USAID
FROM THE AMERICAN PEOPLE

RTI INTERNATIONAL

Early Grade Reading and Mathematics Assessment Baseline Report, 2022

UZBEKISTAN EDUCATION *for* EXCELLENCE PROGRAM

June 14, 2022

FLORIDA STATE UNIVERSITY 1851

STATE
MISSISSIPPI STATE UNIVERSITY

Uzbekistan Education for Excellence Program

Early Grade Reading and Mathematics Assessment Baseline Report, 2022
Cooperative Agreement No. 72011519CA00004

Submitted to:

Lisa Lahalih
Agreement Officer's Representative
USAID/Central Asia/Uzbekistan
3 Moyqorghon Street
Tashkent, Uzbekistan 70093
Tel.: (99871) 140 2486
Fax: (99871) 120 6309
lilahali@usaid.gov

Submitted by:

RTI International
3040 Cornwallis Rd
Research Triangle Park, NC 27709

Dr. Carmen Strigel
Project Manager
cstrigel@rti.org

June 14, 2022

ACKNOWLEDGEMENTS

On behalf of the United States Agency for International Development (USAID) Uzbekistan Education for Excellence Program, the authors of this report wish to acknowledge the contributions of all collaborators who made it possible to complete the Program's Early Grade Reading and Mathematics Assessment (EGRA/EGMA) baseline report. We appreciate the joint work alongside USAID's All Children Reading–Asia (ACR–Asia) Uzbekistan National EGRA/EGMA.

We acknowledge and appreciate the close collaboration of colleagues from the Ministry of Public Education (MoPE) of the Republic of Uzbekistan, who supported us throughout the process and the Republican Education Center. We also express our gratitude to the various officials from the MoPE, including Kasimkhodjaev Ulugbek Patkhidinovich, Sharif Dilmuratov, Laziz Khujakulov, Firdavs Kutliev, Gulfira Toshmirzayeva, Mukhayyo Azamova, Sherzod Shermatov, and Adham Gafurov, for the various roles they played.

We wish to acknowledge the school directors and teachers who supported the assessment in the 140 schools in the EGRA/EGMA baseline. Without your effort and dedication, we would not have been able to achieve our goal of testing more than 3,252 students.

We acknowledge M-Vector for managing the data collection process and assessors Mamadjanova Nurida, Nurmatova Zilola, Iminjanova Sohiba, Mahkamboyeva Umida, Rahimova Minavar, Mahamadjanova Sarvinoz, Xalmatova Gulchehra, Iminjanova Shoira, Tillayeva Gulsanam, and Bannapova Nodira.

We acknowledge the Uzbekistan Education for Excellence Program staff who contributed to the assessment process by providing technical and logistical support for training and data collection. Specifically, we thank Azizbay Mambetkarimov, Temurbek Rakhmatov, Retno Utaira, and Bakhtiyor Mamatkulov.

In addition, we acknowledge the RTI International home office staff who contributed, including Kellie Betts, Rachel Jordan, and Laiba Bahrrahwar for supporting the training of trainers and assessors; Simon King for data cleaning and analysis; and Peter Muyingo, Yasmin Sitabkhan, Jessica Mejia, and Geri Burkholder for writing the report.

Finally, we express our gratitude to Lisa Lahalih, the Program's Agreement Officer's Representative at USAID/Uzbekistan, for her support and encouragement throughout the work.

CONTENTS

EXHIBITS

ACRONYMS AND ABBREVIATIONS

2D	two-dimensional
3D	three-dimensional
ACR–Asia	All Children Reading–Asia
clpm	correct letters per minute
cnwpm	correct nonwords per minute
COVID-19	coronavirus disease 2019
cwpm	correct words per minute
EFL	English as a Foreign Language
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
ICT	Information and Communication Technology
MoPE	Ministry of Public Education
ORF	oral reading fluency
RTI	RTI International
STB	student textbook
TG	teacher guide
TIMSS	Trends in International Mathematics and Science Study
TPD	teacher professional development
ULA	Uzbek Language Arts
USAID	United States Agency for International Development

EXECUTIVE SUMMARY

The Government of Uzbekistan Ministry of Public Education (MoPE) is committed to an ambitious program of systematic and comprehensive reforms. The country aims to create an education system that can produce graduates with critical thinking, problem solving, and practical skills that will enable them to succeed.

To support the MoPE in achieving its reform agenda, the United States Agency for International Development (USAID) initiated the 4-year Uzbekistan Education for Excellence Program (the Program) on December 9, 2019, which will end on December 8, 2023.

This Program aims to provide the expertise and experience needed to help the MoPE to achieve and sustain three overarching results:

(1) Improved Uzbek Language Arts (ULA) and Mathematics outcomes in grades 1–4.

(2) Enhanced Information and Communication Technology (ICT) instruction for grades 1–11; and

(3) Improved English as a Foreign Language (EFL) instruction in grades 1–11.

The Program will be implemented in 919 target schools in Namangan and Sirdaryo Regions.

PURPOSE OF THE EVALUATION

To evaluate the impact of the Program's reading and mathematics components, baseline Early Grade Reading and Mathematics Assessments (EGRA and EGMA, respectively) were conducted in November and December 2021. Overall, 1,623 grade 3 and 1,629 grade 5 students from 140 Program schools participated in the EGRA/EGMA baseline. Their performance will be compared over time as these schools receive reading and mathematics interventions for the first time. The EGRA/EGMA baseline was originally planned to assess students completing grades 2 and 4 at the end of the 2019-2020 school year, in May 2020. However, the assessment was postponed because of COVID-19. A decision was made to assess grades 3 and 5 students at the beginning of the school year in November–December 2021, as proxies for students completing grades 2 and 4.

The main goal of the EGRA/EGMA baseline was to understand the status of learning prior to the intervention to facilitate measuring the intervention's impact on learning outcomes in the future. To this end, the EGRA/EGMA endline will compare data collected in November–December 2021 with data collected in May 2023. Specifically, the changes in learning outcomes in intervention schools will be compared against those observed in control schools to determine the value added by the intervention.

The overall aim of the EGRA/EGMA is to answer the following research question at endline:

- What is the overall impact of the Uzbekistan Education for Excellence Program in grades 2 and 4 on Uzbek language reading and mathematics skills?

This baseline study answered the following research questions:

- What are the pre-intervention levels of mathematics and Uzbek language reading achievement for students at the end of grade 2 and grade 4?
- Are the student achievement levels for the Program and comparison schools equivalent?

SUMMARY OF FINDINGS

The EGRA/EGMA baseline measured basic reading skills, as shown in **Exhibit ES-1**. Students were assessed in different Uzbek language literacy skills; the highest-order task was student comprehension of a grade-level–appropriate text passage. Exhibit ES-1 also provides estimates of student achievement in Program and comparison schools, focusing on baseline equivalence between the two groups.

Exhibit ES-1. Baseline Uzbek Language Literacy Achievement, by Grade, Task, and Treatment

Grade	Task	Treatment	Average
Grade 3	Letter sounds *(correct letters per minute [clpm])*	Program	93.1 [±2.0]
		comparison	91.7 [±3.0]
	Nonwords *(correct words per minute [cwpm])*	Program	38.9 [±1.3]
		comparison	38.0 [±1.8]
	Oral reading fluency *(cwpm)*	Program	47.9 [±1.9]
		comparison	49.5 [±2.9]
	Reading Comprehension *(percent score)*	Program	79.1 [±2.1]
		comparison	80.3 [±2.4]
Grade 5	Nonwords *(cwpm)*	Program	47.2 [±1.6]
		comparison	48.5 [±1.9]
	Oral reading fluency *(cwpm)*	Program	68.3 [±2.4]
		comparison	71.0 [±2.9]
	Silent reading comprehension *(percent score)*	Program	64.8 [±2.0]
		comparison	68.2 [±2.3]

As shown in the baseline results for grade 3 Uzbek language literacy, on average, students demonstrated a high level of mastery of letter sounds, with grade 3 Program students averaging 93.1 clpm. Grade 3 Program students were also able to read a passage of text at an average of 47.9 cwpm and correctly answer an average of 79.1% of the five reading comprehension questions, which is an impressive result, since it indicates that most students answered a great number of questions correctly, that is, at least 3.9 out of the 5 questions. Grade 5 Program students read text at over 1 word per second (68.3 cwpm). However, their average reading comprehension score was 64.8%. **Exhibit ES-2** compares Program and comparison estimates for equivalency. The differences between these estimates are sufficiently small to establish equivalency, as discussed further in Annex A.

Exhibit ES-2 presents students' basic mathematics skills at baseline. Grade 3 and grade 5 students were assessed on different mathematics skills appropriate for their grade. The tasks are described in Section 3.

Exhibit ES-2. Baseline Mathematics Achievement, by Grade, Task, and Treatment

Grade	Task	Treatment	Average
Grade 3	Number discrimination *(percent score)*	Program	97.5 [±0.5]
		comparison	97.5 [±0.7]
	Missing number *(percent score)*	Program	67.9 [±2.0]
		comparison	68.2 [±2.9]

Exhibit ES-2. Baseline Mathematics Achievement, by Grade, Task, and Treatment

Grade	Task	Treatment	Average
	Word problems *(percent score)*	Program	75.4 [±2.1]
		comparison	73.4 [±3.1]
	Addition *(percent score)*	Program	83.2 [±1.8]
		comparison	84.4 [±2.3]
	Subtraction *(percent score)*	Program	74.7 [±2.3]
		comparison	78.9 [±2.6]
	Relational reasoning *(percent score)*	Program	62.2 [±3.0]
		comparison	61.0 [±4.8]
	Three-dimensional (3D) spatial thinking *(percent score)*	Program	62.8 [±2.2]
		comparison	65.4 [±3.3]
Grade 5	**Overall mathematics** *(percent score)*	Program	57.4 [±2.5]
		comparison	53.6 [±2.3]
	Numbers and operations *(percent score)*	Program	61.0 [±2.5]
		comparison	57.6 [±2.4]
	Geometry *(percent score)*	Program	41.7 [±2.4]
		comparison	40.2 [±2.2]
	Measurement *(percent score)*	Program	53.5 [±3.3]
		comparison	46.1 [±3.1]
	Statistics *(percent score)*	Program	60.7 [±3.6]
		comparison	56.2 [±3.3]

Grade 3 students showed mastery of the number discrimination and addition problems, with Program student average scores of 97.5% and 83.2% correct, respectively. Overall students demonstrated proficiency with most tasks focused on basic skills like addition and subtraction. They struggled with tasks requiring higher-order mathematical reasoning skills, such as relational reasoning (62% correct) and complex word problems (50% correct on item 5 and 49% correct on item 6 [see Exhibit 19]).

In Grade 5, Program students scored an average of 57% of problems correctly, suggesting that while students are mastering the foundational skills in the early grades, they struggle with the application of these skills into real-life problems and more complex mathematics.

Presenting results by gender is an important way to investigate equity issues. **Exhibit ES-3** presents achievement on selected tasks by gender.

Exhibit ES-3. Baseline Mathematics and Reading Achievement, by grade and gender

Grade	Subject	Task	Gender	Average	Difference
Grade 3	Reading	Oral reading fluency *(cwpm)*	boys	42.6	girls +10.6
			girls	53.2	
		Reading Comprehension *(percent score)*	boys	77.7	girls +2.8
			girls	80.5	
	Mathematics	Relational reasoning *(percent score)*	boys	62.9	boys +1.4
			girls	61.5	

Exhibit ES-3. Baseline Mathematics and Reading Achievement, by grade and gender

Grade	Subject	Task	Gender	Average	Difference
		3D spatial thinking *(percent score)*	boys	64.9	boys +4.3
			girls	60.6	
Grade 5	Reading	Oral reading fluency *(cwpm)*	boys	61.5	girls +13.5
			girls	75.0	
		Silent reading comprehension *(percent score)*	boys	66.4	boys +3.2
			girls	63.2	
	Mathematics	Overall mathematics *(percent score)*	boys	58.4	boys +2.0
			girls	56.4	

The results by gender present an interesting narrative. Girls outperformed boys by a significant margin for oral reading fluency—+10.6 cwpm and +13.5 cwpm in grades 3 and grades 5, respectively. However, although girls outperformed boys in reading fluency, they did not outperform boys in reading comprehension. All other differences in estimates by gender are relatively modest, except for 3D spatial thinking in grade 3, where boys outscored girls. Baseline differences in student achievement by gender are further explored in **Section 3**.

SECTION 1: BACKGROUND

1.1 PROGRAM OVERVIEW

The Government of Uzbekistan Ministry of Public Education (MoPE) is committed to an ambitious program of systematic and comprehensive reforms. The country aims to create an education system that can produce graduates with critical thinking, problem solving, and practical skills that will enable them to succeed.

To support the MoPE in achieving its reform agenda, the United States Agency for International Development (USAID) initiated the 4-year Uzbekistan Education for Excellence Program (the Program) on December 9, 2019. The Program is implemented by a consortium of implementing partners including RTI International (RTI) as the Consortium lead and Florida State University and Mississippi State University. The RTI Consortium provides the expertise and experience needed to help the MoPE to achieve and sustain three overarching results:

1. Improved Uzbek Language Arts (ULA) and Mathematics outcomes in grades 1–4.
2. Enhanced Information and Communication Technology (ICT) instruction for grades 1–11; and
3. Improved English as a Foreign Language (EFL) instruction in grades 1–11.

1.2 LIFE OF THE PROGRAM ANTICIPATED ACHIEVEMENTS

Over the life of the Program, in close collaboration with the MoPE, the Program will:

- Develop relevant and appropriate student learning standards for ULA, Mathematics, ICT, and EFL.
- Customize or develop and pilot revised student textbooks (STBs) and teacher guides (TGs).
- Design and implement an in-service teacher professional development (TPD) approach.
- Conduct Program monitoring, evaluation, and learning activities, including impact evaluation research.

The new approaches to curriculum product development and support for TPD include a digital platform for instructional materials. These materials and approaches will be used as centerpieces to help enhance teachers' capacity to understand, apply, reflect on, and improve classroom practices. It is expected that the improvements in curriculum products and in teacher capacity will translate into improvements in student achievement over time.

The implementation of activities will also provide the Program and the MoPE with lessons that can be utilized to ensure a solid scale-up of the Program's specific interventions. In addition, the Program's approach to TPD could be introduced to other regions and districts outside of the Program's two target regions. The Program also includes a focus on implementation science to look closely at what is working, how and why, and what effect the changes are having on improving teaching and learning.

This report details the baseline Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) findings of the USAID Uzbekistan Education for Excellence Program.

1.3 DISRUPTION TO EDUCATION CAUSED BY COVID-19

In response to the global coronavirus disease 2019 (COVID-19) pandemic, Uzbekistan suspended in-person schooling on March 16, 2020, for the remainder of the 2019–2020 school year.

Students returned to school on September 14, 2020, at the beginning of the 2020–2021 school year. When any student tested positive for COVID-19, the school would quarantine the entire class for 14 days. Students subsequently returned in-person teaching and learning after the 14-day quarantine period.

Students did not experience lost class time during the 2021–2022 school year.

To ensure education continuity, the MoPE televised video lessons for all subject and grades. At the end of each lesson, homework was assigned to students, who uploaded their assignments to the MoPE's electronic journal system. Teachers then assessed and provided feedback to each student using the same system. The MoPE supplemented the television lessons with online learning, using the Telegram mobile app and a digital platform.[1]

[1] Sankar, D. (2020, September). *Education continuity in COVID-19 pandemic times: Impressions on introducing distance learning in basic education in Uzbekistan.* UNICEF Uzbekistan. https://uzbekistan.un.org/sites/default/files/2020-11/DL%20Rapid%20Assessment%20report%20Oct%206%202020.pdf

SECTION 2: STUDY DESIGN

2.1 PURPOSE OF THE STUDY

This EGRA/EGMA seeks to understand the impact of the USAID Uzbekistan Education for Excellence Program on student learning outcomes in mathematics and Uzbek language reading. Findings from this EGRA/EGMA baseline provide a snapshot of student achievement and facilitate valid comparability between Program and comparison schools. At the end of the Program, a follow-up endline assessment will evaluate the Program's impact on learning outcomes.

2.2 RESEARCH QUESTIONS

The overall goal of the EGRA/EGMA is to evaluate the Program's impact on students' learning. To this end, we will compare data collected in November–December 2021 with data to be collected in May 2023. The secondary goal of the EGRA/EGMA is to provide a snapshot of student achievement. To achieve these two goals, the following research questions will be addressed:

1. What is the overall impact of the Uzbekistan Education for Excellence Program in grades 2 and 4 on Uzbek language reading and mathematics skills?
2. What are the pre-intervention levels of mathematics and Uzbek language reading achievement for students at the end of grades 2 and 4?
3. Are the student achievement levels for the Program and comparison schools equivalent?

This EGRA/EGMA baseline report responds to research questions 2 and 3. The EGRA/EGMA endline will answer research question 1.

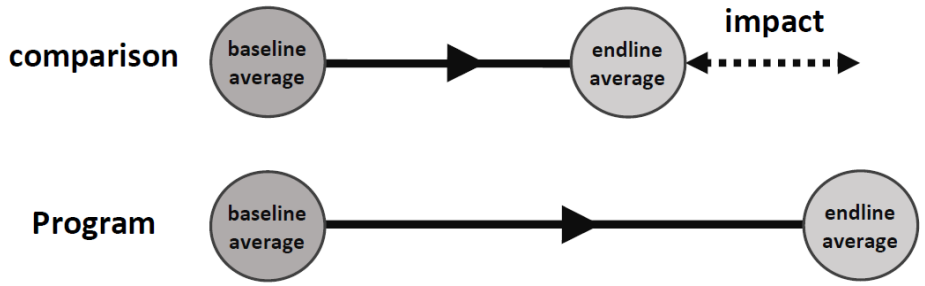
The assessments used to answer the research questions were designed for students at the end of their grade 2 and grade 4 school years. The Program EGRA/EGMA baseline was aligned with and used data from the USAID All Children Reading–Asia (ACR–Asia) National EGRA/EGMA Survey, which was originally scheduled for May 2020, at the end of the 2019–2020 school year. However, because of the global COVID-19 pandemic, the survey had to be postponed. The National EGRA/EGMA Survey was administered in November–December 2021. As these months represented the beginning of the school year, instead of the end, the decision was made to administer the assessments to students in grades 3 and 5 to ensure they would capture the learning of students who had completed 2 and 4 years of schooling.

2.3 MEASURING IMPACT

The EGRA/EGMA are designed to be conducted with Program and comparison groups at two (or more) time points, returning to the same schools and grades and sampling new students each time. This cross-sectional design measures the impact of improved teaching on student learning outcomes while maintaining comparability between the Program and comparison schools.

A typical difference-in-differences analysis will be applied at endline to measure impact. This analysis is a calculation of the difference between the comparison and Program groups' average gains in learning outcomes, as shown in **Exhibit 1**.

Exhibit 1. Impact Calculation



The evaluation is considered to be balanced if the comparison and Program averages are similar, such that we can be confident we are assessing apples to apples. This balance is assessed in **Annex A**, which details a small and acceptable difference between the comparison and Program baseline averages. The Jizzakh Region comparison group was selected from the USAID ACR–Asia-funded Uzbekistan national EGRA/EGMA survey.

2.4 SAMPLING

This EGRA/EGMA baseline aligned with the National Survey and leveraged data collected for that study. The sample design was a two-stage sample: the first stage consisted of sampling schools within a district, and the second stage involved randomized student sampling (up to 12 students per grade) during school visits for data collection. Sampling of schools used a probability-proportional-to-size methodology, meaning that schools with more students were more likely to be selected than smaller schools. This industry-standard sampling methodology facilitates learning outcome estimates that are representative of the desired sub-population.

2.5 SCHOOL AND STUDENT CHARACTERISTICS

School and student sample characteristics are presented in **Exhibit 2**.

Exhibit 2. School Sample Characteristics, by Grade and Treatment Group

Region	Program (Namangan and Sirdaryo Regions)				Comparison (Jizzakh Region)			
	Grade 3		Grade 5		Grade 3		Grade 5	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Number of Schools	140				59			
Number of Students	811	812	820	809	357	354	344	351

The ACR–Asia National EGRA/EGMA Assessments reported results by region and sampled up to 60 schools per region, irrespective of the number of schools in the region. Consequently, as **Exhibit 2** shows, 140 Program schools were sampled from the two targeted regions. In each school, 12 students (6 girls and 6 boys) were randomly selected per grade, for a total of 24 students per school. The Program assessed a total of 3,252 students. As the EGRA/EGMA was designed to compare Program results against just one comparison region, fewer schools were sampled for the comparison group. **Annex A** describes the selection of the comparison region in more detail.

2.6 ASSESSOR TRAINING AND DATA COLLECTION

Assessor training took place November 1–12, 2021. A total of 183 assessors scored at least 90% or above on the final Assessor Accuracy Measure and were involved in collecting data; data collection took place between November 15 and December 14, 2021. During that time, the assessors visited the sampled schools in teams of four, with the trainers acting as field coordinators and supporting the assessor teams.

2.7 SURVEY INSTRUMENTS

2.7.1 Survey Instruments

Exhibit 3 provides an overview of the EGRA and EGMA tasks administered by grade; more detail is provided later in this section.

Exhibit 3. Overview of EGRA/EGMA Task by Grade

Language	Grade 3	Grade 5
EGRA		
Assessed in Uzbek	▪ Letter sound identification ▪ Nonword decoding ▪ Oral reading fluency (grade 2-level text) ▪ Oral reading comprehension (grade 2-level text)	▪ Nonword decoding ▪ Oral reading fluency (grade 4-level text) ▪ Silent reading comprehension (grade 4-level text)
EGMA		
Instructions given in the language of instruction	▪ Quantity discrimination ▪ Missing number ▪ Addition/subtraction ▪ Word problems ▪ Relational reasoning ▪ Three-dimensional (3D) spatial thinking	▪ Numbers and operations ▪ Geometry ▪ Measurement ▪ Statistics

2.7.2 EGRA for Grade 3 and Grade 5 Students

The assessment tasks were designed to measure developmentally appropriate reading and mathematics skills at each grade level. For reading, the goal of learning to read is the ability to construct meaning from written text, or comprehension. Comprehension is a complex skill or composite behavior made possible by the mastery and simultaneous use of a wide array of subskills. EGRA measures both lower-order and higher-order reading skills, including the following. ***Letter sound knowledge*** is one of the earliest skills in learning to read. Students learn the speech sounds associated with each letter of the alphabet and then apply this knowledge to decode (or "sound out") new words. In the **letter sound identification** task **Exhibit 4**), students were presented with a grid of 100 letters in random order and asked to say out loud the sound that each letter represents. The **nonword decoding** task presented students with a grid of 50 nonsensical or pseudowords (nonwords) in random order. Nonwords are constructed from legitimate sound and spelling combinations in the target language but are not actual words; they are used to test the student's ability to apply letter sound knowledge to decode new words that they have never seen before. Over time, strong decoding skills and multiple exposures lead to automatic word recognition manifested by fluent reading.

Fluency is often defined as the ability to read with speed, accuracy, and understanding. Fluency is critical for comprehension, as rapid, effortless word recognition processes enable the reader to focus on the meaning of the text rather than on decoding words letter by letter.

In the **oral reading fluency** task, grade 3 students read aloud a grade 2-level passage of 64 words and grade 5 students read a grade 4-level passage of 103 words.

Reading comprehension is the ultimate goal of reading and refers to the ability to actively engage with and construct meaning from written text.

Exhibit 4. The Letter Sound Identification Task in Uzbek

a M i

d	K	i	z	o	L	s	ng	L	o
N	B	a	Y	M	K	b	R	A	E
u	Y	q	L	b	Z	Q	a	B	U
Sh	x	i	k	j	G	A	i	N	v
r	o'	A	O	H	L	N	p	U	i
A	i	n	h	O	e	i	a	n	f
O'	m	i	Q	T	n	h	O	T	A
L	y	G	N	D	A	o	A	i	s
Ng	H	O	i	t	g	a	d	S	g'
D	R	M	a	R	i	S	g	r	sh

- In the **oral reading comprehension** task, immediately after the student read the text for the oral reading fluency task, the assessor removed the text and orally posed five questions based on the text. The student answered each one orally. Four of the questions were direct, and one was inferential.
- In the **silent reading comprehension** task, students were given up to 4 minutes to read silently a grade 4-level text of approximately 180 words. Afterwards, the assessor orally posed 10 questions based on the text (eight direct questions and two inferential questions), and the student responded orally. The assessor did not remove the text from the student during the questioning, and the student was allowed to refer to the text if desired. Compared to the oral reading comprehension task, the silent reading comprehension task was based on a higher-level passage (grade 4 versus grade 2) that required the ability to process more complex semantic and syntactic relationships among text elements, as well as to read closely and retrieve details within a lengthier text. The task was purposefully constructed this way under the expectation that students in the upper grades should be able to comprehend longer, more complex text.

For the Program's EGRA baseline, students were all assessed in Uzbek, consistent with the ULA Program's design. The higher-order skills of fluency and comprehension build on lower-order skills of letter sound knowledge and decoding. The lower-order skills have been shown to be predictive of later reading achievement (Catts et al., 2001; Schatschneider et al., 2004). Therefore, even if a student cannot yet read a passage with comprehension, EGRA can nonetheless measure their progress toward acquiring the lower-order skills that are steps along the path to that end.

2.7.3 EGMA for Grade 3 and Grade 5 Students

The EGMA instrument was designed to collect information about basic mathematics competencies that grade 2 students should have mastered. The tasks used in a typical EGMA cover early numeracy skills that students need to progress academically. The numeracy skills and abilities demonstrated form a foundation for students to solve more advanced problems and facilitate the acquisition of more advanced mathematics skills in later grades.

The criteria for selecting tasks for a specific EGMA adaptation include, but are not limited to

- skills that are predictive of future academic success and proficiency in mathematics;

- skills that can be improved through classroom instruction; and
- skills that meet international standards of numeracy in the early grades.

The grade 3 EGMA was administered orally with individual students so that each student's ability to complete a task would not depend on their ability to read. In this way, mathematics skills were assessed independently of reading ability.

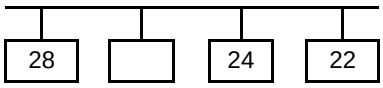
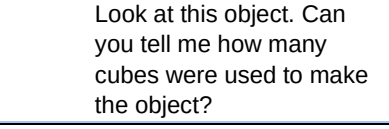
As noted earlier, although the purpose of the grade 5 Written Mathematics Assessment developed for Uzbekistan was the same as for the grade 2 EGMA, the grade 5 assessment took the form of a written group test rather than the standard one-on-one administration.

The tasks included in the grade 3 EGMA are described below, and examples of items in each task are provided in **Exhibit 5**.

These EGMA tasks were adapted for use in Uzbekistan by MoPE technical personnel and other stakeholders (see **Section 2.7.5**)—who constituted an EGMA working group—and were judged to be appropriate assessments for grade 3.

- **Missing number**—Students identified the missing number in 10-number patterns of increasing difficulty.
- **Addition level 2 and subtraction level 2**—Students solved five items of increasing difficulty for each of these tasks. Students were encouraged to use more advanced techniques to solve the problems rather than tallying or counting.
- **Word problems**—Students attempted six-word numeracy problems that increased in difficulty.
- **Relational reasoning**—Students attempted 10 problems that required them to think logically and find a solution using relationships between numbers rather than through sequential addition or subtraction.
- **3D spatial thinking**—Students attempted five problems that required them to manipulate 3D drawings mentally.

Exhibit 5. Examples of Grade 3 Mathematics Items by Task

Quantitative comparison	Which is greater: 623 or 632?
Missing number	
Addition	38 + 26 =
Subtraction	59 - 37 =
Word problem	There are 6 children in the class. 2 are boys. The rest are girls. How many girls are in the class?
Relational reasoning	28 = ☐ + 20 + 6
3D Spatial thinking	 Look at this object. Can you tell me how many cubes were used to make the object?

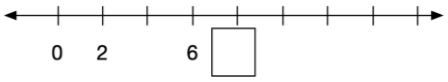
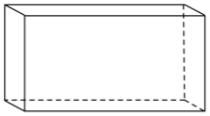
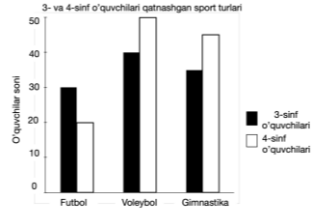
The grade 5 Written Mathematics Assessment was a group-administered test. It measured skills in several mathematics domains aligned to the grade 4 Trends in International

Mathematics and Science Study (TIMSS) Framework, a global mathematics assessment administered in grades 4 and 8. There were 31 items included in the grade 5 mathematics assessment:

- **Number and operations** items included place value, number patterns, addition, subtraction, multiplication, division expressions and word problems, comparing and ordering, fractions, and problems requiring algebraic thinking.
- **Geometry** items covered two-dimensional (2D) and 3D shapes and other geometric figures.
- **Measurement** items included volume, area, time, and estimation of measurement units.
- **Statistics** items asked students to read graphs and answer questions based on the information given. Items were either multiple choice or open answer.

Exhibit 5 provides illustrative examples of the grade 5 mathematics assessment by domain and number of items.

Exhibit 5. Examples of Grade 5 Mathematics Items by Domain

Domain	Number of Items	Example
Number and operations	18	Bo‘sh katakka to‘g‘ri keladigan sonni yozing. 
Geometry	5	Parallelepipedga qarang. Uning nechta **qirrasi** bor? _____ 
Measurement	4	Tomonlari 9 cm va 15 cm bo‘lgan to‘rtburchakning yuzasini toping.  _____
Statistics	4	Diagrammadan foydalanib, savollarga javob bering.  3-sinf o‘quvchilarining nechtasi voleybolga qatnashadi? _____

2.7.5 Tool Adaptation and Piloting

The ERMA/EGMA were adapted collaboratively with the MoPE during a workshop in October 2019. The assessment tasks were aligned with expected grade-level competencies. In November 2019, 21 MoPE staff and methodologists were trained on EGRA/EGMA survey administration. Following the training, teams were deployed to administer the survey in 70 pilot schools in seven regions, 10 schools per region. The ACR–Asia EGRA/EGMA Pilot Study Report was submitted in January 2020 and outlined recommendations to improve the instruments and training in preparation for the National Assessments and the Program's EGRA/EGMA baseline.

2.7.6 Instrument Reliability and Validity

Internal consistency is an appropriate and standard classical evaluation approach for cross-sectional data. Cronbach's alpha values for the mathematics tasks were 0.73 and 0.80 for grades 2 and 4, respectively. For the literacy tasks, grade 2 scored 0.78 and grade 4, 0.73. Cronbach's alpha should be at least 0.70 for adequacy, and coefficients closer to 1 indicate a good assessment.[2]

[2] Aron, A., Coups, E. J., & Aron, E. N. (2010). *Statistics for the behavioral and social sciences: A brief course* (5th edition). Pearson.

SECTION 3: MAIN RESULTS

3.1 GRADE 3 AND GRADE 5 EGRA FINDINGS

The estimates for student performance in Program schools are shown in **Exhibit *6*** It is important to restate that the students were assessed at the start of grades 3 and 5 to demonstrate the ULA and mathematics skills gained by the end of grades 2 and 4, respectively.

Exhibit 6. Reading Achievement by Task and Grade

Grade	Task	Treatment	Average
Grade 3	Letter sounds *(correct letters per minute [clpm])*	Program	93.1 [±2.0]
		comparison	91.7 [±3.0]
	Nonwords *(correct words per minute [cwpm])*	Program	38.9 [±1.3]
		comparison	38.0 [±1.8]
	Oral reading fluency *(cwpm)*	Program	47.9 [±1.9]
		comparison	49.5 [±2.9]
	Reading comprehension *(percent score)*	Program	79.1 [±2.1]
		comparison	80.3 [±2.4]
Grade 5	Nonwords *(cwpm)*	Program	47.2 [±1.6]
		comparison	48.5 [±1.9]
	Oral reading fluency *(cwpm)*	Program	68.3 [±2.4]
		comparison	71.0 [±2.9]
	Silent reading comprehension *(percent score)*	Program	64.8 [±2.0]
		comparison	68.2 [±2.3]

The results demonstrate similar outcomes for the Program and comparison groups. Students in grade 3 demonstrated particular mastery of letter sounds, with an average well over 90 clpm.

Exhibit 7 shows the distribution of correct letters per minute (clpm)among grade 3 students in Program schools. Sound recognition is one of the basic skills a student must have to eventually be able to read fluently and with comprehension. Most students demonstrated strong skills in letter sounds, with over 70% correctly sounding out at least 81 clpm. None of the students scored zero, and very few students' scores were between 1 and 40 clpm.

Exhibit 7. Distribution of Letter Sounds Scores (clpm), Grade 3

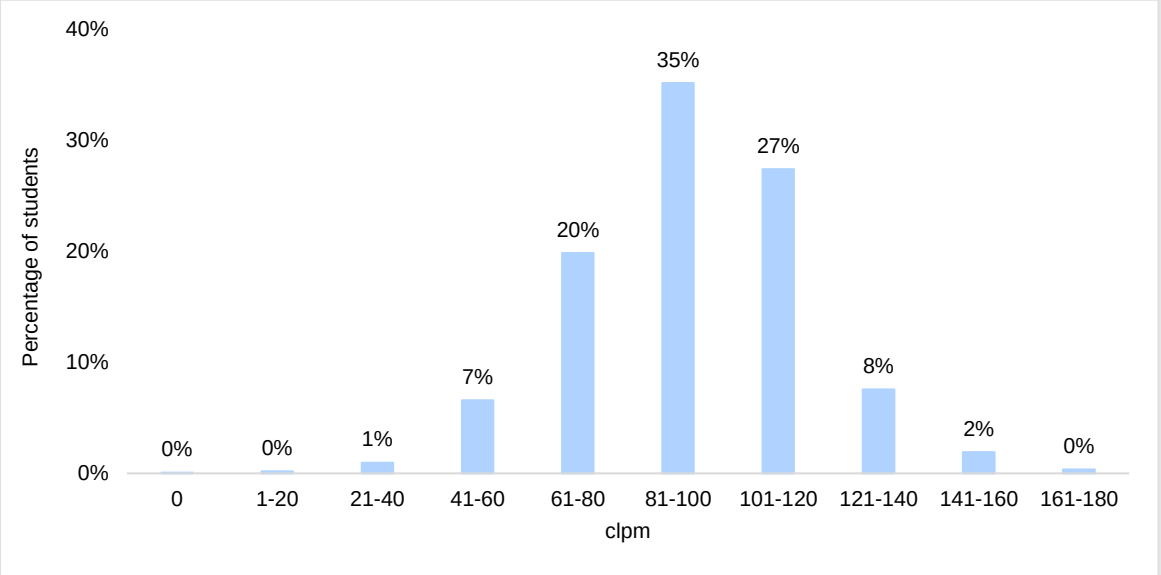
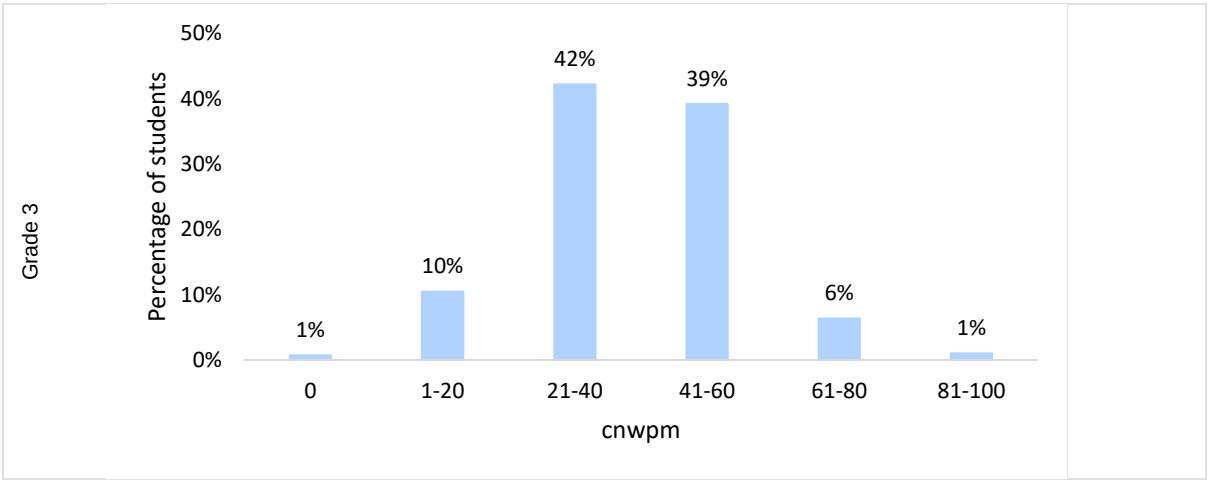
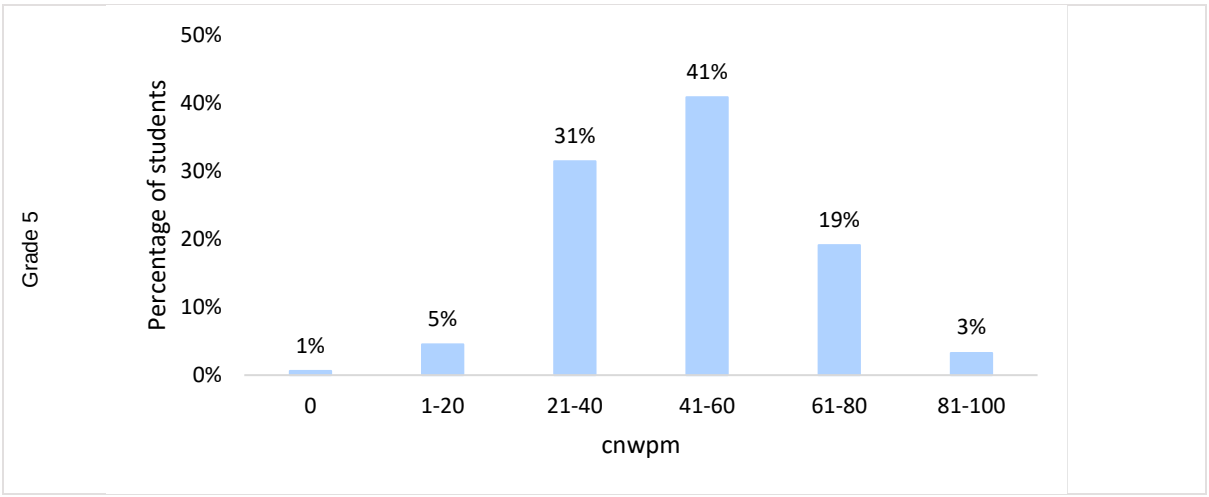


Exhibit 8 presents the distribution of scores for the nonwords task by grade. This task is a more accurate measurement of a student's ability to decode, as students read words that are not real in the language of the assessment (i.e., words they have not already stored in their vocabulary and learned to recognize by sight). The score distribution on the nonwords task indicates that more grade 5 students are progressing to higher ranges of reading, with 63% of them able to decode more than 40 correct nonwords per minute (cnwpm). Only 46% of students in grade 3 decoded more than 40 cnwpm.

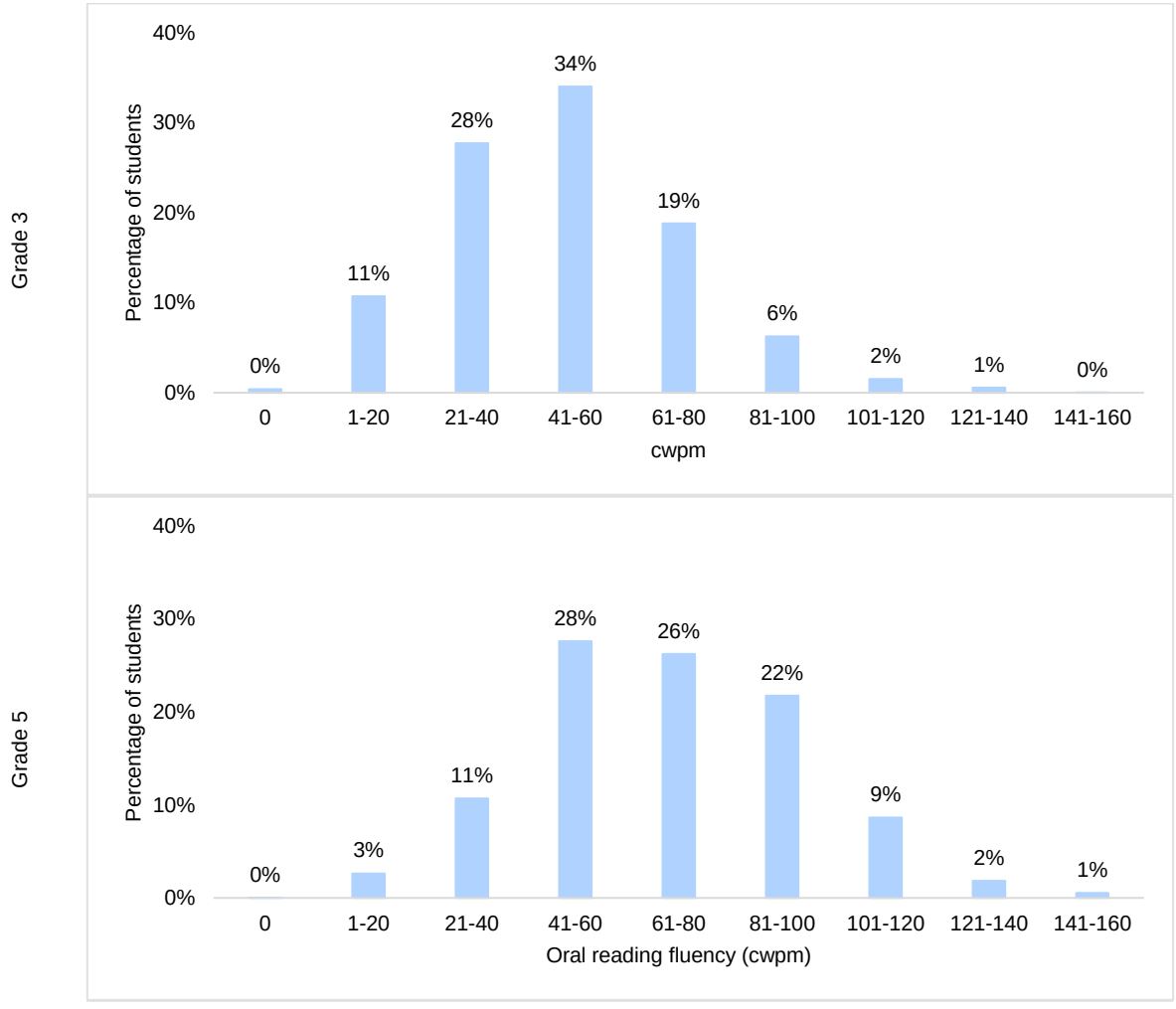
Exhibit 8. Distribution of Nonword Decoding Scores (cnwpm) by Grade





Students' oral reading fluency with connected text was measured using the oral passage reading task. Each grade 3 and grade 5 student was shown a grade 2-level passage and a grade 5-level passage, respectively, in Uzbek language and given 1 minute to read aloud as much of the passage as possible. The distributions of oral reading fluency scores among grade 3 and 5 students are presented in **Exhibit 9**.

Exhibit 9. Distribution of Oral Reading Fluency (cwpm), by Grade



As seen in **Exhibit 9**, none of the grade 3 students scored zero, and 62% read at least 41 cwpm (i.e., one word per 1.5 seconds). Oral reading fluency scores increased from grade 3 to grade 5, with a higher percentage (88%) of grade 5 students able to read at least 41 cwpm. Overall, these results demonstrate high reading achievement in Uzbekistan's schools.

Exhibit 10 presents the grade 3 score distribution for oral reading comprehension. The results indicate high levels of comprehension in grade 3. In fact, 72% of grade 3 students correctly answered 80% or more of the five Uzbek reading comprehension questions. Only 1% of students scored zero on the oral reading comprehension task.

Exhibit 10. Distribution of Reading Comprehension Scores, Grade 3

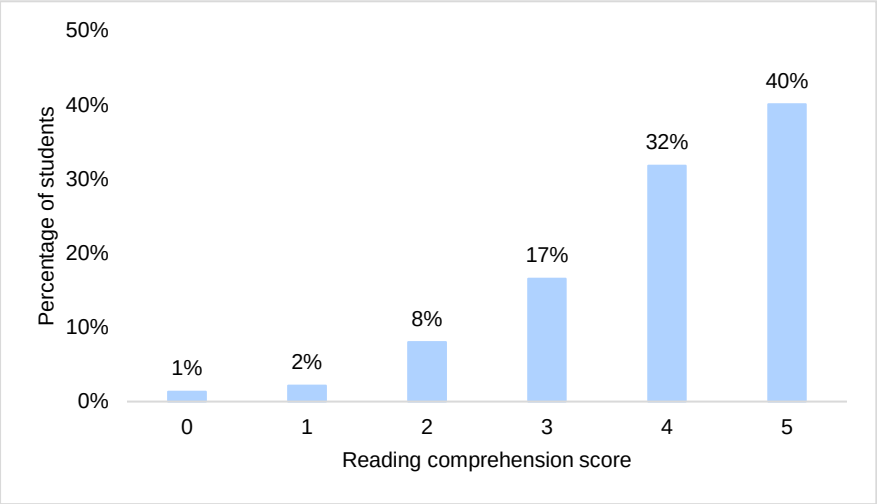
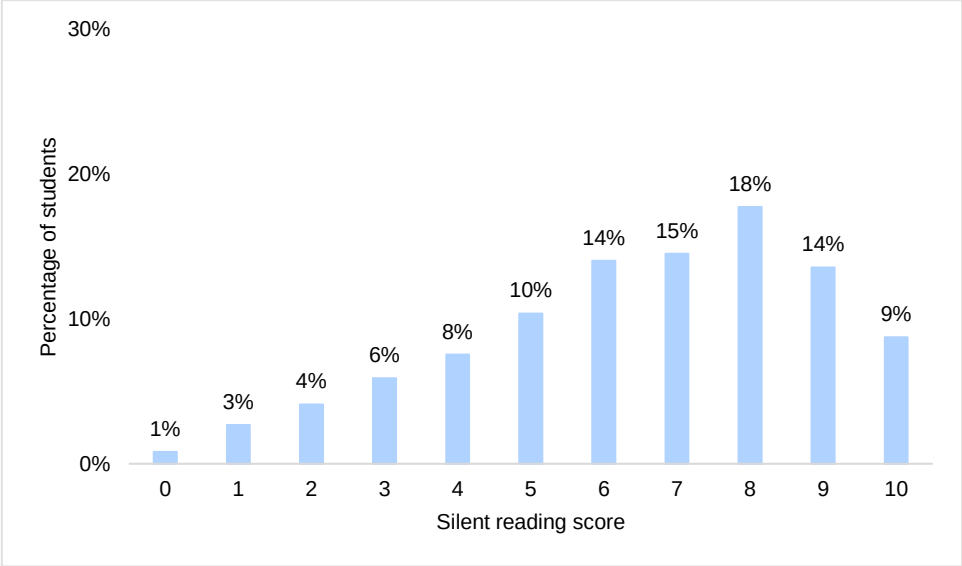


Exhibit 11 shows the distribution of silent reading comprehension scores among grade 5 students. After reading the silent reading passage, each student was verbally asked 10 questions related to the text they had read. The findings indicate that, overall, the assessed students' silent reading comprehension was quite low, with only 41% of students correctly answering at least eight of the 10 questions on the silent reading passage.

Exhibit 11. Distribution of Silent Reading Comprehension Scores, Grade 5



3.2 GRADE 3 AND GRADE 5 EGMA FINDINGS

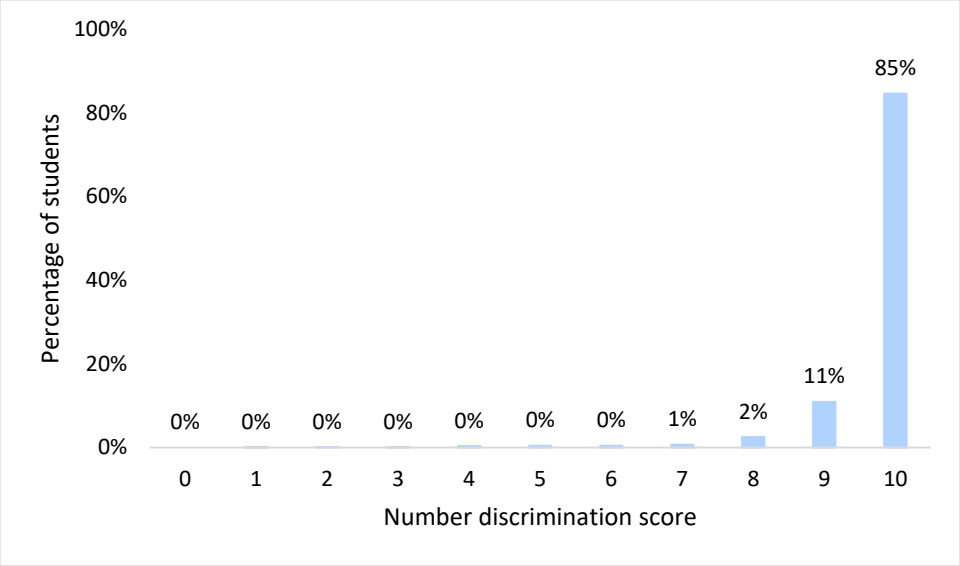
The grade 3 average EGMA scores at baseline are presented in **Exhibit 12**. Student performance was highest in the number discrimination task, with an overall average score of 97.5% in both Program and comparison schools. Students also performed well in the addition task, with average scores of 83.2% and 84.4% in Program and comparison schools, respectively. The lowest performance in mathematics was observed in relational reasoning, with average scores of 62.2% in Program schools and 61.0% in comparison schools.

Exhibit 12. Average Grade 3 Mathematics Achievement, by Task and Treatment

Grade	Task	Treatment	Average
Grade 3	Number discrimination *(percent score)*	Program	97.5 [±0.5]
		comparison	97.5 [±0.7]
	Missing number *(percent score)*	Program	67.9 [±2.0]
		comparison	68.2 [±2.9]
	Word problems *(percent score)*	Program	75.4 [±2.1]
		comparison	73.4 [±3.1]
	Addition *(percent score)*	Program	83.2 [±1.8]
		comparison	84.4 [±2.3]
	Subtraction *(percent score)*	Program	74.7 [±2.3]
		comparison	78.9 [±2.6]
	Relational reasoning *(percent score)*	Program	62.2 [±3.0]
		comparison	61.0 [±4.8]
	3D spatial thinking *(percent score)*	Program	62.8 [±2.2]
		comparison	65.4 [±3.3]

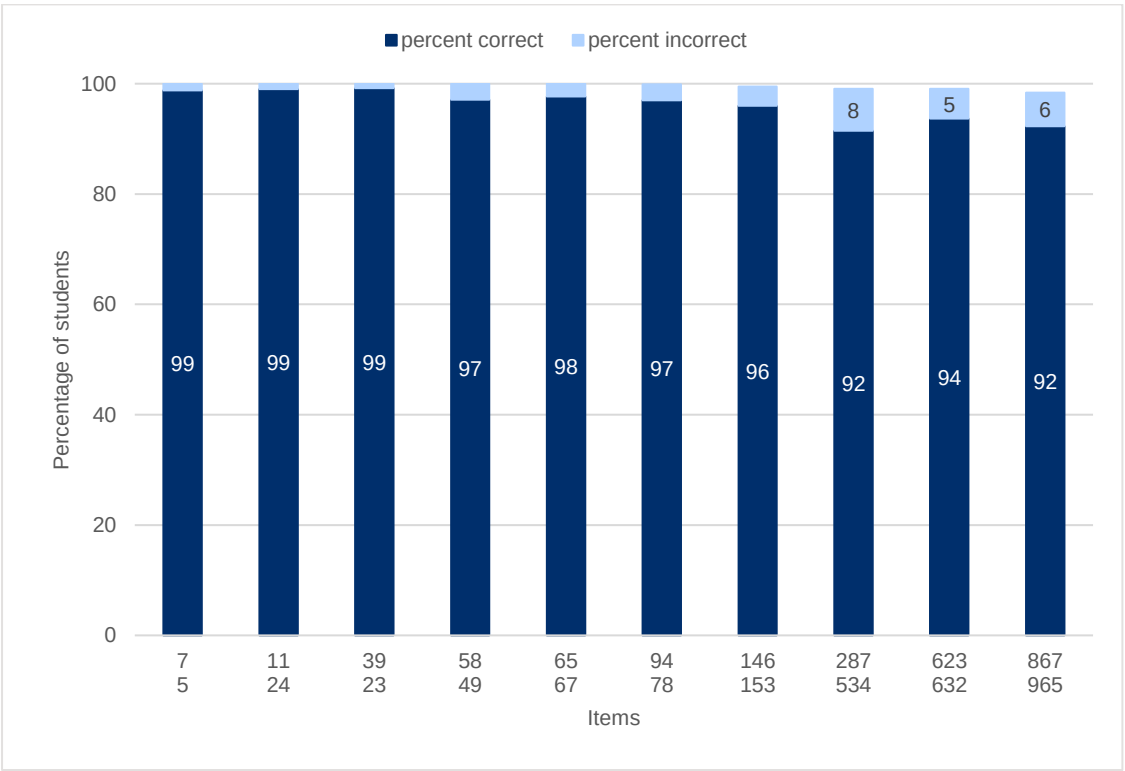
13 shows the distribution of scores for the quantitative comparison task among grade 3 students. This task required students to make a comparison and determine the larger value of a pair of numbers. Students were given 10 items that included a pair of single-digit numbers, five pairs of double-digit numbers, and four pairs of three-digit numbers. The score distribution shows that students performed very well on this task. None of the students got fewer than seven of the test items correct, and 85% correctly answered all items.

Exhibit 13. Distribution of Quantitative Comparison Scores, Grade 3



Findings from the item analysis for the quantitative comparison task are presented in **Exhibit 14**. Although students did well on all the items, they performed slightly better with single- and double-digit numbers (items 1–6) than with three-digit numbers.

Exhibit 14. Quantitative Comparison Scores by Item, Grade 3

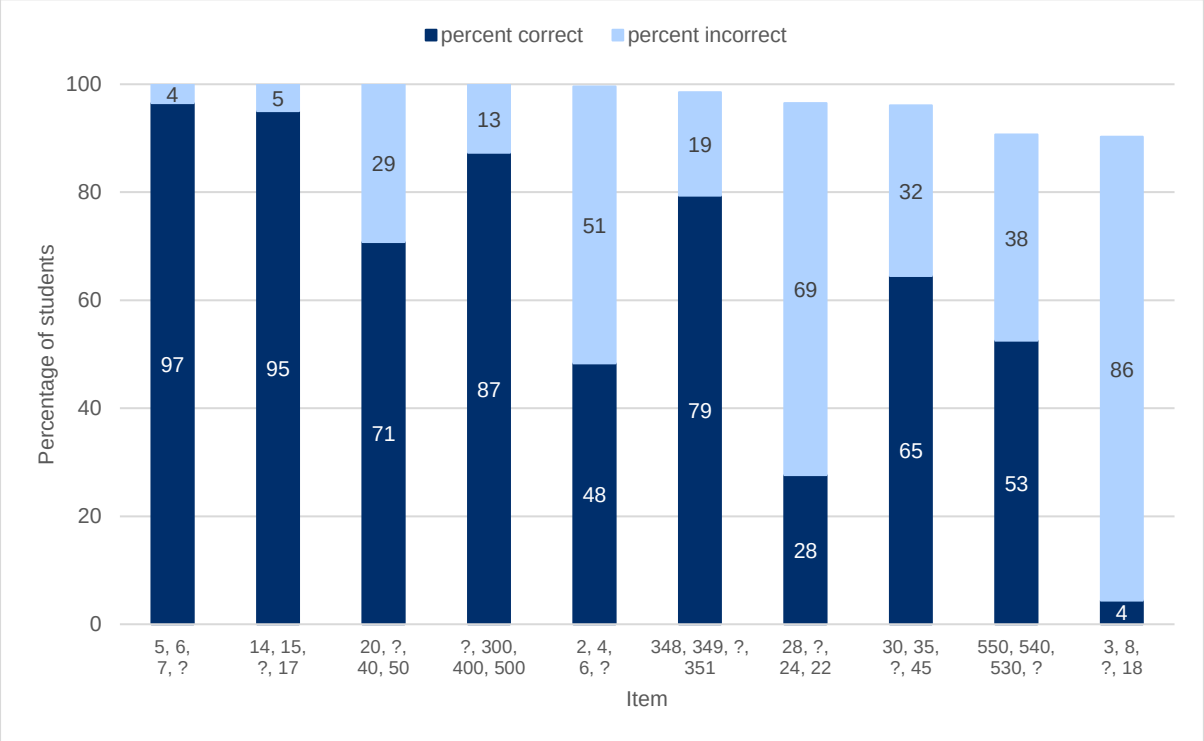


The missing number task was administered to students to assess their ability to recognize and complete number patterns. Student performance on this task was relatively low (**Exhibit 15**): 58% of students correctly answered at least seven out of the 10 items, but only 7% of students correctly answered all 10 questions.

Exhibit 15. Distribution of Missing Number Scores, Grade 3

Exhibit 16 presents the item scores for the missing number task. Students did well with patterns that involved counting forward by ones, fives, tens, and hundreds when presented in a logical sequence. However, most students struggled with patterns that involved counting forward or backward by twos (items 5 and 7) and patterns that involved counting forward by fives (item 10).

Exhibit 16. Missing Number Scores by Item, Grade 3



Students were also assessed on the word problem task (**Exhibit 17**) to measure their ability to interpret, plan, and solve problems. This task had six problems presented to the students as stories. Assessors read aloud each mathematics word problem (with small numerical

values) and asked the students to solve the problem. Students performed well on this task generally. The score distribution shows that 78% of students correctly solved at least four out of the six problems.

Exhibit 17. Distribution of Word Problem Scores, Grade 3

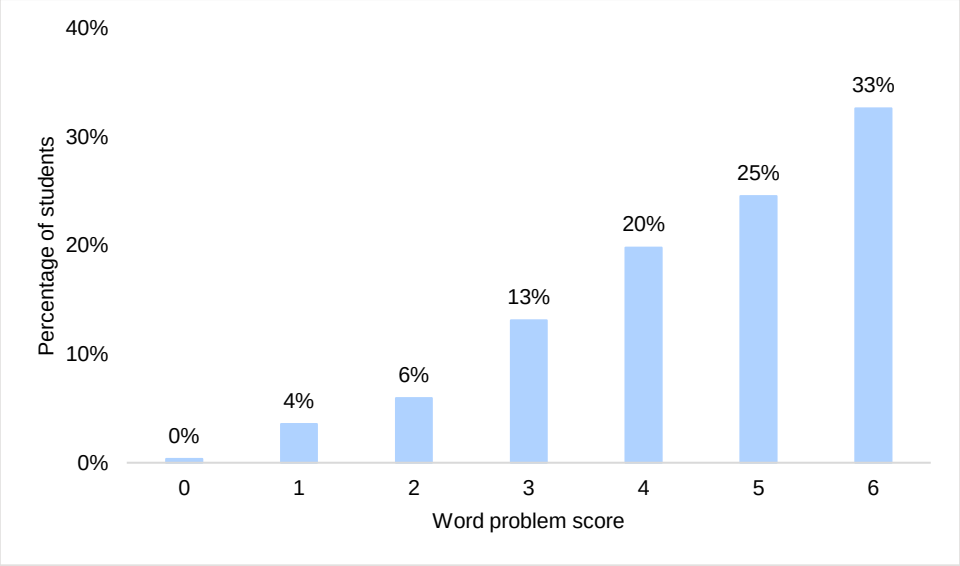
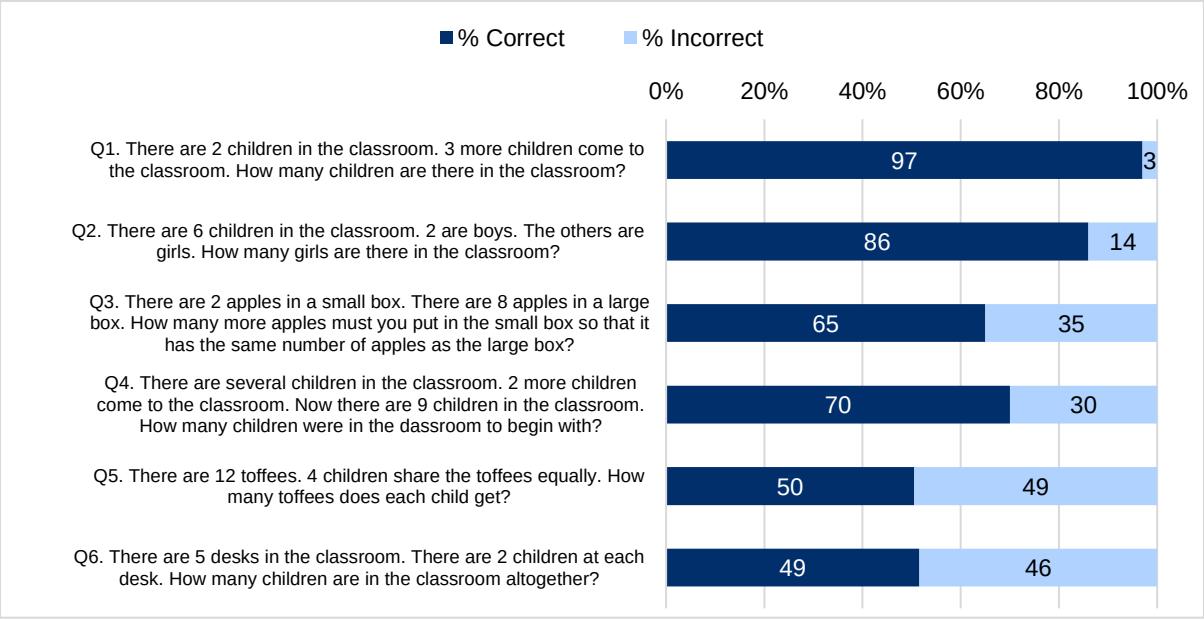


Exhibit 18 shows the item scores for the word problem task. Items in this task increased in difficulty as a student progressed through the assessment. As seen in **Exhibit 19**, students did well on problems that involved simple addition and subtraction (items 1 and 2, respectively). However, students struggled with items that required more critical interpretation (items 3 and 4) and items that required division or multiplication skills (items 5 and 6, respectively).

Exhibit 18. Word Problem Scores by Item, Grade 3



Students were also asked to solve addition and subtraction problems that involved understanding and applying basic addition and subtraction facts (**Exhibit 19**). The addition and subtraction tasks each included five items, all of which involved two-digit numbers.

Exhibit 19. Distributions of Addition and Subtraction Scores, Grade 3

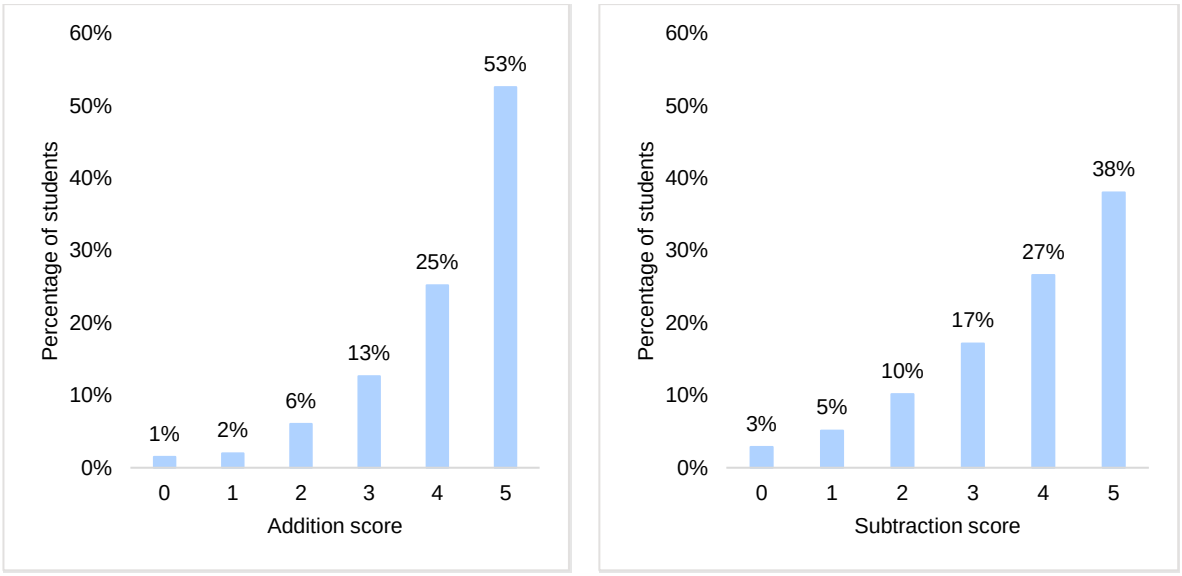


Exhibit 20 shows that students performed better on the addition task than the subtraction task, with 78% able to correctly answer at least four out of the five addition items compared to 65% who correctly answered at least four out of the five subtraction items. The item analysis in **Exhibit 21** indicates that students struggled with addition items that involved carrying tens (items 2 and 5) and subtraction items that involved borrowing tens (items 2 and 5). These findings imply that students have knowledge of basic addition and subtraction facts but cannot use this knowledge to solve more complicated addition and subtraction problems.

Exhibit 20. Addition and Subtraction Scores by Item, Grade 3

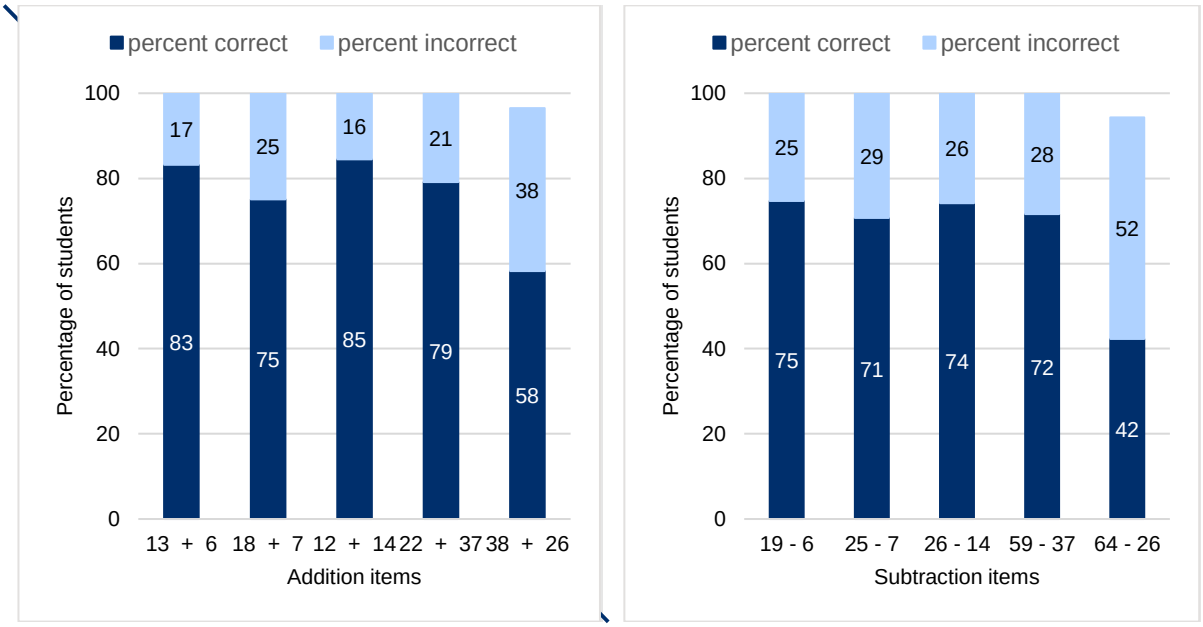
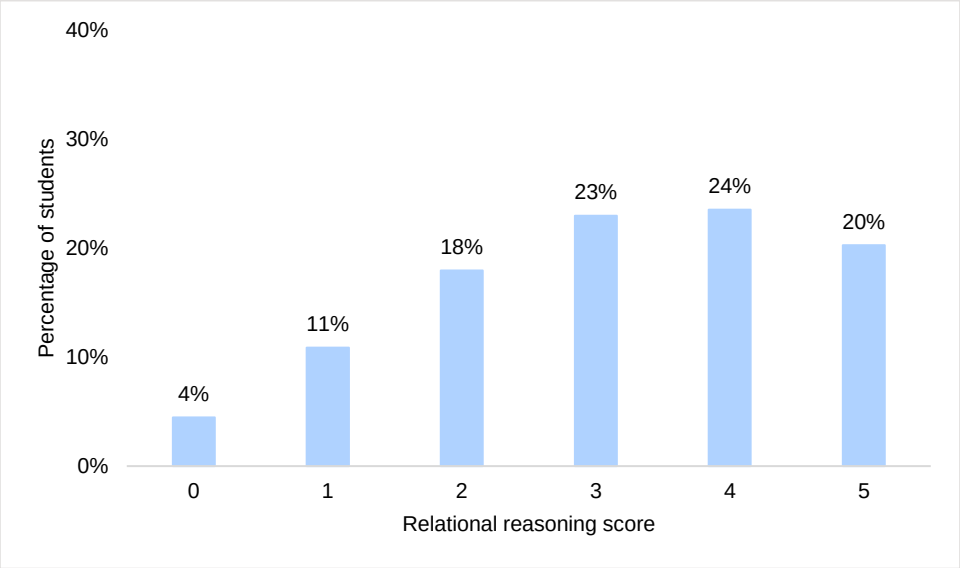


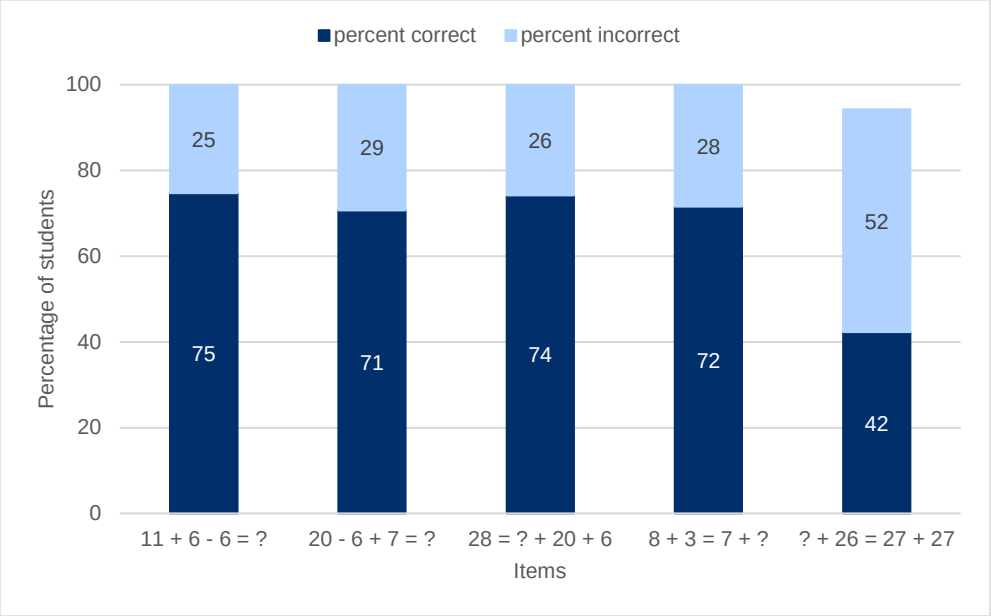
Exhibit 21 presents the distribution of grade 3 student scores on the relational reasoning task. This task measures students' abilities in numeric relational reasoning, which is one of the foundational and predictive early mathematics skills. In this task, students were required to determine the missing value in a numerical expression. Overall, students scored low on this task, with only 44% able to answer correctly at least four out of the five items.

Exhibit 21. Distribution of Relational Reasoning Scores, Grade 3



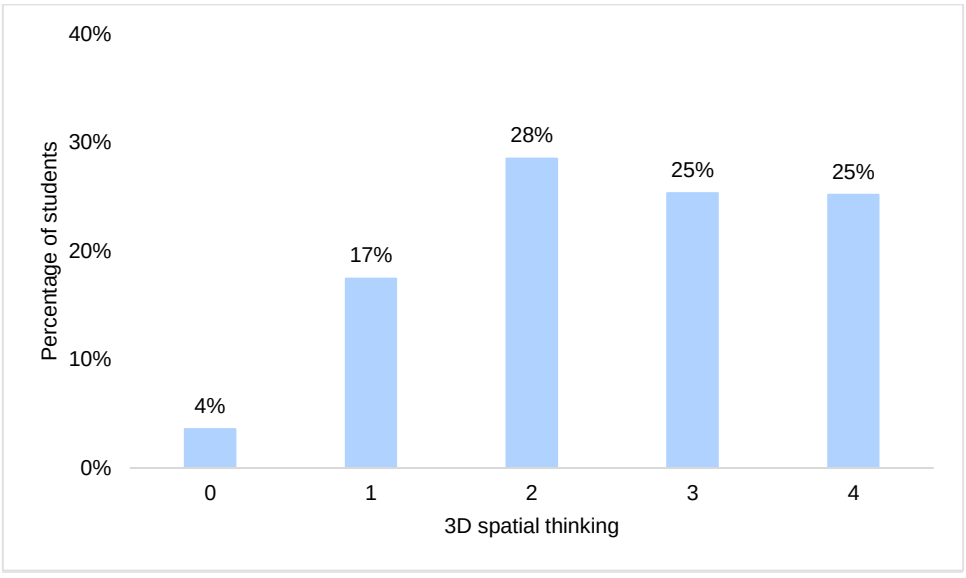
The item analysis for the relational reasoning task (**Exhibit 22)** indicates that students scored high on items for which they had to calculate (by addition or subtraction) the value of an expression presented in a familiar way (items 1 and 2) and the items for which they had to complete the decomposition of a number (item 3). Regarding equivalence expressions in which relationships were presented and the students had to determine the missing number needed to satisfy the relationship, students did well on the single-digit item (item 4), but many struggled with the two-digit item (item 5).

Exhibit 22. Relational Reasoning Scores by Item, Grade 3



Spatial thinking is another foundational skill in early mathematics. Students were tested on this skill to assess their ability to look at, recognize, and manipulate 3D shapes or objects. This task involved four pictures of 3D shapes assembled from cubes. Students were asked to identify how many cubes made up each picture. The distribution of scores in **Exhibit 23** shows that students scored low on this task, with only half (50%) getting at least three out of the four items correct.

Exhibit 23. Distribution of Spatial Thinking Scores, Grade 3



The item analysis (**Exhibit 24**) shows that students did well on items 1 and 2 of the 3D spatial thinking task. These items required simple visualization, as the cubes were wholly or partly visible. However, students scored very low on items 3 and 4, which required intricate visualization as some of the cubes were hidden from direct view. This meant that students had to count the cubes they could see and also mentally manipulate the pictures to know how many cubes were hidden from their view.

Exhibit 24. Spatial Thinking Scores by Item, Grade 3

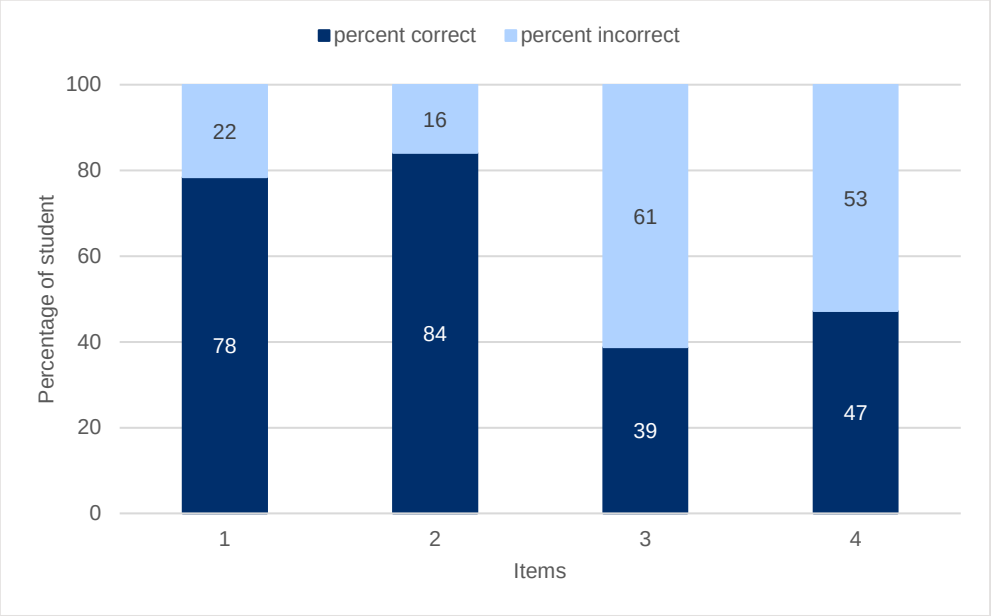


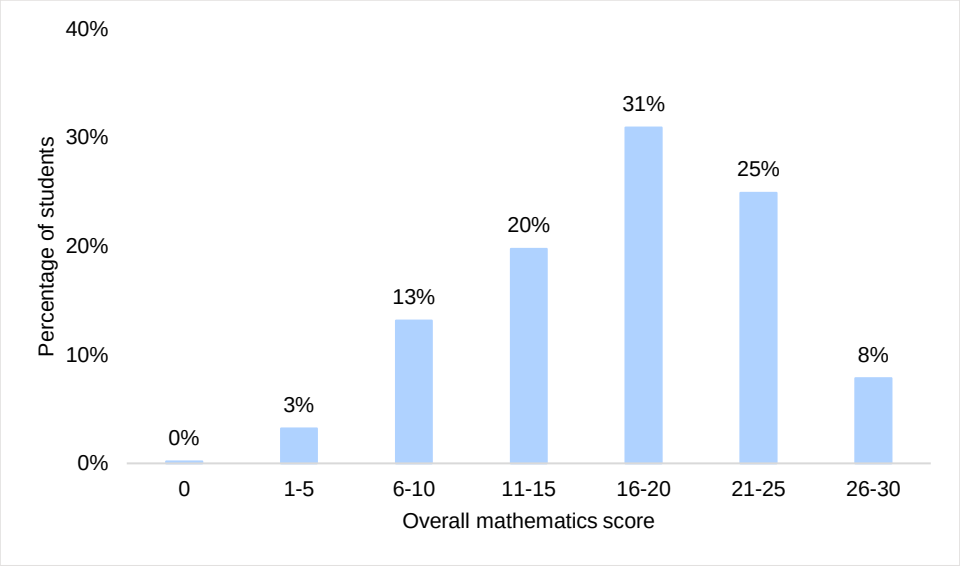
Exhibit 25 shows the average grade 5 mathematics scores by task and treatment. Student performance was strongest in the numbers and operations task, with 61.0% and 57.6% of items answered correctly, on average, in Program and comparison schools, respectively. The lowest scores were in the geometry task, where students at Program schools had an average score of 41.7%, and those from comparison schools had an average score of 40.2%.

Exhibit 25. Average Grade 5 Mathematics Achievement, by Task and Treatment

Grade	Task	Treatment	Average
Grade 5	Overall mathematics *(percent score)*	Program	57.4 [±2.5]
		comparison	53.6 [±2.3]
	Numbers and operations *(percent score)*	Program	61.0 [±2.5]
		comparison	57.6 [±2.4]
	Geometry *(percent score)*	Program	41.7 [±2.4]
		comparison	40.2 [±2.2]
	Measurement *(percent score)*	Program	53.5 [±3.3]
		comparison	46.1 [±3.1]
	Statistics *(percent score)*	Program	60.7 [±3.6]
		comparison	56.2 [±3.3]

The distribution of the overall mathematics scores for grade 5 is presented in **Exhibit 26**. The distribution of scores was left-skewed, indicating that most students (64%) received relatively high scores (i.e., they correctly answered more than half of the test items).

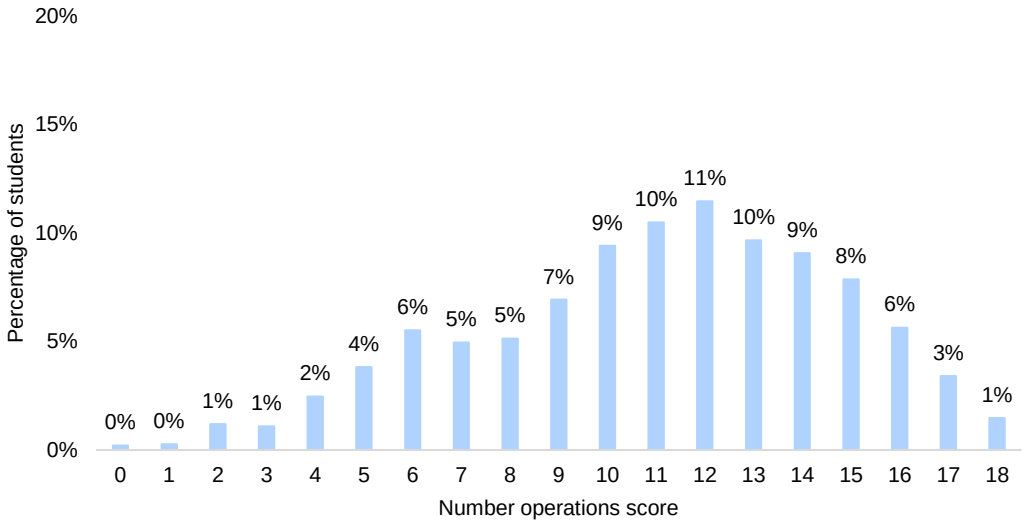
Exhibit 26. Distribution of Overall Mathematics Scores, Grade 5



The distributions of grade 5 student scores by domain are presented below (Exhibit 27–Exhibit 30).

Exhibit 27 shows the distribution of scores for the numbers and operations domain. This domain had a total of 18 items. As in the overall grade 5 mathematics score distribution presented in Exhibit 27, this domain had a left-skewed distribution, demonstrating that most students correctly answered more than half of the items . As seen in Exhibit 28, most students (67%) correctly answered at least 10 out of the 18 test items. None of the students got fewer than two of the test items correct. Overall, students performed well in this domain.

Exhibit 27. Distribution of Numbers and Operations Scores, Grade 5



The geometry domain tested students' knowledge of the features of 3D figures, names of geometric figures, and hierarchy of quadrilaterals. Students were required to count the number of edges and faces of a 3D rectangular prism, identify the name of a geometric figure (a 3D image of a pyramid), and identify the relationship between shapes (rectangles and squares). The distribution of scores in this domain (**Exhibit 28**) was right-skewed, indicating that most students achieved low scores. Specifically, 67% of students got only one

or two items correct out of four, and 12% answered none of the items correctly. Only 22% of students responded correctly to at least three out of the four items. Students struggled most in this domain.

Exhibit 28. Distribution of Geometry Scores, Grade 5

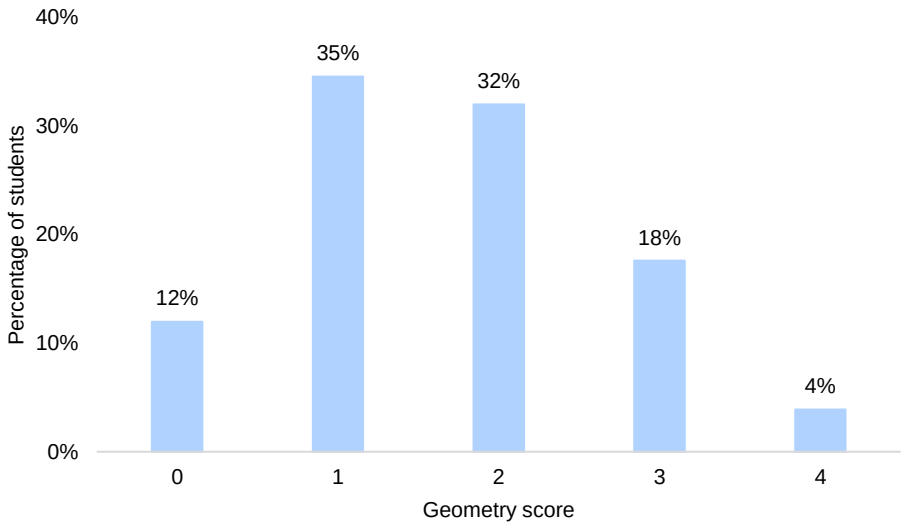
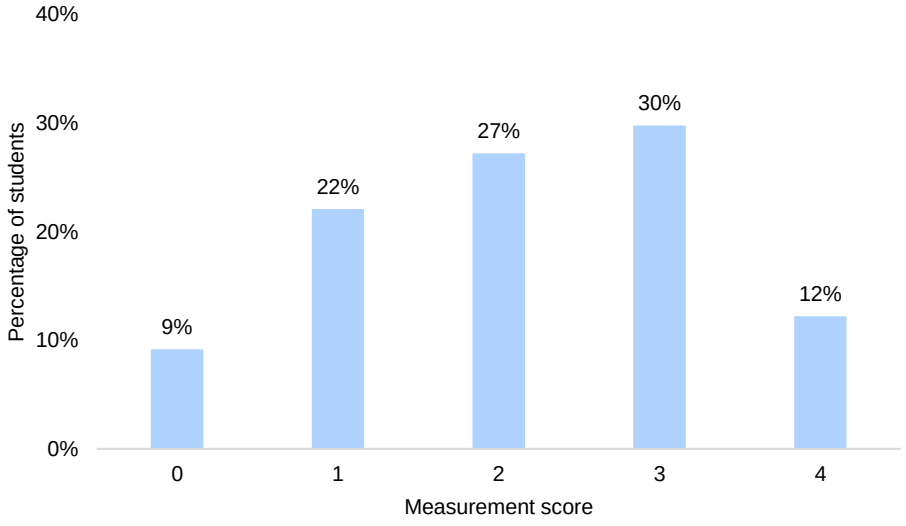


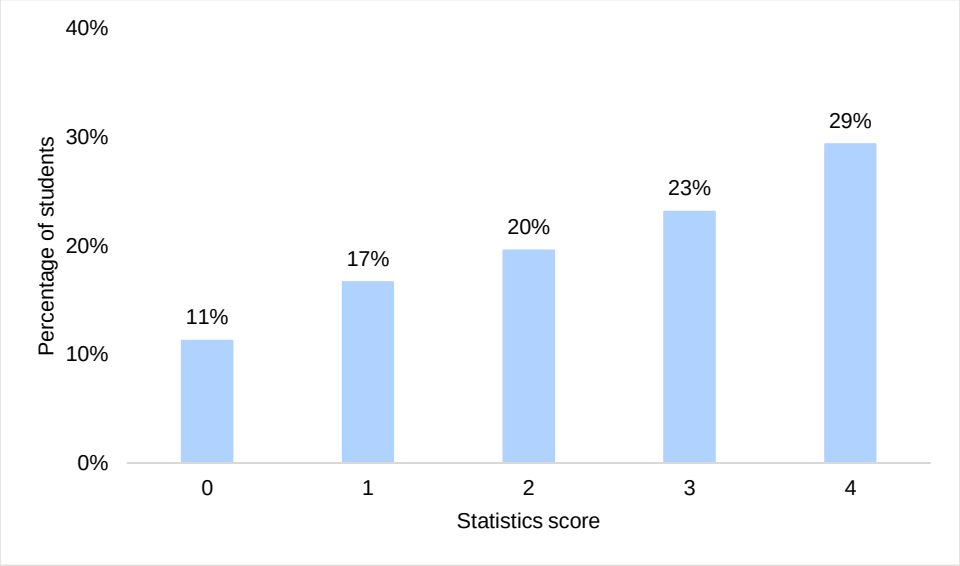
Exhibit 29 presents the distribution of scores in the measurement domain. The items in this domain required students to find the area of a rectangle with the width of one side and the length of one side given, find the perimeter of an irregular figure with right angles, identify the suitable unit of measure (centimeters, meters, or kilometers) for the height of an adult man, and solve a problem about the time duration of an activity with the starting and finishing time given on a clock. The distribution of scores in **Exhibit 29** shows that most students received low scores in this domain. Nearly half (49%) of students answered only one or two items correctly out of four, and 9% got none of the items correct. Forty-two percent of students responded correctly to at least three out of the four items. Students struggled in this domain but to a smaller degree compared to the geometry domain.

Exhibit 29. Distribution of Measurement Scores, Grade 5



The distribution of scores for the statistics domain is presented in **Exhibit 30**. This domain tested students' understanding of the categories and the values they represent on a bar graph. Student performance in this domain was relatively good. Slightly over half (52%) of students correctly answered at least three out of the four test items, and 37% got one or two items correct. Eleven percent of students did not respond correctly to any of the items.

Exhibit 30. Distribution of Statistics Scores, Grade 5



3.3 FINDINGS BY STUDENT GENDER

Exhibit 31 presents the EGRA results for grades 3 and 5 students by task and gender. In grade 3, girls outperformed boys in all tasks, and the differences were statistically significant (p<0.001), except for reading comprehension. The difference in performance between girls and boys in grade 3 was particularly substantial for oral reading fluency, with girls reading 10.6 more cwpm than boys. Grade 5 girls also performed significantly better than grade 5 boys on nonwords and oral reading fluency (p<0.001), and the difference was greatest for oral reading fluency, with girls reading 13.5 more cwpm than boys. For silent reading comprehension, boys outscored girls by 3.2 percentage points (p<0.01).

Exhibit 31. Average Literacy Achievement, by Grade, Task, and Student Gender

Grade	Task	Gender	Average	Difference
Grade 3	**Letter sounds** *(clpm)*	boys	90.2	girls +5.8***
		girls	96.0	
	Nonwords *(cwpm)*	boys	36.1	girls +5.6***
		girls	41.7	
	Oral reading fluency *(cwpm)*	boys	42.6	girls +10.6***
		girls	53.2	
	Reading comprehension *(percent score)*	boys	77.7	girls +2.8
		girls	80.5	
Grade 5	**Nonwords** *(cwpm)*	boys	44.5	girls +5.3***
		girls	49.8	
	Oral reading fluency *(cwpm)*	boys	61.5	girls +13.5***
		girls	75.0	
		boys	66.4	boys +3.2**

Exhibit 31. Average Literacy Achievement, by Grade, Task, and Student Gender

Grade	Task	Gender	Average	Difference
	Silent reading comprehension *(percent score)*	girls	63.2	

*p<0.05, ** p<0.01, *** p<0.001

Grades 3 and 5 EGMA performance by task and gender is highlighted in **Exhibit 32**. There were no statistically significant gender differences in student performance in mathematics in grade 3. Differences in scores between boys and girls were mild across tasks except for 3D spatial thinking, where boys outperformed girls by 4.3 percentage points. Grade 5 estimates show that overall, boys outperformed girls, with the grade 5 boys achieving an average score of 58.4% and the girls achieving an average score of 56.4% (p<0.05). The difference in performance between boys and girls in grade 5 was greatest in the measurement task, with a 3.4 percentage point variance in favor of boys.

Exhibit 32. Average Mathematics Achievement, by Grade, Task, and Student Gender

Grade	Task	Gender	Average	Difference
Grade 3	**Number discrimination** *(percent score)*	boys	97.8	boys +0.6
		girls	97.2	
	Missing number *(percent score)*	boys	67.7	girls +0.4
		girls	68.1	
	Word problems *(percent score)*	boys	76.6	boys +2.3
		girls	74.3	
	Addition *(percent score)*	boys	83.9	boys +1.4
		girls	82.5	
	Subtraction *(percent score)*	boys	74.7	girls +0.1
		girls	74.8	
	Relational reasoning *(percent score)*	boys	62.9	boys +1.4
		girls	61.5	
	3D spatial thinking *(percent score)*	boys	64.9	boys +4.3
		girls	60.6	
Grade 5	**Overall mathematics** *(percent score)*	boys	58.4	boys +2.0*
		girls	56.4	
	Numbers and operations *(percent score)*	boys	62.0	boys +2.0*
		girls	60.0	
	Geometry *(percent score)*	boys	42.1	boys +0.7
		girls	41.4	
	Measurement *(percent score)*	boys	55.2	boys +3.4*
		girls	51.8	
	Statistics *(percent score)*	boys	61.8	boys +2.3
		girls	59.5	

*p<0.05, ** p<0.01, *** p<0.001

3.5 FINDINGS BY STUDENT URBAN/RURAL CLASSIFICATION

Exhibit 33 shows that no difference was observed in grade 3 performance on the letter sounds and nonword decoding tasks between students in rural and urban schools. Grade 3 students in rural schools performed slightly better than those in urban schools on the oral reading fluency task (average score 48.5% [rural] vs. 46.8% [urban]). However, although students in rural schools outperformed students in urban schools in ORF, they did not outperform urban students in reading comprehension (average score 78.4% [rural] vs. 80.6% [urban). In all tasks, grade 5 students in urban schools outscored grade 5 students in rural schools, though the differences were not statistically significant.

Exhibit 33. Average Literacy Achievement, by Grade, Task, and Urbanicity

	Task	Urban Classification	Average
Grade 3	**Letter sounds** *(clpm)*	rural	93.4
		urban	92.6
	Nonwords *(cwpm)*	rural	38.7
		urban	39.4
	Oral reading fluency *(cwpm)*	rural	48.5
		urban	46.8
	Reading comprehension *(percent score)*	rural	78.4
		urban	80.6
Grade 5	**Nonwords** *(cwpm)*	rural	46.4
		urban	48.7
	Oral reading fluency *(cwpm)*	rural	67.2
		urban	70.3
	Silent reading comprehension *(percent score)*	rural	64.1
		urban	66.1

*p<0.05, ** p<0.01, *** p<0.001

Exhibit 34 highlights grades 3 and 5 EGMA performance by task in rural and urban schools. In both grades, the differences in average scores between rural and urban schools were relatively modest across all tasks, except for 3D spatial thinking, on which grade 3 students from urban schools performed better than grade 3 students from rural schools by 4.2 percentage points, and measurement, on which grade 5 students from rural schools outscored grade 5 students from urban schools by 5.6 percentage points. However, neither of these differences was statistically significant.

Exhibit 34. Average Mathematics Achievement, by Grade, Task, and Urbanicity

Grade	Task	Urban Classification	Average	Difference
Grade 3	**Number discrimination** *(percent score)*	rural	97.5	urban +0.1
		urban	97.5	
	Missing number *(percent score)*	rural	67.5	urban +1.2
		urban	68.7	
	Word problems *(percent score)*	rural	75.5	rural +0.2
		urban	75.3	

Exhibit 34. Average Mathematics Achievement, by Grade, Task, and Urbanicity

Grade	Task	Urban Classification	Average	Difference
	Addition *(percent score)*	rural	82.3	urban +2.6
		urban	84.9	
	Subtraction *(percent score)*	rural	74.6	urban +0.4
		urban	75.0	
	Relational reasoning *(percent score)*	rural	61.7	urban +1.3
		urban	63.0	
	3D spatial thinking *(percent score)*	rural	64.2	urban +4.2
		urban	60.0	
Grade 5	**Overall mathematics** *(percent score)*	rural	57.5	rural +0.4
		urban	57.1	
	Numbers and operations *(percent score)*	rural	60.5	urban +1.3
		urban	61.8	
	Geometry *(percent score)*	rural	42.6	rural +2.5
		urban	40.1	
	Measurement *(percent score)*	rural	55.5	rural +5.6
		urban	49.9	
	Statistics *(percent score)*	rural	60.9	rural +0.7
		urban	60.2	

*$p<0.05$, ** $p<0.01$, *** $p<0.001$

SECTION 4: CONCLUSIONS AND RECOMMENDATIONS

This section presents conclusions and recommendation based on the EGRA and EGMA findings.

4.1 EGRA

Most students have mastered the basic skills needed to eventually be able to read fluently and with comprehension in Uzbek. A vast majority of grade 3 students were found to be proficient in letter sounds. Students in grades 3 and 5 also demonstrated high level of skills in decoding words, suggesting that they can apply their knowledge of letter–sound relationships to correctly pronounce words. It is important for students to know how to read, but they also need to be able to comprehend what they are reading. The rate of grade-level comprehension was high in grade 3 but relatively low in grade 5. P

Results by gender show that girls performed significantly better than boys on oral reading fluency, reading 10.6 more cwpm than boys in grade 3 and 13.5 more cwpm than boys in grade 5. However, the comprehension scores were not substantially different between genders in either grade.

By urbanicity, grade 3 students in rural schools performed better than students in urban schools in ORF, but students in urban areas understood more of what they read than students in rural areas. Grade 5 students from urban schools outperformed grade 5 students from rural schools on all tasks, including nonwords, oral passage reading, and silent reading comprehension.

Recommendations:

- Teachers need continued support to implement an effective, evidence-based instructional methodology to strengthen the acquisition of basic reading skills and help students master higher-order skills (e.g., fluency, comprehension) as they advance to higher grades.
- The concept of fluency includes reading with speed, accuracy and understanding. Teacher's emphasis on speed reading in the traditional reading subject, will be expanded to fluency in Uzbek Language Arts. The Program should build on the speed-reading tradition by enhancing teacher's strategies towards attention to greater accuracy and understanding in ULA materials and the training.
- Program TPD activities should emphasize training teachers on the following topics:
  - **Reading fluency—**Teachers need training to carry out reading fluency activities during assessment weeks and revision lessons. This topic was not addressed in Program STBs, and in fact, the STBs contain no relevant activities.
  - **Reading comprehension—**During lessons, teachers must integrate techniques to improve reading comprehension (e.g., questioning, visualization, predicting, reciprocal teaching). In addition to the five questions given in the STB, teachers should be encouraged to ask more reading comprehension questions and provided with examples of such questions. Teachers should also teach students

strategies for working with texts, help them better understand the difference between open and inferential questions, and develop strategies for working with both types of questions.

- **Nonword reading—**The Program could develop examples of handouts with activities for nonword reading (e.g., jigsaw word reading[3], jumbled words[4], crosswords with nonwords[5]) and share on digital platform, in training or in the TGs should include these examples. ULA teachers should be encouraged to include nonword decoding activities in their lessons.

- Effective instruction should be complemented with appropriate TGs, STBs, and supplementary reading materials.

4.2 EGMA

Grade 3 and 5 students were found to be reasonably proficient on tasks that require the application of elementary mathematics skills but struggled to use these skills to solve more advanced and unfamiliar problems.

Specifically, grade 3 students demonstrated high proficiency in basic mathematics skills, such as comparing numbers and solving word problems and numeric expressions involving simple addition and subtraction. However, they scored low on relatively intricate items, such as word problems that required division and multiplication and numeric expressions that required more complex solving techniques (i.e., items that required carrying or borrowing). In addition, student performance was markedly poor on relational reasoning and 3D spatial thinking. These tasks comprised items that required students to apply deductive reasoning rather than elementary addition and subtraction.

Grade 5 students' performance in mathematics was moderate, with an overall average score of 57%. Performance was strongest in the numbers and operations domain. Most of the items in this domain required doing calculations, performing quantitative comparisons involving fractions, and solving algebraic expressions with single- and double-digit numbers. Students struggled most on geometry, followed by measurement. These domains mainly consisted of items that required the application of basic mathematics skills and problem-solving skills to more complex expressions.

Gender differences in mathematics performance were not statistically significant in grade 3, though boys scored higher than girls in the 3D spatial thinking domain. Remarkable differences in mathematics performance emerged in grade 5, where boys performed significantly better than girls overall.

There were no statistically significant differences in EGMA performance between rural and urban schools in grades 3 and 5. However, students in urban schools outperformed their counterparts in rural schools by a relatively large margin (4.2 percentage points) in grade 3 3D spatial thinking, while students in rural schools outscored those in urban schools by 5.6 percentage points in the grade 5 measurement task.

[3] Jigsaw Word Reading – a set of words broken into syllables which students need to reassemble to figure out all possible words.
[4] Jumbled words – a mixed set of letters that students need to use to restore the original word.
[5] Crosswords with nonwords – creating crosswords puzzles using nonsensical or pseudowords.

Recommendations:

Like all other subjects, mathematics is a subject that builds on itself, so teachers must ensure that students have a strong understanding of basic mathematics skills before they are moving on to higher-level concepts. If students are not supported to acquire solid foundational mathematics skills (e.g., addition, subtraction, multiplication, division, spatial reasoning), they are likely to struggle and fall behind with higher-level mathematics concepts.

- Students who are struggling with basic mathematics may need more time to practice and master new concepts, revisit earlier concepts they learned to solidify their knowledge, or be shown alternative approaches to solving mathematics problems.
- Teachers can also enhance students' abilities to apply basic skills to higher-level mathematics ideas through different techniques, such as:
  - Using mathematics manipulatives (e.g., sticks), which make numbers less abstract, in all grades.
  - Connecting mathematics concepts to students' daily lives to help students see the mathematics that is all around them.
  - Using games to practice new and past ideas.
  - Guiding students on how to sketch out word problems using daily life examples.
  - Discussing wrong answers with students.
- Program TPD activities should emphasize training teachers on strategies that can help students improve in the areas where students showed lower results (e.g., relational reasoning, 3D spatial thinking, geometry, and measurement).
- Because the STBs contain exercises designed for students with an average level of achievement, the Program could develop and publish a set of additional examples and tasks for students that need more support and more advanced students.

Overall, once the piloting of the Program STBs and TGs is complete, it will be important to consider revising the materials based on the above recommendations.

ANNEXES

ANNEX A: UZBEKISTAN EDUCATION FOR EXCELLENCE PROGRAM—IMPACT EVALUATION COMPARISON GROUP SELECTION

Summary

The Uzbekistan Education for Excellence Program (the Program) leveraged the Uzbekistan National Assessment data, which are regionally representative, to recommend a comparison region. The purpose of selecting a comparison region was to allow the Program's impact on learning outcomes to be compared against the business-as-usual results achieved in a comparison region. The difference between the outcomes in the Program regions and that in the comparison region would reflect the Program's value-added impact.

The Program recommends selecting Jizzakh Region as the comparison region for the following reasons:

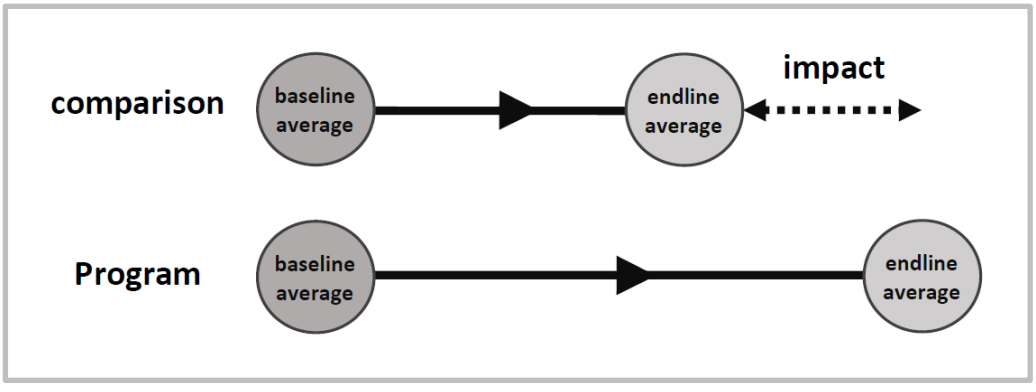
- The baseline average learning outcomes for Jizzakh Region and the Program regions are statistically similar.
- Jizzakh is adjacent to one of the Program regions (Sirdaryo).
- Jizzakh has a sufficient number of schools (over 500).

Why is a comparison region needed?

An impact evaluation measures Program and comparison groups at two (or more) time points, returning to the same schools and grades and sampling new students each time. This cross-sectional design measures the impact of improved teaching on student learning outcomes while maintaining the comparability between the Program and comparison areas.

A typical difference-in-differences analysis will be applied to measure impact. This is a calculation of the difference between the comparison and Program groups' average gains in learning outcomes, as shown in **Exhibit A-1** below.

Exhibit A-1. Impact Calculation



The evaluation is considered to be balanced if the comparison and Program averages are similar, such that we can be confident we are comparing apples to apples.

RTI follows the Institute of Education Sciences guidelines regarding what constitutes an acceptable difference between the Program group and the comparison group. The difference between the Program and comparison group baseline averages is calculated in terms of the

effect size.[6] We compare these differences against Institute of Education Sciences benchmarks of acceptable differences, which are shown in **Exhibit A-2**.

Exhibit A-2. Interpretation of Baseline Differences Between Comparison and Program Groups

	Effect Size Difference	Comment
	difference > 0.25	**unacceptable—**does not satisfy the baseline equivalence requirement
	0.05 ≤ difference ≤ 0.25	**acceptable—**requires statistical adjustment
	difference < 0.05	**acceptable—**satisfies the baseline equivalence requirement

Using **Exhibit A-2**, we can compare the difference between the baseline comparison and Program learning outcome differences, recognizing that a difference between the control and Program averages of less than 0.25 is acceptable.

The differences between the comparison and Program learning outcomes in the various region—accompanied by socioeconomic status data—are shown in **Exhibit A-3**. All statistics are presented relative to effect size differences against Program regions (Namangan and Sirdaryo).

Exhibit A-3. Effect Size Differences in Learning Outcomes and Socioeconomic Index Values Between Program regions and All Other Regions in Uzbekistan

Difference vs. Uzbekistan Program Regions	Oral Reading Fluency		Relational Reasoning (Grade 2)	Overall Mathematics Score (Grade 4)	Socioeconomic Index	
	Grade 2	Grade 4			Grade 2	Grade 4
Tashkent City	0.27	0.23	0.06	0.07	0.69	0.97
Andijan Region	0.26	0.23	0.30	0.33	0.17	0.07
Buxoro viloyati Bukhara Region	0.20	0.37	0.45	0.29	0.05	0.18
Farg'ona viloyati Fergana Region	0.00	0.02	0.45	0.02	0.06	0.02
Jizzakh Region	0.06	0.09	0.16	0.17	0.12	0.04
Navoiy Region	0.44	0.45	0.32	0.44	0.40	0.36
Qashqadaryo viloyati	0.18	0.42	0.12	0.04	0.25	0.06
Samarqand Region	0.05	0.31	0.63	0.11	0.04	0.13
Qashqadaryo Region	0.03	0.23	0.52	0.17	0.04	0.14
Tashkent Region	0.04	0.14	0.05	0.03	0.05	0.30
Xorazm Region	0.07	0.06	0.30	0.06	0.11	0.58
Karakalpakstan	0.14	0.07	0.08	0.12	0.07	0.35

According to the color definitions in **Exhibit A-2**, the only region with small and acceptable differences relative to the Program regions is Jizzakh Region, which had no difference greater than 0.17.

[6] Cohen, *Statistical Power Analysis for the Behavioral Sciences / Jacob Cohen.*