		Quality Education	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Non-reader	276	18.9%	150	20.4%	126	17.3%	233	21.7%*	43	11.5%	83	17.6%	193	19.4%
Beginning reader	779	55.4%	388	55.7%	391	55.1%	526	51.6%	253	65.4%*	216	53.2%	563	56.3%
Progressing reader	161	11.8%	84	12.0%	77	11.7%	113	11.7%	48	12.1%	52	14.2%	109	10.9%
Proficient reader	184	13.9%	77	11.9%	107	16.0%*	143	15.0%	41	11.1%	62	15.0%	122	13.5%
Totals	1400	100.0%	699	100.0%	701	100.0%	1015	100.0%	385	100.0%	413	100.0%	987	100.0%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Note: 88 students did not complete the ORF subtask and therefore are missing from the proficiency bands.

Table D.2. Proportion of Students Per Proficiency Band, by Province and Gender

Proficiency Band	Province A				Province B			
	Boys		Girls		Boys		Girls	
	n	%	n	%	n	%	n	%
Non-reader	121	21.7%	112	21.6%	29	16.8%	14	6.2%
Beginning reader	278	54.0%	248	49.1%	110	60.3%	143	70.4%
Progressing reader	57	11.5%	56	12.0%	27	13.2%	21	10.9%
Proficient reader	59	12.7%	84	17.3%	18	9.7%	23	12.5%
Totals	515	100.0%	500	100.0%	184	100.0%	201	100.0%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.3. Proportion of Students Per Proficiency Band, by School Type and Gender

Proficiency Band	AO				QE			
	Boys		Girls		Boys		Girls	
	n	%	n	%	n	%	n	%
Non-reader	52	18.1%	31	17.0%	98	21.6%	95	17.4%
Beginning reader	113	52.7%	103	53.9%	275	57.2%	288	55.5%
Progressing reader	31	14.8%	21	13.4%	53	10.6%	56	11.1%
Proficient reader	33	14.4%	29	15.8%	44	10.7%	78	16.1%
Totals	229	100.0%	184	100.0%	470	100.0%	517	100.0%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.4. Proportion of Students Per Proficiency Band, by Province and School Type

Proficiency Band	Province A				Province B			
	AO		QE		AO		QE	
	n	%	n	%	n	%	n	%
Non-reader	73	21.5%	160	21.7%	10	7.7%	33	13.1%
Beginning reader	134	45.8%	392	54.0%*	82	71.9%	171	62.6%
Progressing reader	40	15.4%*	73	10.3%	12	11.2%	36	12.5%
Proficient reader	50	17.3%	93	14.0%	12	9.2%	29	11.9%
Totals	297	100.0%	718	100.0%	116	100.0%	269	100.0%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.5. Proportion of Students Per Proficiency Band by District

Proficiency Band	Province A										Province B	
	District 1		District 2		District 3		District 4		District 5		District 6	
	n	%	n	%	n	%	n	%	n	%	n	%
Non-reader	60	22.4%	10	7.4%	81	35.7%*	65	16.5%	17	59.9%*	43	11.5%

Proficiency Band	Province A										Province B	
	District 1		District 2		District 3		District 4		District 5		District 6	
	n	%	n	%	n	%	n	%	n	%	n	%
Beginning reader	168	57.8%	74	69.0%*	92	46.7%	183	48.4%	9	27.0%	253	65.4%
Progressing reader	30	8.3%	9	9.2%	18	7.5%	54	15.9%	2	5.9%	48	12.1%
Proficient reader	37	11.5%	16	14.4%	16	10.1%	72	19.1%	2	7.2%	41	11.1%
Totals	295	100.0%	109	100.0%	207	100.0%	374	100.0%	30	100.0%	385	100.0%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.6. Proportion of Students Per Proficiency Band, by District and Gender

Proficiency Band	Province A														Province B									
	District 1		District 2		District 3		District 4		District 5		District 6		Boys		Girls									
	Boys	Girls	Boys	Girls	Boys	Girls																		
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Non-reader	36	22.1%	24	22.7%	8	11.1%	2	2.9%	40	40.3%	41	31.5%	30	15.6%	35	17.5%	7	49.0%	10	70.9%	29	16.8%	14	6.2%
Beginning reader	90	60.0%	78	55.6%	40	65.8%	34	72.7%	39	40.5%	53	52.2%	104	55.1%	79	41.1%	5	31.5%	4	22.5%	110	60.3%	143	70.4%
Progressing reader	15	9.0%	15	7.7%	7	11.3%	2	6.7%	9	8.8%	9	6.3%	25	14.0%	29	18.1%	1	5.3%	1	6.6%	27	13.2%	21	10.9%
Proficient reader	12	8.9%	25	14.1%	7	11.8%	9	17.6%	7	10.3%	9	9.9%	31	15.3%	41	23.3%	2	14.3%	0	0.0%	18	9.7%	23	12.5%
Totals	153	100%	142	100%	62	100%	47	100%	95	100%	112	100%	190	100%	184	100%	15	100%	15	100%	184	100%	201	100%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.7. Proportion of Students Per Proficiency Band, by School Type and

Proficiency Band	Access Only				Quality Education									
	N/A		Wave D		Wave A		Wave B		Wave C		Wave D		N/A	
	n	%	n	%	n	n	%	%	n	%	n	%	n	%
Non-reader	77	16.5%	6	61.0%*	77	22.4%*	23	28.9%*	3	2.0%	57	21.1%*	33	13.1%

Proficiency Band	Access Only								Quality Education							
	N/A		Wave D		Wave A		Wave B		Wave C		Wave D		N/A			
	n	%	n	%	n	n	%	%	n	%	n	%	n	%		
Beginning reader	216	54.6%	0	0.0%	177	55.5%	63	49.4%	30	58.6%	122	53.1%	171	62.6%		
Progressing reader	51	14.3%	1	9.5%	29	8.3%	15	9.9%	9	15.5%	20	12.1%	36	12.5%		
Proficient reader	59	14.6%	3	29.5%	42	13.7%	18	11.8%	9	23.9%	24	13.7%	29	11.9%		
Totals	403	100%	10	100%	325	100%	119	100%	51	100%	223	100%	269	100%		

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.8. Proportion of Students Per Proficiency Band, by Number of Moves and New Schools Due to Conflict, Attendance in Past Week, and Province

Proficiency Band	Moves Due to Conflict								New Schools Attended								Attendance in Past Week							
	Province A				Province B				Province A				Province B				Province A				Province B			
	1-5 times		6+ times		1-5 times		6+ times		1-3 schools		4+ schools		1-3 schools		4+ schools		No Days Missed		1+ Days Missed		No Days Missed		1+ Days Missed	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Non-reader	103	17.8%	130	26.0%*	30	12.7%	13	8.6%	65	18.1%	168	23.4%	19	11.7%	24	11.2%	164	19.9%	69	27.9%*	30	9.1%	13	22.8%*
Beginning reader	278	51.8%	248	51.4%	180	65.4%	73	65.4%	167	51.3%	359	51.8%	121	63.4%	132	67.1%	395	49.7%	131	58.4%*	208	65.9%	45	62.9%
Progressing reader	62	13.4%	51	9.9%	30	9.9%	18	16.8%	44	15.2%*	68	10.1%	26	12.9%	22	11.3%	92	12.9%*	21	7.5%	43	12.5%	5	10.1%
Proficient reader	85	17.0%	58	12.8%	30	11.9%	11	9.3%	49	15.4%	94	14.8%	21	11.9%	20	10.4%	130	17.4%*	13	6.2%	37	12.6%*	4	4.1%
Totals	528	100%	487	100%	270	100%	115	100%	325	100%	689	100%	187	100%	198	100%	781	100%	234	100%	318	100%	67	100%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.9. Proportion of Students Per Proficiency Band, by District and Number of Moves Due to Conflict

Proficiency Bands	Province A												Province B											
	District 1		District 2		District 3		District 4		District 5		District 6													
	1-5 times		6+ times		1-5 times		6+ times		1-5 times		6+ times		1-5 times		6+ times									
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%								
Non-reader	29	19.6%	31	26.1%	7	10.3%	3	4.6%	27	26.5%	54	41.0%	28	12.3%	37	21.8%	12	58.9%	5	67.0%	30	12.7%	13	8.6%
Beginning reader	86	56.5%	82	59.5%	37	61.3%	37	76.5%	38	45.6%	54	47.4%	110	51.9%	73	44.0%	7	26.2%	2	33.0%	180	65.4%	73	65.4%
Progressing reader	16	8.0%	14	8.8%	4	6.9%	5	11.4%	10	11.3%	8	5.3%	30	18.3%	24	12.9%	2	6.7%	0	0.0%	30	9.9%	18	16.8%
Proficient reader	24	16.0%*	13	5.7%	13	21.5%*	3	7.6%	8	16.6%*	8	6.3%	38	17.5%	34	21.3%	2	8.2%	0	0.0%	30	11.9%	11	9.3%
Totals	155	100%	140	100%	61	100%	48	100%	83	100%	124	100%	206	100%	168	100%	23	100%	7	100%	270	100%	115	100%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.10. Proportion of Students Per Proficiency Band, by Feelings of Safety

Proficiency Band	Feeling Safe at School				Feeling Safe to and from School				Feeling Safe at Home															
	Province A		Province B		Province A		Province B		Province A		Province B													
	Unsafe		Safe		Unsafe		Safe		Unsafe		Safe													
	n	%	n	%	n	%	n	%	n	%	n	%												
Non-reader	10	30.4%	222	21.5%	4	29.8%*	36	11.1%	19	17.1%	214	22.3%	4	12.8%	36	11.9%	3	17.4%	230	21.8%	1	19.5%	39	11.8%
Beginning reader	22	59.6%	502	51.2%	7	41.0%	225	66.5%*	55	44.9%	469	52.3%	25	61.0%	207	65.8%	9	44.7%	515	51.5%	5	80.5%	227	65.0%
Progressing reader	2	3.3%	111	12.0%	2	11.1%	39	11.3%	18	17.4%	95	11.1%	8	13.2%	33	11.0%	5	33.3%*	108	11.4%	0	0.0%	41	11.5%
Proficient reader	2	6.8%	141	15.3%	3	18.1%	36	11.1%	19	20.5%	124	14.3%	5	13.1%	34	11.3%	1	4.6%	142	15.2%	0	0.0%	39	11.7%
Totals	36	100%	976	100%	16	100%	336	100%	111	100%	902	100%	42	100%	310	100%	18	100%	995	100%	6	100%	346	100%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Table D.11. Proportion of Students Per Proficiency Band, by Province and Time to Travel to School

Proficiency Band	Province A						Province B					
	Time to get to school						Time to get to school					
	Between 1-15 min		Between 16-30 min		31 or more min		Between 1-15 min		Between 16-30 min		31 or more min	
	n	%	n	%	n	%	n	%	n	%	n	%
Non-reader	183	20.9%	44	24.4%	6	37.1%	34	12.5%	6	11.4%	0	0.0%
Beginning reader	412	50.3%	102	55.6%	10	62.9%	189	65.8%	37	61.1%	6	79.8%
Progressing reader	97	12.3%	16	10.5%	0	0.0%	31	10.9%	8	11.5%	2	20.2%*
Proficient reader	126	16.5%*	17	9.5%	0	0.0%	29	10.8%	10	16.0%	0	0.0%
Totals	818	100.0%	179	100.0%	16	100.0%	283	100.0%	61	100.0%	8	100.0%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Table D.12. Proportion of Students Per Proficiency Band, by Feelings of Hunger and Tiredness

Proficiency Band	Hunger								Tiredness							
	Province A				Province B				Province A				Province B			
	Never Hungry		Experience Hunger		Never Hungry		Experience Hunger		Never Tired		Experience Tiredness		Never Tired		Experience Tiredness	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Non-reader	109	18.3%	124	26.0%*	18	11.9%	22	12.0%	142	19.5%	91	26.3%*	17	7.9%	23	17.0%*
Beginning reader	304	54.4%*	220	47.9%	106	67.6%	126	63.6%	383	54.5%*	141	45.2%	138	70.6%*	94	58.9%
Progressing reader	69	13.8%*	43	9.4%	13	9.3%	28	12.7%	78	12.3%	35	10.8%	18	9.8%	23	13.1%
Proficient reader	72	13.6%	71	16.8%	12	11.1%	27	11.7%	90	13.7%	53	17.7%	21	11.8%	18	11.0%
Totals	554	100.0%	458	100.0%	149	100.0%	203	100.0%	693	100.0%	320	100.0%	194	100.0%	158	100.0%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table D.13. Proportion of Students Per Proficiency Band between Provinces

Proficiency Band	Province A		Province B		Totals	
	n	%	n	%	n	%
Non-reader	233	83.3%	43	16.7%	276	100.0%

Proficiency Band	Province A		Province B		Totals	
	n	%	n	%	n	%
Beginning reader	526	67.5%	253	32.5%	779	100.0%
Progressing reader	113	71.9%	48	28.1%	161	100.0%
Proficient reader	143	78.0%	41	22.0%	184	100.0%

Table D.14. Proportion of Students Per Proficiency Band between School Types

Proficiency Band	Access Only		Quality Education		Totals	
	n	%	n	%	n	%
Non-reader	83	27.6%	193	72.4%	276	100.0%
Beginning reader	216	28.4%	563	71.6%	779	100.0%
Progressing reader	52	35.4%	109	64.6%	161	100.0%
Proficient reader	62	31.9%	122	68.1%	184	100.0%

Table D.15. Proportion of Students Per Proficiency Band between Districts

Proficiency Band	District												Totals	
	District 1		District 2		District 3		District 4		District 5		District 6			
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Non-reader	60	17.4%	10	3.0%	81	26.8%	65	29.7%	17	6.4%	43	16.7%	276	100.0%
Beginning reader	168	15.3%	74	9.6%	92	12.0%	183	29.6%	9	1.0%	253	32.5%	779	100.0%
Progressing reader	30	10.3%	9	6.0%	18	9.0%	54	45.7%	2	1.0%	48	28.1%	161	100.0%
Proficient reader	37	12.1%	16	8.0%	16	10.3%	72	46.6%	2	1.0%	41	22.0%	184	100.0%

Table D.16. Proportion of Students Per Proficiency Band, comparing 2017 and 2019 Results by Province and Gender

Proficiency Band	Total				Province								Gender							
	2019		2017		Province A				Province B				Male		Female					
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%				
Non-reader	276	18.9%*	138	14.2%	233	21.7%*	71	15.3%	43	11.5%	67	13.0%	150	20.4%	86	18.1%	126	17.3%*	52	10.4%

Proficiency Band	Province										Gender									
	Total				Province A				Province B				Male		Female					
	2019		2017		2019		2017		2019		2017		2019		2017					
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%				
Beginning reader	779	55.4%*	340	34.5%	526	51.6%*	138	29.7%	253	65.4%*	202	39.3%	388	55.7%*	189	39.2%	391	55.1%*	151	30.1%
Progressing reader	161	11.8%	441	45.3%*	113	11.7%	228	49.1%*	48	12.1%	213	41.4%*	84	12.0%	181	38.1%*	77	11.7%	260	52.1%*
Proficient reader	184	13.9%*	59	6.0%	143	15.0%*	27	5.8%	41	11.1%*	32	6.2%	77	11.9%*	22	4.6%	107	16.0%*	37	7.4%
Total	1400	100.0%	978	100.0%	1015	100.0%	464	100.0%	385	100.0%	514	100.0%	699	100.0%	478	100.0%	701	100.0%	500	100.0%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Annex E: Mean Fluency Rates and Accuracy

Table E.1. EGRA Fluency Rates and Accuracy Scores, by Province, Gender and School Type

Subtask	Province						Gender				School Type			
	Total		Province A		Province B		Male		Female		AO		QE	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Fluency Rates														
Letter sound identification fluency (CLSPM)	1479	27.9	1064	26.4	415	31.7*	744	26.0	735	29.8*	430	27.7	1049	27.9
Nonword reading fluency (CNWPM)	1479	3.9	1064	3.7	415	4.2	744	3.9	735	3.9	430	3.8	1049	3.9
Oral reading fluency (CWPM)	1479	14.9	1064	15.2	415	14.2	744	13.8	735	16.1*	430	16.2	1049	14.4
Accuracy Scores														
Letter Sounds % Correct Out of 100 Total Items	1479	51.9%	1064	49.2%	415	58.8%*	744	49.0%	735	54.9%*	430	51.4%	1049	52.1%

Subtask	Province						Gender				School Type			
	Total		Province A		Province B		Male		Female		AO		QE	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Nonword % Correct Out of 50 Total Items	1479	14.9%	1064	14.1%	415	16.8%*	744	15.2%	735	14.5%	430	14.5%	1049	15.0%
ORF % Correct Out of 82 Total Items	1479	31.6%	1064	32.5%	415	29.2%	744	29.8%	735	33.5%*	430	34.2%*	1049	30.5%
Listening Comp % Correct Out of 6 Total Items	1479	79.6%	1064	80.2%	415	77.9%	744	81.9%*	735	77.2%	430	81.9%*	1049	78.6%
Reading Comp % Correct Out of 5 Total Items	1479	29.4%	1064	29.9%	415	28.0%	744	27.0%	735	31.8%	430	31.1%	1049	28.7%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Table E.2. EGMA Fluency Rates and Accuracy Scores, by Province, Gender and School Type

Subtask	Province						Gender				School Type			
	Total		Province A		Province B		Male		Female		AO		QE	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Fluency Rates														
Number recognition fluency (CNPM)	1479	24.2	1064	24.7*	415	22.9	744	25.9*	735	22.4	430	25.8*	1049	23.5
Addition 1 fluency (CAPPm)	1479	7.9	1064	8.0	415	7.9	744	8.5*	735	7.4	430	8.4*	1049	7.8
Subtraction 1 fluency (CSPPM)	1479	5.4	1064	5.3	415	5.6	744	6.0*	735	4.7	430	5.8*	1049	5.2
Accuracy Scores														
Number recognition: % correct out of 20 total items	1479	82.0%	1064	82.2%	415	81.4%	744	84.2%*	735	79.7%	430	83.9%*	1049	81.2%
Number discrimination:	1479	67.8%	1064	65.8%	415	72.8%*	744	72.5%*	735	62.9%	430	70.6%*	1049	66.6%

Subtask	Province						Gender				School Type			
	Total		Province A		Province B		Male		Female		AO		QE	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
% correct out of 10 total items														
Missing number: % correct out of 10 total items	1479	45.4%	1064	44.6%	415	47.5%*	744	48.5%*	735	42.3%	430	48.5%*	1049	44.1%
Addition 1: % correct out of 20 total items	1479	70.5%	1064	70.3%	415	71.1%	744	73.4%*	735	67.5%	430	72.7%*	1049	69.6%
Addition 2: % correct out of 5 total items	1479	47.5%	1064	47.3%	415	47.8%	744	51.6%*	735	43.3%	430	49.0%	1049	46.9%
Subtraction 1: % correct out of 20 total items	1479	51.2%	1064	51.0%	415	51.8%	744	55.5%*	735	46.8%	430	53.5%*	1049	50.3%
Subtraction 2: % correct out of 5 total items	1479	25.6%	1064	25.5%	415	25.9%	744	30.0%*	735	21.2%	430	28.8%*	1049	24.3%
Word Problems: % correct out of 3 total items	1479	51.5%	1064	50.2%	415	54.9%*	744	53.9%*	735	49.1%	430	56.1%*	1049	49.7%

Note: An asterisk (*) indicates mean differences that are statistically significant at the $p < 0.05$ level.

Table E.3. EGRA Fluency Rates and Accuracy Scores, by Province and Gender

Subtask	Total				Province A				Province B			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Fluency Rates												
Letter sound identification fluency (clspm)	744	26.0	735	29.8	539	24.5	525	28.2*	205	29.7	210	33.9*
Nonword reading fluency (cnwpm)	744	3.9	735	3.9	539	3.7	525	3.7	205	4.2	210	4.3
Oral reading fluency (cwpm)	744	13.8	735	16.1	539	13.8	525	16.6*	205	13.6	210	14.8
Accuracy Scores												

Subtask	Total				Province A				Province B			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Letter Sounds % Correct Out of 100 Total Items	744	49.0%	735	54.9%	539	46.5%	525	52.0%*	205	55.3%	210	62.4%*
Nonword % Correct Out of 50 Total Items	744	15.2%	735	14.5%	539	14.6%	525	13.7%	205	16.7%	210	16.9%
ORF % Correct Out of 82 Total Items	744	29.8%	735	33.5%	539	30.7%	525	34.4%	205	27.4%	210	31.1%
Listening Comp % Correct Out of 6 Total Items	744	81.9%	735	77.2%	539	82.6%*	525	77.9%	205	80.1%	210	75.6%
Reading Comp % Correct Out of 5 Total Items	744	27.0%	735	31.8%	539	27.3%	525	32.5%*	205	26.2%	210	29.9%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Table E.4. EGMA Fluency Rates and Accuracy Scores, by Province and Gender

Subtask	Total				Province A				Province B			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Fluency Rates												
Number recognition fluency (cnrpm)	744	25.9	735	22.4	539	26.6*	525	22.8	205	24.3*	210	21.5
Addition 1 fluency (caddpm)	744	8.5	735	7.4	539	8.7*	525	7.3	205	8.2	210	7.6
Subtraction 1 fluency (csubpm)	744	6.0	735	4.7	539	5.9*	525	4.6	205	6.1*	210	5.0
Accuracy Scores												
Number recognition: % correct out of 20 total items	744	84.2%	735	79.7%	539	84.6%*	525	79.9%	205	83.4%*	210	79.3%

Subtask	Total				Province A				Province B			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Number discrimination: % correct out of 10 total items	744	72.5%	735	62.9%	539	70.9%*	525	60.7%	205	76.6%*	210	68.8%
Missing number: % correct out of 10 total items	744	48.5%	735	42.3%	539	47.3%*	525	41.7%	205	51.4%*	210	43.6%
Addition 1: % correct out of 20 total items	744	73.4%	735	67.5%	539	73.5%*	525	67.0%	205	73.2%	210	69.0%
Addition 2: % correct out of 5 total items	744	51.6%	735	43.3%	539	52.2%*	525	42.4%	205	50.0%	210	45.5%
Subtraction 1: % correct out of 20 total items	744	55.5%	735	46.8%	539	55.8%*	525	46.0%	205	54.7%*	210	48.9%
Subtraction 2: % correct out of 5 total items	744	30.0%	735	21.2%	539	29.9%*	525	21.1%	205	30.3%*	210	21.2%
Word Problems: % correct out of 3 total items	744	53.9%	735	49.1%	539	52.8%	525	47.6%*	205	56.7%	210	53.1%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Table E.5. EGRA Fluency Rates and Accuracy Scores, by School Type and Gender

Subtask	Total				AO				QE			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Fluency Rates												
Letter sound identification fluency (clspm)	744	26.0	735	29.8	242	27.1	188	28.6	502	25.5	547	30.2*
Nonword reading fluency (cnwpm)	744	3.9	735	3.9	242	3.9	188	3.5	502	3.8	547	4.0

Subtask	Total				AO				QE			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Oral reading fluency (cwpm)	744	13.8	735	16.1	242	16.2	188	16.2	502	12.5	547	16.1*
Accuracy Scores												
Letter Sounds % Correct Out of 100 Total Items	744	49.0%	735	54.9%	242	50.0%	188	53.4%	502	48.5%	547	55.4%*
Nonword % Correct Out of 50 Total Items	744	15.2%	735	14.5%	242	15.1%	188	13.6%	502	15.3%	547	14.8%
ORF % Correct Out of 82 Total Items	744	29.8%	735	33.5%	242	34.0%	188	34.5%	502	27.6%	547	33.1%*
Listening Comp % Correct Out of 6 Total Items	744	81.9%	735	77.2%	242	82.5%	188	81.1%	502	81.5%*	547	75.9%
Reading Comp % Correct Out of 5 Total Items	744	27.0%	735	31.8%	242	29.4%	188	33.4%	502	25.9%	547	31.2%*

Note: As asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Table E.6. EGMA Fluency Rates and Accuracy Scores, by Province and Gender

Subtask	Total				AO				QE			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Fluency Rates												
Number recognition fluency (cnrpm)	744	25.9	735	22.4	242	28.1*	188	22.7	502	24.8*	547	22.4
Addition 1 fluency (caddpm)	744	8.5	735	7.4	242	9.0*	188	7.6	502	8.3*	547	7.3
Subtraction 1 fluency (csubpm)	744	6.0	735	4.7	242	6.5*	188	4.8	502	5.7*	547	4.7
Accuracy Scores												

Subtask	Total				AO				QE			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Number recognition: % correct out of 20 total items	744	84.2%	735	79.7%	242	85.4%*	188	81.7%	502	83.6%*	547	79.0%
Number discrimination: % correct out of 10 total items	744	72.5%	735	62.9%	242	74.2%*	188	65.7%	502	71.7%*	547	62.0%
Missing number: % correct out of 10 total items	744	48.5%	735	42.3%	242	51.5%*	188	44.4%	502	47.0%*	547	41.5%
Addition 1: % correct out of 20 total items	744	73.4%	735	67.5%	242	76.2%*	188	68.0%	502	72.0%*	547	67.4%
Addition 2: % correct out of 5 total items	744	51.6%	735	43.3%	242	53.3%*	188	43.1%	502	50.7%*	547	43.3%
Subtraction 1: % correct out of 20 total items	744	55.5%	735	46.8%	242	58.1%*	188	47.3%	502	54.2%*	547	46.7%
Subtraction 2: % correct out of 5 total items	744	30.0%	735	21.2%	242	32.2%*	188	24.2%	502	29.0%*	547	20.1%
Word Problems: % correct out of 3 total items	744	53.9%	735	49.1%	242	57.1%	188	54.8%	502	52.3%*	547	47.3%

Note: An asterisk (*) indicates mean differences that are statistically significant at the $p < 0.05$ level.

Table E.7. EGRA Fluency Rates and Accuracy Scores, by District

	Province A										Province B	
	District 1		District 2		District 3		District 4		District 5		District 6	
	n	Mean										
Fluency Rates												
Letter sound identification fluency (CLSPM)	301	25.2	112	32.6	240	21.7	381	28.0	30	21.1	415	31.8
Non-word reading fluency (CNWPM)	301	3.4	112	5.9	240	1.4	381	4.6	30	2.1	415	4.2
Oral reading fluency (CWPM)	301	12.9	112	14.4	240	12.1	381	18.1	30	7.6	415	14.2
Average Percentage Correct of Total												
Letter Sounds % Correct Out of 100 Total Items	301	48.1%	112	61.7%	240	40.9%	381	51.2%	30	42.1%	415	58.8%
Nonword % Correct Out of 50 Total Items	301	13.2%	112	23.3%	240	5.6%	381	16.9%	30	8.5%	415	16.8%
ORF % Correct Out of 82 Total Items	301	29.0%	112	32.8%	240	23.8%	381	38.5%	30	17.4%	415	29.2%
Listening Comp % Correct Out of 6 Total Items	301	80.3%	112	93.9%	240	69.7%	381	83.2%	30	61.3%	415	77.9%
Reading Comp % Correct Out of 5 Total Items	301	28.9%	112	34.8%	240	19.8%	381	34.8%	30	17.0%	415	28.0%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Table E.8. EGMA Fluency Rates and Accuracy Scores, by District

	Province A										Province B	
	District 1		District 2		District 3		District 4		District 5		District 6	
	n	Mean										
Fluency Rates												
Number recognition fluency (CNPM)	301	23.9	112	23.8	240	21.7	381	27.0	30	18.6	415	22.9
Addition 1 fluency (CAPPm)	301	8.2	112	7.5	240	6.9	381	8.6	30	6.8	415	7.9

	Province A										Province B	
	District 1		District 2		District 3		District 4		District 5		District 6	
	n	Mean										
Subtraction 1 fluency (CSPPM)	301	5.5	112	5.2	240	4.6	381	5.6	30	4.1	415	5.6
Accuracy Scores												
Number recognition: % correct out of 20 total items	301	84.2%	112	86.3%	240	76.8%	381	83.6%	30	73.4%	415	81.4%
Number discrimination: % correct out of 10 total items	301	74.3%	112	70.7%	240	54.1%	381	66.9%	30	59.3%	415	72.8%
Missing number: % correct out of 10 total items	301	46.0%	112	41.6%	240	37.2%	381	48.6%	30	36.5%	415	47.6%
Addition 1: % correct out of 20 total items	301	74.0%	112	71.1%	240	61.9%	381	72.8%	30	63.5%	415	71.1%
Addition 2: % correct out of 5 total items	301	55.4%	112	52.6%	240	37.1%	381	47.8%	30	42.9%	415	47.8%
Subtraction 1: % correct out of 20 total items	301	54.2%	112	51.6%	240	44.7%	381	53.1%	30	39.4%	415	51.8%
Subtraction 2: % correct out of 5 total items	301	30.6%	112	28.2%	240	17.8%	381	26.7%	30	20.8%	415	25.9%
Word Problems: % correct out of 3 total items	301	47.8%	112	55.1%	240	45.2%	381	52.7%	30	46.5%	415	54.9%

Note: An asterisk (*) indicates mean differences that are statistically significant at the $p < 0.05$ level.

Table E.9. EGRA Fluency Rates and Accuracy Scores, by District and Gender

Subtask	Province A																				Province B			
	District 1				District 2				District 3				District 4				District 5				District 6			
	Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Fluency Rates																								
Letter sound identification fluency (clspm)	155	22.4	146	28.0*	64	30.0	48	35.7	113	19.8	127	23.3	192	26.2	189	29.9	15	23.2	15	18.9	205	29.7	210	33.9*
Nonword reading fluency (cnwpm)	155	3.0	146	3.7	64	5.5	48	6.4	113	1.8	127	1.1	192	4.5	189	4.7	15	3.3	15	0.9	205	4.2	210	4.3
Oral reading fluency (cwpm)	155	11.3	146	14.5	64	12.1	48	17.1	113	10.9	127	13.1	192	16.5	189	19.9*	15	11.5	15	3.6	205	13.6	210	14.8
Accuracy Scores																								
Letter Sounds % Correct Out of 100 Total Items	155	43.6%	146	52.7%*	64	56.1%	48	68.7%*	113	37.7%	127	43.7%	192	49.2%	189	53.2%	15	46.5%	15	37.7%	205	55.3%	210	62.4%*
Nonword % Correct Out of 50 Total Items	155	12.0%	146	14.3%	64	21.9%	48	25.0%	113	6.9%	127	4.4%	192	17.4%	189	16.4%	15	13.3%	15	3.6%	205	16.7%	210	16.8%
ORF % Correct Out of 82 Total Items	155	26.1%	146	31.9%	64	27.5%	48	39.0%*	113	23.7%	127	24.0%	192	36.2%	189	40.9%	15	26.0%	15	8.7%	205	27.4%	210	31.1%
Listening Comp % Correct Out of 6 Total Items	155	80.1%	146	80.4%	64	95.5%	48	92.0%	113	71.2%	127	68.3%	192	85.5%*	189	80.9%	15	81.3%*	15	41.0%	205	80.1%	210	75.6%
Reading Comp % Correct Out of 5 Total Items	155	25.9%	146	31.9%	64	27.9%	48	43.3%*	113	18.4%	127	20.9%	192	31.7%	189	38.0%*	15	28.1%*	15	5.8%	205	26.2%	210	29.9%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table E.10. EGMA Fluency Rates and Accuracy Scores, by District and Gender

Subtask	Province A																				Province B			
	District 1				District 2				District 3				District 4				District 5				District 6			
	Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Fluency Rates																								
Number recognition fluency (cnrpm)	155	24.7	146	23.1	64	23.5	48	24.2	113	23.5	127	20.1	192	29.6*	189	24.3	15	22.7*	15	14.4	205	24.3*	210	21.4
Addition 1 fluency (caddpm)	155	8.7*	146	7.6	64	7.1	48	8.1	113	7.7*	127	6.2	192	9.5*	189	7.6	15	8.2*	15	5.4	205	8.2	210	7.6
Subtraction 1 fluency (csubpm)	155	5.8	146	5.2	64	5.4	48	5.0	113	5.1*	127	4.1	192	6.5*	189	4.6	15	4.4	15	3.8	205	6.1*	210	5.0
Accuracy Scores																								
Number recognition: % correct out of 20 total items	155	84.3%	146	84.1%	64	86.4%	48	86.2%	113	80.1%*	127	73.8%	192	86.0%*	189	81.0%	15	87.5%*	15	59.2%	205	83.4%*	210	79.3%
Number discrimination: % correct out of 10 total items	155	77.9%*	146	70.7%	64	69.4%	48	72.2%	113	59.1%*	127	49.6%	192	73.2%*	189	60.3%	15	74.1%*	15	44.3%	205	76.6%*	210	68.8%
Missing number: % correct out of 10 total items	155	47.6%	146	44.4%	64	42.0%	48	41.1%	113	38.5%	127	36.2%	192	52.4%*	189	44.6%	15	47.8%*	15	25.1%	205	51.4%*	210	43.6%
Addition 1: % correct out of 20 total items	155	75.8%	146	72.1%	64	67.7%	48	75.2%	113	66.4%*	127	57.9%	192	77.1%*	189	68.3%	15	73.0%	15	53.9%	205	73.2%	210	69.0%
Addition 2: % correct out of 5 total items	155	58.8%	146	52.0%	64	53.1%	48	52.0%	113	42.4%*	127	32.4%	192	53.2%*	189	42.0%	15	57.2%*	15	28.5%	205	50.0%	210	45.5%
Subtraction 1: % correct out of 20 total items	155	56.2%	146	52.3%	64	53.0%	48	49.9%	113	49.1%*	127	40.8%	192	60.2%*	189	45.6%	15	41.0%	15	37.7%	205	54.7%*	210	48.9%
Subtraction 2: % correct out of 5 total items	155	32.3%	146	29.0%	64	29.5%	48	26.6%	113	21.7%*	127	14.3%	192	32.4%*	189	20.7%	15	32.9%*	15	8.5%	205	30.3%*	210	21.2%
Word Problems: % correct out of 3 total items	155	51.5%	146	44.2%	64	54.6%	48	55.8%	113	46.6%	127	44.0%	192	55.6%*	189	49.6%	15	50.9%	15	42.1%	205	56.7%	210	53.1%

Subtask	Province A												Province B				
	District 1		District 2		District 3		District 4		District 5		District 6						
	Boys	Girls	Boys	Girls													
n	%	n	%	N	%	n	%	n	%	n	%	n	%	n	%	n	%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Table E.11. EGMA Fluency Rates and Accuracy Scores, Comparing 2017 and 2019 by Province and Gender

Subtask	Total				Province								Gender							
					Province A				Province B				Male				Female			
	2019		2017		2019		2017		2019		2017		2019		2017		2019		2017	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Fluency Rates																				
Number recognition fluency (CNPM)	1479	24.2	978	33.8	1064	24.7	464	33.7*	415	22.9	514	34.0*	744	25.9	478	34.1*	735	22.4	500	33.6*
Addition 1 fluency (CAPPm)	1479	7.9	978	11.4	1064	8.0	464	10.9*	415	7.9	514	11.9*	744	8.5	478	11.3*	735	7.4	500	11.5*
Subtraction 1 fluency (CSPPM)	1479	5.4	978	9.1	1064	5.3	464	8.9*	415	5.6	514	9.3*	744	6.0	478	9.4*	735	4.7	500	8.8*
Accuracy Scores																				
Number recognition: % correct out of 20 total items	1479	82.0%	978	87.7%	1064	82.2%	464	89.0%*	415	81.4%	514	86.5%*	744	84.2%	478	87.3%*	735	79.7%	500	88.1%*
Number discrimination: % correct out of 10 total items	1479	67.8%	978	76.8%	1064	65.8%	464	75.5%*	415	72.8%	514	78.0%*	744	72.5%	478	78.2%*	735	62.9%	500	75.4%*
Missing number: % correct out of 10 total items	1479	45.4%	978	54.1%	1064	44.6%	464	51.6%*	415	47.5%	514	56.6%*	744	48.5%	478	54.9%*	735	42.3%	500	53.3%*

Subtask	Province												Gender							
	Total				Province A				Province B				Male				Female			
	2019		2017		2019		2017		2019		2017		2019		2017		2019		2017	
	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean
Addition 1: % correct out of 20 total items	1479	70.5%	978	56.3%	1064	70.3%*	464	54.2%	415	71.1%*	514	58.4%	744	73.4%*	478	56.0%	735	67.5%*	500	56.6%
Addition 2: % correct out of 5 total items	1479	47.5%	978	55.7%	1064	47.3%	464	51.9%*	415	47.8%	514	59.6%*	744	51.6%	478	55.7%*	735	43.3%	500	55.8%*
Subtraction 1: % correct out of 20 total items	1479	51.2%	978	45.4%	1064	51.0%*	464	44.3%	415	51.8%*	514	46.4%	744	55.5%*	478	46.8%	735	46.8%*	500	44.0%
Subtraction 2: % correct out of 5 total items	1479	25.6%	978	36.9%	1064	25.5%	464	32.2%*	415	25.9%	514	41.5%*	744	30.0%	478	37.3%*	735	21.2%	500	36.5%*
Word Problems: % correct out of 3 total items					1064	50.2%*	464	40.7%	415	54.9%	514	45.8%	744	53.9%	478	44.3%	735	49.1%	500	42.3%

Note: An asterisk (*) indicates mean differences that are statistically significant at the p<0.05 level.

Annex F: Zero Scores

Table F.1. Proportion of EGRA and EGMA Zero Scores by Subgroups

Subtask	Gender				Province				School Type			
	Male		Female		Province A		Province B		AO		QE	
	n	%	n	%	n	%	n	%	n	%	n	%
EGRA												
Letter Sound Identification	89	12.7%	75	10.4%	137	13.1%*	27	7.7%	68	14.6%*	96	10.3%
Nonword Reading	363	47.0%	354	49.2%	550	52.1%*	167	37.8%	229	53.6%*	488	45.8%
Oral Reading Fluency	150	20.4%	126	17.3%	233	21.7%*	43	11.5%	83	17.6%	193	19.4%
Reading Comprehension	321	43.1%	280	38.7%	438	41.0%	163	40.9%	163	37.4%	438	42.4%
Listening Comprehension	12	1.5%	22	3.2%*	25	2.2%	9	2.7%	7	2.0%	27	2.5%
EGMA												
Number Recognition	2	0.1%	1	0.2%	3	0.2%	0	0.0%	1	0.0%	2	0.2%
Number Discrimination	7	0.6%	8	1.1%	13	1.0%	2	0.4%	5	0.8%	10	0.8%
Missing Number	21	2.1%	17	1.9%	30	2.3%	8	1.4%	10	1.0%	28	2.4%
Addition 1	34	4.0%	39	5.6%	56	5.3%	17	3.4%	19	3.3%	54	5.4%
Addition 2	90	11.7%	109	15.8%	145	13.5%	54	14.2%	48	10.1%	151	15.2%*
Subtraction 1	55	6.7%	78	11.2%	100	9.0%	33	8.9%	33	7.1%	100	9.7%
Subtraction 2	241	30.5%	333	46.2%	410	37.2%	164	41.0%	148	32.6%	426	40.6%*
Word Problems	106	13.2%	120	16.3%	179	16.3%	47	10.9%	53	10.5%	173	16.5%*

Note: An asterisk (*) indicates mean differences that are statistically significant at the $p < 0.05$ level.

Table F.2. Proportion of EGRA and EGMA Zero Scores by Province and Gender

Subtask	Total				Province A				Province B			
	Boys		Girls		Boys		Girls		Boys		Girls	
	n	%	n	%	n	%	n	%	n	%	n	%
EGRA												
Letter Sound Identification	89	12.7%	75	10.4%	72	13.5%	65	12.6%	17	10.6%	10	4.7%
Nonword Reading	363	47.0%	354	49.2%	280	50.6%	270	53.6%	83	37.9%	84	37.8%
Oral Reading Fluency	150	20.4%	126	17.3%	121	21.7%	112	21.6%	29	16.8%	14	6.2%
Reading Comprehension	321	43.1%	280	38.7%	236	43.1%	202	38.8%	85	43.2%	78	38.5%
Listening Comprehension	12	1.5%	22	3.2%	10	1.6%	15	2.7%	2	1.2%	7	4.3%
EGMA												
Number Recognition	2	0.1%	1	0.2%	2	0.1%	1	0.3%	0	0.0%	0	0.0%
Number Discrimination	7	0.6%	8	1.0%	5	0.6%	8	1.5%	2	0.7%	0	0.0%
Missing Number	21	2.1%	17	1.9%	17	2.4%	13	2.1%	4	1.4%	4	1.5%
Addition 1	34	4.0%	39	5.5%	25	4.2%	31	6.4%	9	3.5%	8	3.4%
Addition 2	90	11.6%	109	15.8%	65	10.9%	80	16.2%	25	13.6%	29	14.8%
Subtraction 1	55	6.7%	78	11.2%	40	6.2%	60	11.7%	15	8.0%	18	9.9%
Subtraction 2	241	30.5%	333	46.2%	180	30.1%	230	44.5%	61	31.7%	103	50.7%
Word Problems	106	13.2%	120	16.3%	84	14.5%	95	18.1%	22	9.9%	25	11.8%

Table F.3. Proportion of EGRA and EGMA Zero Scores, by School Type and Gender

Subtask	AO				QE			
	Boys		Girls		Boys		Girls	
	n	%	n	%	n	%	n	%
EGRA								
Letter Sound Identification	40	14.6%	28	14.6%	49	11.7%	47	9.0%
Nonword Reading	125	49.2%	104	59.7%	238	46.0%	250	45.7%
Oral Reading Fluency	52	18.1%	31	17.0%	98	21.6%	95	17.4%
Reading Comprehension	99	39.2%	64	35.0%	222	45.0%	216	40.0%
Listening Comprehension	3	1.4%	4	2.7%	9	1.6%	18	3.3%
EGMA								
Number Recognition	1	0.1%	0	0.0%	1	0.1%	1	0.3%
Number Discrimination	3	0.7%	2	0.9%	4	0.6%	6	1.1%
Missing Number	8	1.5%	2	0.3%	13	2.5%	15	2.4%
Addition 1	10	2.4%	9	4.6%	24	4.8%	30	5.9%
Addition 2	26	9.0%	22	11.5%	64	12.9%	87	17.2%
Subtraction 1	19	6.4%	14	8.1%	36	6.9%	64	12.3%
Subtraction 2	72	28.2%	76	38.7%	169	31.7%	257	48.7%
Word Problems	31	10.1%	22	11.0%	75	14.8%	98	18.1%

Table F.4. Proportion of EGRA and EGMA Zero Scores by District

Subtasks	Province A										Province B	
	District 1		District 2		District 3		District 4		District 5		District 6	
	n	%	n	%	n	%	n	%	n	%	n	%
EGRA												
Letter Sound Identification	40	12.0%	5	3.7%	43	17.4%	46	13.6%	3	13.2%	27	7.7%
Nonword Reading	144	52.5%	24	19.8%	176	72.7%	182	47.7%	24	82.6%	167	37.8%
Oral Reading Fluency	60	22.4%	10	7.4%	81	35.7%	65	16.5%	17	59.9%	43	11.5%
Reading Comprehension	111	40.9%	35	26.1%	146	58.2%	127	34.7%	19	66.4%	163	40.9%
Listening Comprehension	5	2.3%	1	0.7%	13	4.3%	4	1.1%	2	8.6%	9	2.7%
EGMA												
Number Recognition	2	0.2%	0	0.0%	0	0.0%	0	0.0%	1	5.3%	0	0.0%
Number Discrimination	2	0.2%	0	0.0%	9	3.5%	1	0.2%	1	5.3%	2	0.4%
Missing Number	9	1.9%	6	4.2%	10	3.3%	4	1.3%	1	5.3%	8	1.4%
Addition 1	10	1.8%	2	1.3%	22	8.6%	19	6.0%	3	8.1%	17	3.4%
Addition 2	26	6.9%	12	8.8%	58	21.9%	45	13.7%	4	10.9%	54	14.2%
Subtraction 1	19	3.9%	7	6.2%	32	12.6%	35	9.5%	7	18.6%	33	8.9%
Subtraction 2	99	29.5%	47	38.0%	111	44.6%	137	36.0%	16	52.5%	164	41.0%
Word Problems	52	17.3%	12	12.0%	56	24.1%	50	12.4%	9	27.4%	47	10.9%

Table F.5. Proportion of EGRA and EGMA Zero Scores by District and Gender

Subtask	Province A												Province B											
	District 1				District 2				District 3				District 4				District 5				District 6			
	Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Letter Sound Identification	25	12.0%	15	12.1%	3	4.3%	2	2.9%	22	19.4%	21	15.6%	21	14.3%	25	12.8%	1	5.3%	2	21.3%	17	10.6%	10	4.7%
Nonword Reading	83	55.5%	61	49.5%	14	20.7%	10	18.8%	85	72.0%	91	73.3%	87	45.1%	95	50.3%	11	74.8%	13	90.6%	83	37.9%	84	37.8%
Oral Reading Fluency	36	22.1%	24	22.7%	8	11.1%	2	2.9%	40	40.3%	41	31.5%	30	15.6%	35	17.5%	7	49.0%	10	70.9%	29	16.8%	14	6.2%
Reading Comprehension	64	44.2%	47	37.6%	26	36.7%	9	13.1%	73	61.5%	73	55.3%	66	35.8%	61	33.6%	7	49.0%	12	84.1%	85	43.2%	78	38.5%
Listening Comprehension	2	1.9%	3	2.6%	0	0.0%	1	1.5%	6	4.0%	7	4.5%	2	0.9%	2	1.3%	0	0.0%	2	17.2%	2	1.2%	7	4.3%
Number Recognition	2	0.5%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	10.6%	0	0.0%	0	0.0%
Number Discrimination	2	0.5%	0	0.0%	0	0.0%	0	0.0%	3	2.3%	6	4.5%	0	0.0%	1	0.3%	0	0.0%	1	10.6%	2	0.7%	0	0.0%
Missing Number	7	3.3%	2	0.6%	4	6.0%	2	2.0%	5	4.0%	5	2.7%	1	0.7%	3	1.9%	0	0.0%	1	10.6%	4	1.4%	4	1.5%
Addition 1	9	2.8%	1	0.8%	0	0.0%	2	2.9%	9	7.4%	13	9.7%	7	4.7%	12	7.4%	0	0.0%	3	16.2%	9	3.5%	8	3.4%
Addition 2	16	7.3%	10	6.5%	5	7.1%	7	10.9%	26	20.1%	32	23.5%	17	9.6%	28	18.0%	1	5.8%	3	16.2%	25	13.6%	29	14.8%
Subtraction 1	13	4.5%	6	3.2%	4	4.8%	3	8.0%	10	8.6%	22	16.0%	9	5.4%	26	13.9%	4	21.0%	3	16.2%	15	8.0%	18	9.9%
Subtraction 2	51	28.5%	48	30.4%	26	35.1%	21	41.6%	48	41.0%	63	47.8%	48	23.9%	89	48.8%	7	44.8%	9	60.3%	61	31.7%	103	50.7%
Word Problems	25	14.2%	27	20.4%	7	11.6%	5	12.5%	25	23.5%	31	24.7%	23	10.8%	27	14.1%	4	25.3%	5	29.6%	22	9.9%	25	11.8%

Note: An asterisk (*) indicates differences that are statistically significant at the p<0.05 level.

Annex G: 2017 Versus 2019 Analysis

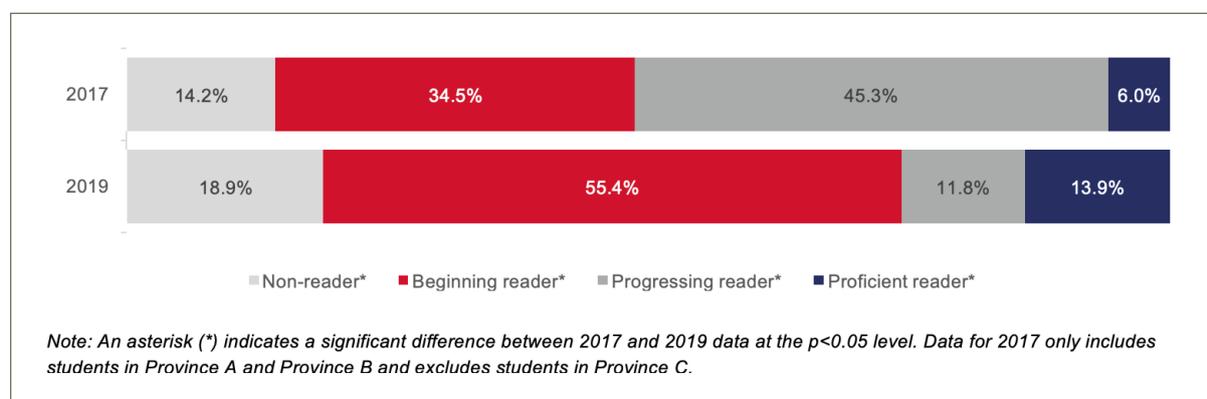
To compare scores, STS obtained analytical data files from Chemonics International for the 2017 EGRA and EGMA and used the scores and weights provided in these files for analysis.⁵⁷ For EGRA performance, STS compared the proportion of students in each proficiency category. STS did not compare mean scores by subtask because of changes to the letter sound identification, nonword reading and ORF subtasks (see Annex C). Conversely, analysis compares average accuracy scores and fluency rates for EGMA subtasks because only minor changes were made.

In reading, students tested in 2017 at the end of Grade 3 were more likely to be classified as progressing readers, while students tested in 2019 at the beginning of Grade 3 were more likely to be classified as beginning readers (Figure 2). This is not surprising considering the difference in school year progression between the 2017 and 2019 students. The effect of the differences in the tools, however, cannot be isolated. Comparing students at different timepoints in the school year and with slightly different tools means that the comparison of the previous project’s 2017 results and Manahel 2019 results is not a useful one.

READING OUTCOMES – 2017 AND 2019

Students tested in 2017 at the end of Grade 3 were more likely to be classified as progressing readers. Students tested in 2019 at the beginning of Grade 3 were more likely to be classified as beginning readers. The proportion of students who were non-readers, beginning readers or proficient readers was significantly higher in 2019 than in 2017 (Figure 74). However, the proportion of students who were progressing readers was significantly higher in 2017.

Figure 74. Percentage of Students by Reading Proficiency Level and Year



⁵⁷ Weights provided in the data file were applied. Results for students in Province C were excluded from all analyses.

Trends for students in Province A matched overall trends. However, in Province B, there were significantly more beginning readers and proficient readers in 2019 and significantly more progressing readers in 2017. There was no statistically significant difference in the proportion of zero scores in 2017 and 2019 (Annex F).

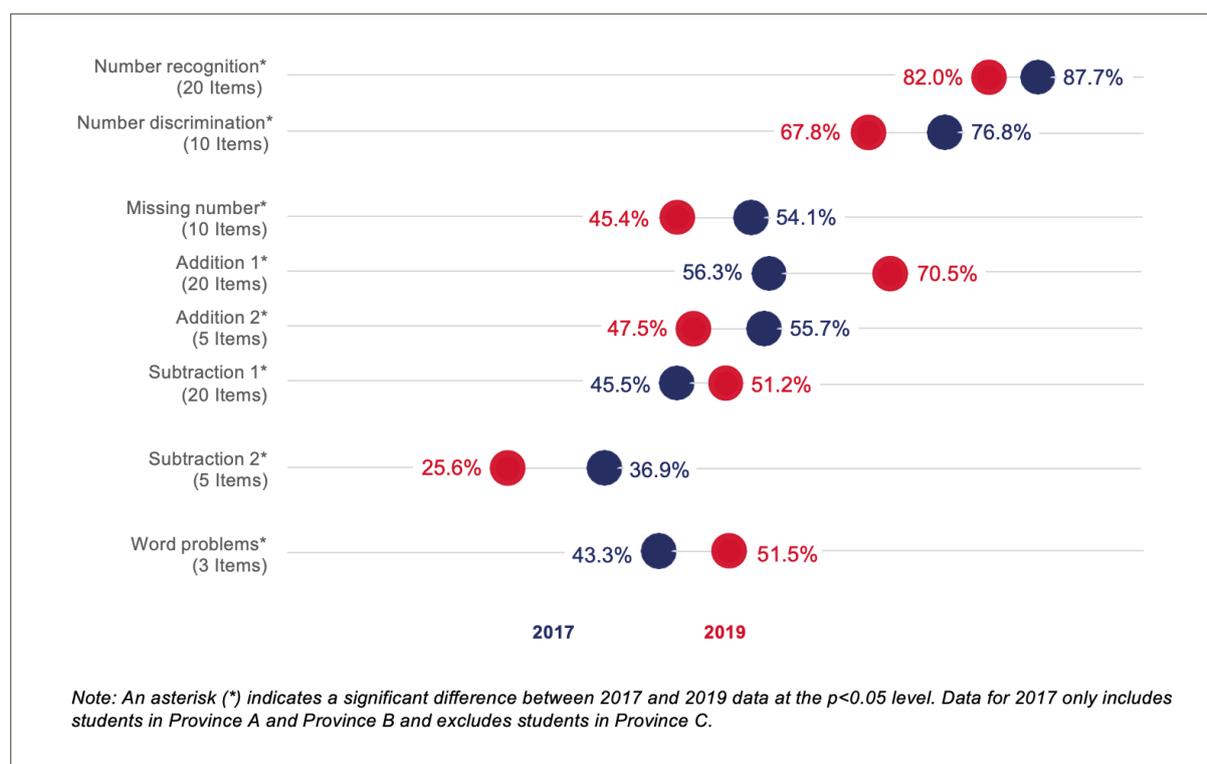
Reading Outcomes by Gender – 2017 and 2019

Boys were more likely to be beginning or proficient readers in 2019 than in 2017, while the opposite was true for progressing readers. Boys in 2019 and in 2017 were equally likely to be non-readers. Girls were more likely to be non-readers, beginning readers or proficient readers in 2019 than in 2017, but the opposite held true for progressing readers (Annex D).

MATHEMATICS OUTCOMES – 2017 AND 2019

Students at the end of Grade 3 in 2017 had higher accuracy scores on nearly all EGMA subtasks than students at the beginning of Grade 3 in 2019 (Figure 75). On addition level 1, students at the beginning of Grade 3 in 2019 outperformed students at the end of Grade 3 in 2017.

Figure 75. Mathematics Percentage of Items Correct (Accuracy Scores) by Subtask



Students in Province B have higher performance than students in Province A on five subtasks in 2017 and on three subtasks in 2019. In 2017, students in Province B had significantly higher accuracy scores on missing number, addition levels 1 and 2, subtraction level 2 and word problems. In 2019, students in Province B had significantly higher accuracy scores on number discrimination, missing number and word problems.

Mathematics Outcomes by Gender – 2017 and 2019

By gender, boys had higher performance than girls in both 2017 and 2019, but the gap has widened. In 2017, boys had significantly higher performance on two subtasks – number discrimination and subtraction level 1 – but in 2019, they outperformed girls on all mathematics subtasks.

The comparison in mathematics performance across the two studies is equally problematic as in reading. In addition to the developmental difference in students in the two cohorts, the higher performance of beginning-of-year students in addition, subtraction and word problems suggests differences in population – which may be differences in the curriculum, schooling experiences, demographics, or a number of other factors.

The inability of the evaluation to isolate true differences in learning from differences in developmental stage, tools, and population means that no meaningful conclusions can be drawn from the comparison of 2017 and 2019 data. Challenges to this comparison are described further below.

CAVEATS FOR 2017 AND 2019 STUDY COMPARISON

As noted above, while the results of the individual studies in 2017 and 2019 are sound, comparisons between the two should be made with caution. The seemingly higher student performance in 2017 does not point directly to conclusions about the performance of the program in 2019, as several factors must be considered. These factors include the different developmental stages at which students were assessed in 2017 and 2019; the changes in the tool; and the changes in the conflict between 2017 and 2019, including high levels of displacement and new schools attended in the 2019 sample. The factors that render the comparison between 2017 and 2019 inappropriate are outlined in the table below.

Table 16. Changes between 2017 and 2019 Assessment

Comparison factor	2017 context	2019 context
Student grade level	End of grade 3	Beginning of grade 3
EGRA instruments	Missing modifiers in letter sounds, nonwords, and ORF subtasks	Added modifiers in letter sounds, nonwords, and ORF subtasks
Average number of moves due to conflict	Not reported or not available	3.9
Average number of new schools attended	Not reported or not available	2.3
Prevalence of other education interventions in Manahel schools sampled	Not reported or not available	Not available

Annex H: Manahel Levels Analysis

Table H.1. Summary of Results by Manahel Level and Subgroup

	Manahel Level 1: Learner can recognize letter names		Manahel Level 2: Learner can read letters, words and sentences with short vowels (Al Madd)		Manahel Level 3 and 4: Learner can read and sound out all letters, words and sentences with the Sokoon modifier; Learner can read letters, words, and sentences with long vowels (Al Madd)		Manahel Level 7 and 8: Learner reads sentences with comprehension; Learner reads paragraphs with comprehension		Significant Differences			
	n	%	n	%	n	%	n	%	L1	L2	L3 & L4	L7 & L8
	Overall	321	21.7	11	0.8	10	0.7	82	5.5			
Province												
Province A	193	19.7%	9	0.8%	8	0.8%	61	6.0%	Province B> Province			
Province B	106	26.8%	3	0.6%	2	0.3%	17	4.4%				
District												
D1: District 1	57	17.9%	4	1.4%	4	1.4%	13	4.3%	D2>D1/D3/ D5; D6>D3			D4>D3
D2: District 2	29	32.6%	0	0.0%	0	0.0%	9	6.5%				
D3: District 3	30	12.9%	0	0.0%	0	0.0%	4	2.2%				
D4: District 4	75	21.6%	5	1.3%	4	1.2%	34	8.5%				
D5: District 5	2	4.3%	0	0.0%	0	0.0%	1	2.6%				
D6: District 6	106	26.8%	3	0.6%	2	0.3%	17	4.4%				
School Type												
AO	83	21.1%	5	1.5%	4	1.3%	28	6.5%		AO>QE		
QE	216	21.9%	7	0.5%	6	0.4%	50	5.1%				

	Manahel Level 1: Learner can recognize letter names		Manahel Level 2: Learner can read letters, words and sentences with short vowels (Al Madd)		Manahel Level 3 and 4: Learner can read and sound out all letters, words and sentences with the Sokoon modifier; Learner can read letters, words, and sentences with long vowels (Al Madd)		Manahel Level 7 and 8: Learner reads sentences with comprehension; Learner reads paragraphs with comprehension		Significant Differences			
	n	%	n	%	n	%	n	%	L1	L2	L3 & L4	L7 & L8
QE only												
QE & new QE	187	22.2%	7	0.6%	6	0.5%	39	4.5%				QE ML > QE & new QE
QE-Mobile Library	29	20.6%	0	0.0%	0	0.0%	11	8.0%				
Proficiency Band												
Non-reader	3	1.3%	0	0.0%	0	0.0%	0	0.0%	Proficient, progressing reader > non-reader, beginning reader; beginning reader > non-reader			
Beginning reader	110	15.2%	0	0.0%	0	0.0%	0	0.0%				
Progressing reader	74	48.6%	2	2.0%	1	1.7%	0	0.0%				
Proficient reader	110	60.1%	10	4.1%	9	3.7%	78	42.3%				
Number of moves due to war (IDP moves)												
1-5 times	189	24.2%	11	1.3%	9	1.1%	47	6.2%	1-5 moves > 6 or more	1-5 moves > 6 or more	1-5 moves > 6 or more	
6+ times	110	18.4%	1	0.1%	1	0.1%	31	4.7%				
Household size												
1-5 people	69	22.7%	4	0.9%	3	0.7%	20	6.9%				
6+ people	230	21.4%	8	0.7%	7	0.7%	58	5.1%				
Attendance												

	Manahel Level 1: Learner can recognize letter names		Manahel Level 2: Learner can read letters, words and sentences with short vowels (Al Madd)		Manahel Level 3 and 4: Learner can read and sound out all letters, words and sentences with the Sokoon modifier; Learner can read letters, words, and sentences with long vowels (Al Madd)		Manahel Level 7 and 8: Learner reads sentences with comprehension; Learner reads paragraphs with comprehension		Significant Differences			
	n	%	n	%	n	%	n	%	L1	L2	L3 & L4	L7 & L8
No days missed in past 5 days	252	23.8%	10	0.8%	9	0.8%	69	6.1%	not missed any days > missed 1 or more			
Missed 1+days in past 5 days	47	13.8%	2	0.5%	1	0.2%	9	3.4%				

Annex I: Accuracy Scores for Low, Moderate, and High Performers

Table I.1. Proportion of Students in Low, Moderate and High-Performing Groups for Reading Mechanics, by Subtask and Subgroup

	Letter Sound Identification						Nonword Reading						Oral Reading Fluency					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Province																		
Province A	409	38.9%	435	41.3%	209	19.9%	925	88.1%	115	10.9%	10	1.0%	646	63.7%	237	23.4%	131	12.9%
Province B	106	25.8%	195	47.3%	111	27.0%	358	87.4%	48	11.7%	4	0.9%	281	70.9%	78	19.7%	37	9.4%
District																		
District 1	81	39.3%	87	42.5%	37	18.2%	189	91.4%	15	7.3%	3	1.4%	139	68.1%	50	24.6%	15	7.3%
District 2	26	23.3%	48	43.8%	36	32.8%	84	77.8%	24	22.2%	0	0.0%	73	68.1%	22	20.4%	12	11.5%
District 3	115	50.5%	83	36.4%	30	13.1%	223	97.7%	5	2.3%	0	0.0%	146	74.2%	36	18.1%	15	7.7%
District 4	173	35.9%	204	42.4%	105	21.7%	403	84.1%	69	14.3%	7	1.6%	264	55.3%	128	26.7%	86	17.9%
District 5	15	51.8%	12	43.9%	1	4.3%	26	92.8%	2	7.2%	0	0.0%	24	84.3%	2	5.9%	3	9.8%
District 6	106	25.8%	195	47.3%	111	27.0%	358	87.4%	48	11.7%	4	0.9%	281	70.9%	78	19.7%	37	9.4%
School Shifts																		
Full time	277	35.7%	326	42.0%	173	22.3%	651	84.0%	115	14.8%	10	1.2%	474	63.9%	169	22.8%	99	13.3%
Morning Shift	221	36.8%	256	42.5%	124	20.7%	550	92.1%	44	7.4%	3	0.5%	393	67.6%	125	21.5%	63	10.8%
Evening Shift	14	17.3%	46	58.1%	19	24.5%	72	93.1%	4	5.3%	1	1.6%	52	66.3%	21	27.1%	5	6.5%
Implementation Waves																		
Wave A	121	34.6%	156	44.4%	74	21.1%	317	90.6%	28	8.0%	5	1.4%	227	69.7%	61	18.6%	38	11.7%
Wave B	49	43.7%	35	30.9%	29	25.4%	101	89.4%	12	10.6%	0	0.0%	77	68.3%	23	20.3%	13	11.5%
Wave C	12	27.0%	29	65.2%	3	7.8%	33	72.9%	12	25.9%	1	1.1%	23	52.0%	17	38.9%	4	9.2%
Wave D	118	45.8%	96	37.0%	44	17.1%	227	89.0%	28	11.0%	0	0.0%	153	62.4%	64	26.1%	28	11.5%
N/A	215	30.7%	315	45.0%	170	24.3%	606	86.9%	83	11.9%	8	1.2%	448	65.5%	151	22.1%	85	12.4%
School Type																		
AO	142	33.4%	193	45.4%	91	21.3%	370	87.0%	49	11.4%	6	1.5%	258	61.9%	100	24.0%	59	14.1%
QE	264	33.2%	347	43.8%	183	23.0%	696	87.9%	88	11.1%	7	0.9%	523	68.5%	157	20.6%	84	11.0%

	Letter Sound Identification						Nonword Reading						Oral Reading Fluency					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
QE-Mobile Library	85	43.4%	70	36.0%	40	20.7%	168	87.2%	25	12.8%	0	0.0%	106	58.2%	52	28.5%	24	13.2%
New QE	25	49.1%	19	37.2%	7	13.7%	49	97.4%	1	2.6%	0	0.0%	40	84.1%	6	13.1%	1	2.7%
Access Only versus Quality Education																		
AO	142	33.4%	193	45.4%	91	21.3%	370	87.0%	49	11.4%	6	1.5%	258	61.9%	100	24.0%	59	14.1%
QE	374	35.9%	436	42.0%	230	22.1%	913	88.2%	114	11.0%	7	0.7%	669	67.3%	216	21.7%	109	11.0%
Quality Education – Fixed versus Mobile Libraries																		
QE and new QE	289	34.2%	366	43.4%	189	22.4%	745	88.5%	90	10.6%	7	0.9%	563	69.4%	163	20.1%	85	10.5%
QE Mobile Library	85	43.4%	70	36.0%	40	20.7%	168	87.2%	25	12.8%	0	0.0%	106	58.2%	52	28.5%	24	13.2%
Household Size																		
1-5 people	108	32.3%	150	45.0%	76	22.7%	287	86.9%	41	12.3%	3	0.9%	203	62.7%	79	24.4%	42	12.9%
6+ people	408	36.0%	479	42.4%	244	21.6%	996	88.2%	122	10.8%	11	1.0%	724	66.6%	236	21.8%	126	11.6%
Number of School Moves																		
1-5 times	278	33.9%	343	41.8%	200	24.3%	696	85.1%	108	13.3%	13	1.6%	500	62.7%	193	24.2%	105	13.1%
6+ times	238	36.9%	286	44.4%	121	18.7%	588	91.4%	55	8.5%	1	0.1%	427	69.7%	122	19.9%	63	10.3%
Attendance																		
Missed 0 school days in past 5 days	386	33.0%	504	43.2%	278	23.8%	1009	86.9%	140	12.0%	12	1.0%	698	62.4%	269	24.0%	152	13.6%
Missed 1+ school day(s) in past 5 days	130	43.6%	126	42.2%	42	14.2%	275	91.7%	23	7.8%	2	0.5%	229	78.6%	47	16.0%	16	5.4%

Table I.2. Proportion of Students in Low, Moderate and High-Performing Groups for Understanding and Comprehension, by Subtask and Subgroup

	Listening Comprehension						Reading Comprehension					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%
Province												
Province A	112	10.6%	200	18.9%	745	70.5%	775	73.3%	214	20.3%	68	6.5%
Province B	48	11.5%	105	25.4%	261	63.1%	323	78.2%	72	17.3%	18	4.5%
District												
District 1	19	9.0%	40	19.5%	149	71.5%	154	74.2%	44	21.2%	10	4.6%
District 2	1	0.8%	11	9.8%	98	89.4%	76	69.4%	26	24.0%	7	6.5%
District 3	45	19.5%	62	27.0%	123	53.4%	192	83.7%	31	13.4%	7	2.9%
District 4	40	8.2%	80	16.7%	362	75.1%	329	68.2%	109	22.7%	44	9.2%
District 5	8	29.1%	6	22.9%	13	48.1%	24	84.3%	4	13.1%	1	2.6%
District 6	48	11.5%	105	25.4%	261	63.1%	323	78.2%	72	17.3%	18	4.5%
School Shift												
Full time	96	12.3%	144	18.5%	538	69.2%	570	73.3%	161	20.7%	47	6.1%
Morning Shift	58	9.6%	142	23.6%	403	66.8%	462	76.6%	103	17.0%	38	6.4%
Evening Shift	4	4.9%	17	21.9%	58	73.2%	57	72.6%	22	27.4%	0	0.0%
School Wave												
Wave A	44	12.5%	81	23.0%	227	64.5%	271	77.0%	63	18.0%	18	5.0%
Wave B	10	8.5%	25	22.4%	79	69.1%	80	70.8%	27	23.5%	7	5.7%
Wave C	9	19.2%	8	16.9%	29	63.9%	26	58.6%	13	29.2%	5	12.2%
Wave D	30	11.4%	45	17.5%	184	71.1%	192	74.0%	50	19.2%	18	6.8%
N/A	68	9.7%	146	20.8%	487	69.5%	529	75.4%	133	19.0%	39	5.6%
School Type												
AO	43	10.1%	70	16.4%	314	73.5%	318	74.5%	80	18.7%	29	6.8%
QE	87	10.8%	183	23.0%	528	66.2%	596	74.7%	161	20.1%	41	5.1%
QE-Mobile Library	24	12.3%	36	18.3%	136	69.4%	139	71.2%	41	20.8%	16	8.0%
New QE	6	12.4%	16	31.8%	28	55.8%	45	89.0%	5	9.6%	1	1.4%
Access Only versus Quality Education												
AO	43	10.1%	70	16.4%	314	73.5%	318	74.5%	80	18.7%	29	6.8%

	Listening Comprehension						Reading Comprehension					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%
QE	117	11.2%	235	22.5%	692	66.3%	780	74.7%	206	19.8%	57	5.5%
Quality Education – Fixed versus Mobile Libraries												
QE and new QE	93	10.9%	199	23.5%	556	65.6%	641	75.6%	166	19.5%	42	4.9%
QE Mobile Library	24	12.3%	36	18.3%	136	69.4%	139	71.2%	41	20.8%	16	8.0%
Household Size												
1-5 people	15	4.5%	82	24.5%	238	71.0%	239	71.6%	71	21.3%	24	7.1%
6+ people	145	12.8%	223	19.6%	768	67.6%	859	75.6%	214	18.9%	63	5.5%
Number of School Moves												
1-5 times	59	7.1%	157	19.1%	608	73.8%	595	72.2%	175	21.2%	54	6.5%
6+ times	101	15.6%	148	22.8%	398	61.5%	503	77.8%	111	17.1%	33	5.1%
Attendance												
Missed 0 school days in past 5 days	126	10.8%	224	19.2%	820	70.1%	845	72.2%	249	21.3%	76	6.5%
Missed 1+ school day(s) in past 5 days	34	11.3%	81	26.9%	185	61.8%	253	84.3%	36	12.2%	11	3.5%

Table I.6. Proportion of Students in Low, Moderate and High-Performing Groups for Whole Numbers, by Subtask and Subgroup

	Number Identification						Number Discrimination						Missing Number					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Province																		
Province A	50	4.8%	379	36.0%	624	59.2%	268	25.5%	418	39.7%	367	34.9%	594	56.4%	381	36.2%	78	7.4%
Province B	22	5.4%	169	40.8%	222	53.8%	65	15.8%	182	44.0%	166	40.1%	208	50.3%	170	41.0%	36	8.7%
District																		
District 1	7	3.5%	66	31.9%	134	64.6%	26	12.3%	91	43.5%	92	44.1%	114	54.9%	81	38.7%	13	6.4%
District 2	0	0.3%	39	35.1%	71	64.6%	14	12.5%	61	55.5%	35	32.1%	65	59.0%	40	36.2%	5	4.9%
District 3	14	6.4%	108	47.6%	105	46.1%	97	42.8%	79	34.7%	51	22.5%	159	70.0%	59	26.0%	9	4.0%

	Number Identification						Number Discrimination						Missing Number					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
District 4	25	5.2%	153	32.0%	302	62.8%	123	25.5%	174	36.2%	184	38.3%	234	48.9%	198	41.2%	48	9.9%
District 5	3	11.9%	13	45.7%	12	42.4%	9	31.9%	14	48.9%	5	19.1%	21	75.0%	4	14.7%	3	10.3%
District 6	22	5.4%	169	40.8%	222	53.8%	65	15.8%	182	44.0%	166	40.1%	208	50.3%	170	41.0%	36	8.7%
School Shift																		
Full time	37	4.7%	290	37.6%	444	57.7%	186	24.1%	300	39.0%	284	36.9%	418	54.3%	288	37.3%	64	8.4%
Morning Shift	30	5.1%	227	37.9%	342	57.0%	125	20.8%	262	43.7%	213	35.5%	336	56.1%	218	36.4%	45	7.5%
Evening Shift	5	6.4%	23	29.4%	51	64.2%	19	24.3%	31	38.9%	29	36.9%	36	45.4%	39	49.9%	4	4.7%
School Wave																		
Wave A	19	5.4%	134	38.1%	199	56.5%	93	26.4%	146	41.5%	113	32.1%	221	62.9%	114	32.3%	17	4.8%
Wave B	10	9.2%	36	31.4%	67	59.3%	29	25.5%	40	35.6%	44	38.9%	58	51.0%	49	43.0%	7	6.0%
Wave C	0	0.4%	11	24.3%	34	75.3%	10	23.4%	15	32.5%	20	44.1%	22	48.6%	15	32.7%	8	18.6%
Wave D	10	3.8%	103	40.1%	144	56.1%	64	25.0%	115	44.9%	77	30.1%	147	57.2%	90	35.2%	20	7.6%
N/A	33	4.8%	265	37.8%	402	57.4%	137	19.6%	284	40.5%	280	39.9%	354	50.5%	283	40.5%	63	9.0%
School Type																		
AO	18	4.2%	146	34.3%	262	61.5%	88	20.8%	157	36.9%	180	42.4%	210	49.4%	172	40.4%	43	10.2%
QE	46	5.8%	293	36.8%	458	57.4%	178	22.4%	332	41.6%	287	36.0%	450	56.5%	292	36.7%	55	6.9%
QE-Mobile Library	8	4.0%	82	42.2%	104	53.8%	47	24.3%	89	46.0%	58	29.7%	104	53.6%	75	38.6%	15	7.8%
New QE	1	1.4%	27	54.0%	22	44.6%	20	39.5%	22	44.1%	8	16.5%	37	74.2%	12	23.3%	1	2.6%
Access Only versus Quality Education																		
AO	18	4.2%	146	34.3%	262	61.5%	88	20.8%	157	36.9%	180	42.4%	210	49.4%	172	40.4%	43	10.2%
QE	55	5.2%	403	38.6%	584	56.1%	245	23.5%	443	42.5%	353	33.9%	591	56.8%	379	36.4%	71	6.8%
Quality Education – Fixed versus Mobile Libraries																		
QE and new QE	47	5.5%	321	37.8%	480	56.6%	198	23.4%	354	41.8%	296	34.9%	488	57.5%	304	35.9%	56	6.6%
QE Mobile Library	8	4.0%	82	42.2%	104	53.8%	47	24.3%	89	46.0%	58	29.7%	104	53.6%	75	38.6%	15	7.8%
Household Size																		
1-5 people	12	3.6%	113	34.2%	205	62.2%	68	20.5%	132	39.9%	131	39.6%	169	51.1%	133	40.2%	29	8.7%

	Number Identification						Number Discrimination						Missing Number					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
6+ people	61	5.3%	435	38.3%	641	56.4%	266	23.4%	468	41.2%	403	35.4%	633	55.6%	419	36.8%	86	7.5%
Number of School Moves																		
1-5 times	40	4.9%	302	36.9%	476	58.2%	159	19.4%	332	40.6%	327	40.0%	432	52.8%	313	38.2%	74	9.0%
6+ times	33	5.1%	246	37.9%	370	57.0%	175	27.0%	268	41.3%	206	31.8%	370	57.0%	239	36.8%	40	6.2%
Attendance																		
Missed 0 school days in past 5 days	54	4.6%	414	35.6%	695	59.7%	259	22.3%	454	39.1%	449	38.6%	610	52.4%	449	38.6%	105	9.0%
Missed 1+ school day(s) in past 5 days	19	6.1%	134	44.1%	151	49.8%	74	24.4%	146	47.9%	84	27.7%	192	63.1%	103	33.7%	10	3.2%

Table I.7. Proportion of Students in Low, Moderate and High-Performing Groups for Addition, by Subtask and Subgroup

	Addition Level 1						Addition Level 2					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%
Province												
Province A	144	13.6%	489	46.5%	420	39.9%	580	55.0%	379	36.0%	95	9.0%
Province B	39	9.4%	220	53.1%	155	37.5%	213	51.4%	163	39.3%	38	9.3%
District												
District 1	15	7.1%	110	52.8%	83	40.1%	89	42.9%	90	43.1%	29	13.9%
District 2	12	11.2%	57	51.6%	41	37.2%	54	49.2%	44	39.8%	12	11.1%
District 3	50	22.0%	113	49.8%	64	28.2%	159	69.8%	59	25.7%	10	4.5%
District 4	62	12.9%	195	40.7%	223	46.4%	259	54.0%	181	37.6%	40	8.4%
District 5	5	16.2%	14	50.7%	9	33.1%	19	67.5%	6	22.1%	3	10.3%
District 6	39	9.4%	220	53.1%	155	37.5%	213	51.4%	163	39.3%	38	9.3%
School Shift												
Full time	110	14.3%	352	45.7%	308	40.0%	434	56.4%	276	35.9%	60	7.8%
Morning Shift	68	11.4%	300	50.1%	230	38.5%	308	51.5%	226	37.7%	65	10.8%

	Addition Level 1						Addition Level 2					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%
Evening Shift	2	2.5%	45	57.5%	31	40.0%	37	47.5%	33	41.8%	8	10.7%
School Wave												
Wave A	55	15.7%	163	46.5%	133	37.8%	199	56.7%	120	34.0%	33	9.3%
Wave B	19	16.6%	55	48.6%	40	34.8%	59	51.5%	44	38.8%	11	9.7%
Wave C	2	4.8%	21	46.2%	22	49.0%	18	40.0%	20	44.5%	7	15.5%
Wave D	34	13.3%	121	47.0%	102	39.7%	144	56.0%	98	38.1%	15	5.9%
N/A	72	10.2%	349	49.9%	279	39.9%	373	53.2%	260	37.1%	67	9.6%
School Type												
AO	49	11.5%	199	46.8%	178	41.7%	231	54.2%	148	34.8%	47	11.0%
QE	93	11.7%	402	50.4%	302	37.9%	417	52.3%	305	38.3%	75	9.4%
QE-Mobile Library	30	15.4%	83	43.0%	81	41.6%	105	54.5%	77	39.8%	11	5.8%
New QE	10	20.3%	25	49.5%	15	30.2%	39	77.8%	11	22.2%	0	0.0%
Access Only versus Quality Education												
AO	49	11.5%	199	46.8%	178	41.7%	231	54.2%	148	34.8%	47	11.0%
QE	133	12.8%	510	49.0%	398	38.2%	562	53.9%	394	37.8%	86	8.3%
Quality Education – Fixed versus Mobile Libraries												
QE and new QE	104	12.2%	427	50.3%	317	37.4%	456	53.8%	317	37.3%	75	8.9%
QE Mobile Library	30	15.4%	83	43.0%	81	41.6%	105	54.5%	77	39.8%	11	5.8%
Household Size												
1-5 people	38	11.6%	162	49.1%	130	39.3%	173	52.2%	127	38.3%	31	9.4%
6+ people	144	12.7%	547	48.1%	446	39.2%	620	54.5%	415	36.5%	102	9.0%
Number of School Moves												
1-5 times	88	10.8%	373	45.6%	357	43.6%	402	49.1%	335	41.0%	81	9.9%
6+ times	94	14.5%	336	51.8%	219	33.7%	391	60.2%	206	31.8%	52	8.0%
Attendance												
Missed 0 school days in past 5 days	139	11.9%	538	46.2%	486	41.8%	610	52.5%	443	38.1%	110	9.4%
Missed 1+ school day(s) in past 5 days	44	14.3%	171	56.3%	89	29.3%	183	60.0%	98	32.3%	23	7.7%

Table I.8. Proportion of Students in Low, Moderate and High-Performing Groups for Subtraction and Word Problems, by Subtask and Subgroup

	Subtraction Level 1						Subtraction Level 2						Word Problems					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Province																		
Province A	291	27.7%	668	63.4%	94	9.0%	863	81.9%	160	15.2%	31	2.9%	514	48.8%	374	35.5%	165	15.7%
Province B	112	27.0%	259	62.5%	44	10.6%	331	79.9%	66	16.0%	17	4.1%	191	46.2%	132	32.0%	90	21.8%
District																		
District 1	47	22.8%	140	67.4%	20	9.8%	154	73.8%	50	24.1%	4	2.1%	106	51.1%	77	37.1%	25	11.8%
District 2	27	24.6%	76	68.5%	8	6.9%	88	79.8%	17	15.0%	6	5.2%	50	44.9%	36	32.7%	25	22.4%
District 3	87	38.4%	127	56.0%	13	5.6%	209	92.1%	15	6.8%	3	1.1%	124	54.6%	71	31.1%	33	14.3%
District 4	115	24.0%	312	65.0%	53	11.0%	388	81.0%	76	15.7%	16	3.3%	223	46.4%	176	36.7%	81	16.9%
District 5	14	51.1%	13	46.0%	1	2.9%	24	85.2%	2	7.4%	2	7.4%	12	42.1%	14	48.7%	3	9.2%
District 6	112	27.0%	259	62.5%	44	10.6%	331	79.9%	66	16.0%	17	4.1%	191	46.2%	132	32.0%	90	21.8%
School Shift																		
Full time	208	27.0%	492	63.9%	70	9.1%	637	82.7%	110	14.3%	23	3.0%	374	48.5%	276	35.8%	121	15.6%
Morning Shift	167	27.8%	368	61.4%	65	10.8%	474	79.2%	104	17.4%	21	3.4%	281	46.9%	192	32.1%	126	21.0%
Evening Shift	24	30.2%	52	65.7%	3	4.1%	65	82.4%	10	13.1%	4	4.6%	43	55.0%	28	36.1%	7	8.9%
School Wave																		
Wave A	113	32.2%	213	60.5%	26	7.3%	299	85.1%	43	12.1%	10	2.8%	188	53.4%	111	31.5%	53	15.1%
Wave B	25	22.2%	79	69.8%	9	8.0%	87	76.7%	21	18.4%	6	4.9%	58	51.1%	39	34.3%	17	14.5%
Wave C	4	8.8%	32	71.5%	9	19.7%	35	77.1%	10	22.9%	0	0.0%	21	46.5%	18	40.6%	6	12.9%
Wave D	63	24.4%	179	69.6%	15	6.0%	216	84.2%	37	14.5%	3	1.3%	127	49.5%	93	36.2%	37	14.3%
N/A	198	28.3%	423	60.4%	79	11.3%	557	79.5%	115	16.4%	29	4.1%	311	44.5%	245	35.0%	144	20.5%
School Type																		
AO	119	28.0%	254	59.7%	53	12.4%	334	78.4%	77	18.1%	15	3.5%	176	41.2%	165	38.7%	85	20.0%
QE	221	27.7%	505	63.3%	71	8.9%	655	82.1%	113	14.2%	29	3.6%	404	50.7%	260	32.6%	133	16.7%
QE-Mobile Library	49	25.1%	134	69.1%	11	5.8%	155	80.3%	35	18.0%	3	1.8%	98	50.6%	65	33.7%	30	15.7%

	Subtraction Level 1						Subtraction Level 2						Word Problems					
	0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct		0-40% correct		41-80% correct		81-100% correct	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
New QE	14	28.1%	33	66.3%	3	5.7%	50	98.6%	1	1.4%	0	0.0%	28	55.2%	16	31.2%	7	13.5%
Access Only versus Quality Education																		
AO	119	28.0%	254	59.7%	53	12.4%	334	78.4%	77	18.1%	15	3.5%	176	41.2%	165	38.7%	85	20.0%
QE	284	27.3%	672	64.5%	85	8.2%	860	82.6%	149	14.3%	32	3.1%	530	50.9%	341	32.8%	170	16.4%
Quality Education – Fixed versus Mobile Libraries																		
QE and new QE	235	27.7%	538	63.5%	74	8.7%	705	83.1%	114	13.5%	29	3.4%	432	50.9%	276	32.5%	140	16.5%
QE Mobile Library	49	25.1%	134	69.1%	11	5.8%	155	80.3%	35	18.0%	3	1.8%	98	50.6%	65	33.7%	30	15.7%
Household Size																		
1-5 people	77	23.3%	219	66.4%	34	10.3%	264	79.9%	53	16.1%	13	4.0%	151	45.9%	125	37.8%	54	16.4%
6+ people	326	28.7%	707	62.2%	104	9.2%	930	81.8%	173	15.2%	34	3.0%	554	48.7%	381	33.5%	202	17.7%
Number of School Moves																		
1-5 times	209	25.6%	521	63.7%	88	10.7%	636	77.8%	150	18.4%	32	3.9%	368	45.0%	290	35.5%	160	19.5%
6+ times	193	29.8%	405	62.4%	51	7.8%	558	85.9%	76	11.6%	16	2.4%	337	52.0%	216	33.2%	96	14.8%
Attendance																		
Missed 0 school days in past 5 days	303	26.1%	743	63.9%	116	10.0%	931	80.0%	188	16.2%	44	3.8%	544	46.8%	403	34.6%	216	18.6%
Missed 1+ school day(s) in past 5 days	99	32.7%	183	60.1%	22	7.3%	263	86.4%	38	12.4%	4	1.2%	162	53.1%	103	33.9%	40	13.0%

Table I.9. Comparisons of Column Proportions for EGRA Subtasks, by Subgroup^{b, c}

	Letter Sound Identification			Nonword Reading			Oral Reading Fluency			Listening Comprehension			Reading Comprehension		
	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct
	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
Province															
Province A	B C											B			
Province B		A	A								C				
District															
District 1															
District 2			A		A	. ^a						A B			
District 3	B C			B		. ^a				C	C		B		
District 4					A		A	A				A B		A	A
District 5	C					. ^a				C					
District 6		A	A								C				
School Shift															
Full time					A						B				
Morning Shift				B											
Evening Shift		A	A												. ^a
School Wave															
Wave A															
Wave B	B					. ^a									
Wave C		C			A		A								
Wave D	B C					. ^a									
N/A		A	A												
School Type															
AO												B			
QE															
QE-Mobile Library	B					. ^a									

	Letter Sound Identification			Nonword Reading			Oral Reading Fluency			Listening Comprehension			Reading Comprehension		
	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct	0-40% correct	41-80% correct	81-100% correct
	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
New QE				B		. ^a									
Access Only versus Quality Education															
AO														B	
QE													C		
Quality Education – Fixed versus Mobile Libraries															
QE and new QE						. ^a	B								
QE Mobile Library						. ^a		A							
Household Size															
1-5 people													A	A	
6+ people											BC				
Number of School Moves															
1-5 times					A	A							A	AB	
6+ times				BC							BC		C		
Attendance															
Missed 0 school days in past 5 days			AB						A	A				B	A
Missed 1+ school day(s) in past 5 days	C	C							BC				C		B

Note: Results are based on 2-sided tests. For each significant pair, the key to the category with the smaller column proportion appears in the category with the larger column proportion.

Note: Significance level for upper case letters (A, B, C) = .05"

a. This category is not used in comparisons because its column proportion is equal to zero or 1.

b. Tests are adjusted for all pairwise comparisons within a row of each innermost sub-table using the Bonferroni correction.

c. Cell counts of some categories are not integers. They were rounded to the nearest integers before performing column proportions tests.

Table I.10. Comparisons of Column Proportions for EGMA Subtasks, by Subgroup^{b, c}

	Number Identification			Number Discrimination			Missing Number			Addition Level 1			Addition Level 2			Subtraction Level 1			Subtraction Level 2			Word Problems		
	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%
	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
Province																								
Province A				BC						B												C	C	
Province B					A	A					A													AB
District																								
District 1					A	A					A			A	A								A	
District 2			A		A																			
District 3		C		BC			BC			BC	C		BC			BC			B					
District 4			B	B				A	A			B												
District 5							B									B								
District 6					A	A					A													AB
School Shift																								
Full time										B														
Morning Shift																								AB
Evening Shift											A	A												
School Wave																								
Wave A							BC																	
Wave B	B																							
Wave C									AB							A	A				a			
Wave D																								
N/A					A			A										B						A
School Type																								
AO					B			A										B					A	A
QE																								
QE-Mobile Library																								

	Number Identification			Number Discrimination			Missing Number			Addition Level 1			Addition Level 2			Subtraction Level 1			Subtraction Level 2			Word Problems		
	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%	0-40%	41-80%	81-100%
	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
New QE				C			B						B		.a				B		.a			
Access Only versus Quality Education																								
AO						B			A										B				A	A
QE					C		C										C					B	C	
Quality Education – Fixed versus Mobile Libraries																								
QE and new QE																								
QE Mobile Library																								
Household Size																								
1-5 people																								
6+ people																								
Number of School Moves																								
1-5 times						A						AB		A							A			A
6+ times				C						C	C		B						B			C		
Attendance																								
Missed 0 school days in past 5 days			B			B			AB			AB										A		A
Missed 1+ school day(s) in past 5 days		C			C		C	C		C	C								C			C		

Note: Results are based on 2-sided tests. For each significant pair, the key to the category with the smaller column proportion appears in the category with the larger column proportion.

Note: Significance level for upper case letters (A, B, C) = .05"

a. This category is not used in comparisons because its column proportion is equal to zero or 1.

b. Tests are adjusted for all pairwise comparisons within a row of each innermost sub-table using the Bonferroni correction.

c. Cell counts of some categories are not integers. They were rounded to the nearest integers before performing column proportions tests.

Annex J: Correlations Between Subtasks

Table J.1. Score Correlations with EGRA and EGMA Fluency Rates

	Proficiency Band	Letter sound identification fluency (CLSPM)	Nonword reading fluency (CNWPM)	Oral reading fluency (CWPM)	Number recognition fluency (CNPM)	Addition 1 fluency (CAPPM)	Subtraction 1 fluency (CSPPM)
Proficiency Band	1	.602**	.541**	.709**	.572**	.388**	.318**
Letter sound identification fluency (CLSPM)	.602**	1	.502**	.500**	.514**	.389**	.307**
Non-word reading fluency (CNWPM)	.541**	.502**	1	.595**	.393**	.288**	.282**
Oral reading fluency (CWPM)	.709**	.500**	.595**	1	.580**	.388**	.292**
Number recognition fluency (CNPM)	.572**	.514**	.393**	.580**	1	.624**	.456**
Addition 1 fluency (CAPPM)	.388**	.389**	.288**	.388**	.624**	1	.590**
Subtraction 1 fluency (CSPPM)	.318**	.307**	.282**	.292**	.456**	.590**	1
Letter Sounds: % correct of 100	.612**	.924**	.502**	.494**	.503**	.393**	.320**
Nonword Reading: % correct of 50	.588**	.564**	.937**	.638**	.438**	.310**	.301**
ORF: % correct of 82	.874**	.654**	.625**	.776**	.620**	.443**	.334**
Listening Comprehension: % correct of 6	.285**	.223**	.228**	.171**	.206**	.240**	.227**
Reading Comprehension: % correct of 5	.838**	.617**	.557**	.657**	.573**	.427**	.334**
Number recognition: % correct of 20	.434**	.419**	.314**	.392**	.736**	.544**	.440**
Number discrimination: % correct of 10	.413**	.371**	.296**	.371**	.610**	.529**	.430**
Missing number: % correct of 10	.400**	.393**	.342**	.400**	.577**	.540**	.487**
Addition 1: % correct of 20	.384**	.410**	.291**	.357**	.549**	.890**	.520**
Addition 2: % correct of 5	.370**	.355**	.312**	.326**	.478**	.608**	.483**
Subtraction 1: % correct of 20	.380**	.359**	.300**	.336**	.530**	.687**	.818**

	Proficiency Band	Letter sound identification fluency (CLSPM)	Nonword reading fluency (CNWPM)	Oral reading fluency (CWPM)	Number recognition fluency (CNPM)	Addition 1 fluency (CAPPM)	Subtraction 1 fluency (CSPPM)
Subtraction 2: % correct of 5	.333**	.296**	.296**	.314**	.453**	.505**	.542**
Word Problems: % correct of 3	.261**	.255**	.232**	.222**	.305**	.433**	.402**

Note: Two asterisks (**) indicates correlation is significant at the 0.01 level (2-tailed).

Table J.2. Score Correlations with EGRA Accuracy Scores

	Mechanics of Reading			Understanding	Comprehension
	Letter Sounds % Correct Out of 100 Total Items	Nonword Reading % Correct Out of 50 Total Items	Oral Reading Fluency % Correct Out of 82 Total Items	Listening Comprehension % Correct Out of 6 Total Items	Reading Comprehension % Correct Out of 5 Total Items
Proficiency Band	.612**	.588**	.874**	.285**	.838**
Letter sound identification fluency (CLSPM)	.924**	.564**	.654**	.223**	.617**
Non-word reading fluency (CNWPM)	.502**	.937**	.625**	.228**	.557**
Oral reading fluency (CWPM)	.494**	.638**	.776**	.171**	.657**
Number recognition fluency (CNPM)	.503**	.438**	.620**	.206**	.573**
Addition 1 fluency (CAPPM)	.393**	.310**	.443**	.240**	.427**
Subtraction 1 fluency (CSPPM)	.320**	.301**	.334**	.227**	.334**
Letter Sounds: % correct of 100	1	.570**	.656**	.247**	0.625
Nonword Reading: % correct of 50	.570**	1	.682**	.255**	0.613
ORF: % correct of 82	.656**	.682**	1	.271**	0.89
Listening Comprehension: % correct of 6	.247**	.255**	.271**	1	.358**
Reading Comprehension: % correct of 5	.625**	.613**	.890**	.358**	1
Number recognition: % correct of 20	.448**	.343**	.455**	0.249	0.445
Number discrimination: % correct of 10	.398**	.355**	.422**	0.229	0.422
Missing number: % correct of 10	.409**	.383**	.448**	0.283	0.445

	Mechanics of Reading			Understanding	Comprehension
	Letter Sounds % Correct Out of 100 Total Items	Nonword Reading % Correct Out of 50 Total Items	Oral Reading Fluency % Correct Out of 82 Total Items	Listening Comprehension % Correct Out of 6 Total Items	Reading Comprehension % Correct Out of 5 Total Items
Addition 1: % correct of 20	.436**	.317**	.431**	0.275	0.422
Addition 2: % correct of 5	.383**	.343**	.390**	0.265	0.415
Subtraction 1: % correct of 20	.381**	.328**	.397**	0.289	0.396
Subtraction 2: % correct of 5	.294**	.321**	.367**	0.26	0.391
Word Problems: % correct of 3	.274**	.260**	.267**	.293**	.306**

Note: Two asterisks (**) indicates correlation is significant at the 0.01 level (2-tailed).

Table J.3. Score Correlations with EGMA Accuracy Scores

	Whole Numbers			Operations				Word Problems
	Number recognition : % correct out of 20	Number discrimination: % correct out of 10	Missing number: % correct out of 10	Addition 1: % correct out of 20	Addition 2: % correct out of 5	Subtraction 1: % correct out of 20	Subtraction 2: % correct out of 5	Word Problems: % correct out of 3
Proficiency Band	.434**	.413**	.400**	.384**	.370**	.380**	.333**	.261**
Letter sound identification fluency (CLSPM)	.419**	.371**	.393**	.410**	.355**	.359**	.296**	.255**
Non-word reading fluency (CNWPM)	.314**	.296**	.342**	.291**	.312**	.300**	.296**	.232**
Oral reading fluency (CWPM)	.392**	.371**	.400**	.357**	.326**	.336**	.314**	.222**
Number recognition fluency (CNPM)	.736**	.610**	.577**	.549**	.478**	.530**	.453**	.305**
Addition 1 fluency (CAPPM)	.544**	.529**	.540**	.890**	.608**	.687**	.505**	.433**
Subtraction 1 fluency (CSPPM)	.440**	.430**	.487**	.520**	.483**	.818**	.542**	.402**
Letter Sounds: % correct of 100	.448**	.398**	.409**	.436**	.383**	.381**	.294**	.274**
Nonword Reading: % correct of 50	.343**	.355**	.383**	.317**	.343**	.328**	.321**	.260**
ORF: % correct of 82	.455**	.422**	.448**	.431**	.390**	.397**	.367**	.267**
Listening Comprehension : % correct of 6	.249**	.229**	.283**	.275**	.265**	.289**	.260**	.293**
Reading Comprehension : % correct of 5	.445**	.422**	.445**	.422**	.415**	.396**	.391**	.306**

	Whole Numbers			Operations				Word Problems
	Number recognition : % correct out of 20	Number discrimination : % correct out of 10	Missing number: % correct out of 10	Addition 1: % correct out of 20	Addition 2: % correct out of 5	Subtraction 1: % correct out of 20	Subtraction 2: % correct out of 5	Word Problems: % correct out of 3
Number recognition: % correct of 20	1	.709**	.580**	.586**	.502**	.546**	.434**	.357**
Number discrimination: % correct of 10	.709**	1	.587**	.534**	.512**	.515**	.430**	.356**
Missing number: % correct of 10	.580**	.587**	1	.533**	.570**	.564**	.561**	.420**
Addition 1: % correct of 20	.586**	.534**	.533**	1	.645**	.666**	.467**	.431**
Addition 2: % correct of 5	.502**	.512**	.570**	.645**	1	.574**	.585**	.367**
Subtraction 1: % correct of 20	.546**	.515**	.564**	.666**	.574**	1	.616**	.470**
Subtraction 2: % correct of 5	.434**	.430**	.561**	.467**	.585**	.616**	1	.392**
Word Problems: % correct of 3	0.357	0.356	.420**	.431**	.367**	.470**	.392**	1

Note: Two asterisks (**) indicates correlation is significant at the 0.01 level (2-tailed).

Annex K: Bibliography

- 'A Better Tomorrow: Syria's Children Have Their Say,' *Save the Children*, 2019, www.savethechildren.org/content/dam/usa/reports/emergency-response/syria-anniversary-report.pdf
- 'Education,' *United Nations Children's Fund*, www.unicef.org/syria/education
- 'Exposure to War Stressors Questionnaire,' *Children and War Foundation*, www.childrenandwar.org/projectsresources/measures
- "Global Proficiency Framework for Reading and Mathematics - Grades 2 to 6," *Global Alliance for Monitoring Learning*, October 2019, http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/Global-Proficiency-Framework-18Oct2019_KD.pdf
- 'Humanitarian Needs Overview for Syria,' *United Nations Office for the Coordination of Humanitarian Affairs*, www.unocha.org/syria
- 'Internally Displaced People,' Syria, *United National High Commissioner for Refugees*, www.unhcr.org/sy/internally-displaced-people
- Nicolai, Susan, Sebastien Hine, and Joseph Wales, 'Education in emergencies and protracted crises: Toward a strengthened response,' *Overseas Development Institute*, July 2015, <https://www.odi.org/publications/9688-education-emergencies-and-protracted-crises-toward-strengthened-response>
- RTI International, Early Grade Mathematics Assessment (EGMA) Toolkit, 2014.
- RTI International, Early Grade Reading Assessment (EGRA) Toolkit, Second Edition. Washington, DC: United States Agency for International Development, 2015.
- 'School Enrolment, Primary (% net) - Syrian Arab Republic,' *World Bank*, data.worldbank.org/indicator/SE.PRM.NENR?locations=SY
- 'Syria Regional Refugee Response,' Operational Portal, *United Nations High Commissioner for Refugees*, last modified December 1, 2019, data2.unhcr.org/en/situations/syria
- 'Whole of Syria Facts and Figures, Education, Mid-Year 2019,' *United Nations Children's Fund*, <https://www.unicef.org/syria/media/4231/file/UNICEFFactsandFiguresEducation2020Q2202019.pdf>
- 'Why is there a war in Syria?,' *British Broadcasting Corporation*, February 25, 2019, www.bbc.com/news/world-middle-east-35806229

Disclaimer

This document has been redacted to protect the individuals involved in the Syria Education Programme. All names of people and locations have either been altered or removed, as has any information that may identify people or locations.



The Syria Education Programme is funded by UK aid from the UK government.

Published in May 2021 by the Syria Education Programme.
© 2021, the Syria Education Programme.

Please consider the environment before printing this report.