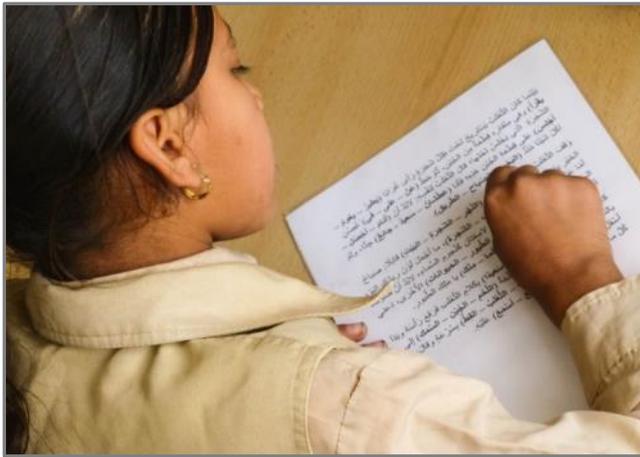




EdData II

Egypt Grade 3 Early Grade Reading Assessment Baseline



EdData II Technical and Managerial Assistance, Task Number 15
Contract Number: AID-278-BC-00019
May 2013

This publication was produced for review by the United States Agency for International Development. It was prepared by RTI International.

Egypt Grade 3 Early Grade Reading Assessment Baseline

EdData II
Task Order No. 15

Prepared for
USAID/Washington
Data for Education Programming in Asia and the Middle East (DEP-AME) contract
DRAFT FOR REVIEW AND COMMENT

Prepared by
Robert J. LaTowsky, Christopher Cummiskey, and Penelope Collins
RTI International
3040 Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194

RTI International is a trade name of Research Triangle Institute.

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

Table of Contents

Section	Page
List of Figures.....	v
List of Tables	vi
Abbreviations	vii
Acknowledgments	viii
Executive Summary	ix
Background.....	1
USAID Support for Early Grade Reading in Egypt.....	1
The MOE Early Grade Reading Program	1
Purpose and Design of the National EGRA Baseline for Grade 3.....	2
Why Test Early Grade Reading?	2
Previous EGRAs in Egypt	3
Purposes and Uses of this National Baseline for Grade 3	4
What EGRA Measures.....	5
EGRA Measures for the National Grade 3 Baseline	6
The Electronic and Paper EGRA Instruments.....	8
Implementing this National Baseline for Grade 3.....	9
Developing and Testing the Instrument.....	9
Training the Assessors	9
Field Implementation	10
National EGRA Baseline Findings for Grade 3.....	10
Summary Scores and Levels of Student Performance	10
Subtask Analysis.....	14
Letter Sound Knowledge	14
Nonword Reading.....	15
Oral Reading Fluency.....	16
Reading Comprehension.....	17
Listening Comprehension.....	18
Maze Reading Comprehension	20
Gender Differences in Reading Performance	21
Geographical Differences in Reading Performance	22
Rural/Urban Differences in Reading Performance.....	22
Reading Performance across Regions.....	23
Conclusions and Recommendations.....	25
Performance Comparison with Previous Egypt EGRAs.....	25
Nonreaders	26

Promote Technical Standards for EGRAs in Modern Standard Arabic for Specific Grades.....	27
Recommendations and Next Steps: Outcomes of the Policy Dialogue Workshop..	27
Annex A: The National EGRA Baseline Instrument for Grade 3	28
Annex B: Sample Design and Weighting.....	35
Annex C: Brief Reports of Previous Egypt EGRAs.....	40

List of Figures

Figure 1: Percentage of Grade 3 Students Performing at or above Benchmark on the EGRA Subtasks	12
Figure 2: Distribution of Grade 3 Students on the Letter Sound Knowledge Subtask	15
Figure 3: Distribution of Grade 3 Students on the Nonword Decoding Subtask... 16	
Figure 4: Distribution of Grade 3 Students on the Oral Reading Fluency Subtask	17
Figure 5: Distribution of Grade 3 Students on the Reading Comprehension Subtask	18
Figure 6: Distribution of Grade 3 Students on the Listening Comprehension Subtask	19
Figure 7: Distribution of Grade 3 Student on the Maze Comprehension Subtask	21
Figure 8: Literacy Achievement for Boys and Girls	22
Figure 9: Literacy Achievement for Students Who Attended Urban and Rural Schools	23
Figure 10: Regional Differences in Literacy Achievement for Students.....	24
Figure 11: Average Scores on Comparable Subtasks for all Egypt EGRAs	25

List of Tables

Table ES1: Summary Scores for All EGRA Subtasks.....	x
Table 1: Summary of Egypt EGRA Average Scores for Grade 3.....	11
Table 2: Summary of EGRA Scores for the Number of Items Attempted.....	13

Abbreviations

CCIMD	Center for Curriculum and Instructional Materials Development, MOE
clpm	correct letter sounds per minute
cnonwpm	correct nonwords per minute
cwpm	correct words per minute
EGLU	Early Grade Learning Unit, MOE
EGRA	Early Grade Reading Assessment
GILO	Girls' Improved Learning Outcomes Project, USAID Egypt
GOE	Government of Egypt
idara	district-level ministry administration, GOE
MOE	Ministry of Education
muderiya	governorate-level ministry administration, GOE
MSA	Modern Standard Arabic
NCEEE	National Center for Examinations and Education Evaluation, MOE
NGO	nongovernmental organization
ORF	oral reading fluency
PAT	Professional Academy of Teachers, Egypt
RTI	Research Triangle Institute, a trademark of RTI International
USAID	United States Agency for International Development

Acknowledgments

The authors wish to acknowledge with appreciation and respect the contributions and support of education staff from USAID/Egypt and Egypt’s Ministry of Education (MOE) in the implementation of this national Early Grade Reading Assessment (EGRA) baseline for Grade 3. At USAID/Egypt we thank Amena Chenzaie for leading the assessment launch, Hala ElSerafy for her skilled liaison with the MOE, and Mitch Kirby of USAID/Washington for his direction under the Data for Education Programming in Asia and the Middle East (DEP-AME) contract. Dr. Ali El-Alfi, Director of the Central Department for Basic Education, and Hanaa Qassem Hassanein, Head of the Early Reading Unit, at MOE were instrumental to EGRA’s successful implementation and to mobilizing ministry support at all levels, from start to finish. Dr. Emad al-Ba’aly, Advisor for International Cooperation at MOE, and Dr. Mohamed Ragab Fadlallah, Director of the National Center for Examinations and Education Evaluation, provided timely support in the planning and official approval for this national baseline. The success of this EGRA owes a debt to the more than fifty MOE staff who directly implemented the assessment as trained assessors and assessor team leaders and the Infonex consultants—Nahla Ibrahim Mohamed, Nagwa Ibrahim Hassan, Ayman Badr, Rania Refaat Daniel, and Eman Samir Fawzy—who capably managed and directed them in each of the five regions. Scores of school, idara, muderiya, and central ministry staff facilitated the field assessments in schools and the random pre-selection of sample Grade 3 students by providing complete enrollment lists for each school. The Arabic instrument for this EGRA was developed by Dr. Fathi al-Ashry and his contribution is appreciated. Adnan Al-Harazi, Osama M. Al-Fagih, and Hamdi Alweshah of Prodigy Systems, Inc., expertly supported the development and adaptation of the electronic instrument on iPads and final compilation of the database. Dr. Robert J. LaTowsky, Managing Director of Infonex and former Director for M&E, Communications and Reporting at the Girls’ Improved Learning Outcomes Project (GILO), was Field Director for this EGRA and a lead author of this report. Dr. Chris Cummiskey of RTI was responsible for data quality control, the calculation of scores, and statistical analysis of the data. Dr. Penelope Collins contributed expert input on the component skills of reading and report narrative. Dr. Samir Shafik and Ahmed Ismail of GILO ably assisted both the strategic planning and assessor training, and Sabry Mekhael managed procurement of the iPads and technical components. At RTI International, Medina Korda expertly served as the overall task leader and Michelle Ward-Brent backstopped and guided this assessment, from initiation to conclusion; David Harbin did the same for its financial and contractual management. We sincerely thank them all for their contributions to the success of this assessment.

Executive Summary

This report presents key findings of the first national baseline of early grade reading skills implemented in Egypt. The three previous Early Grade Reading Assessments (EGRAs) in 2009 and 2011 were regional samples for Grades 1 and 2. This EGRA for Grade 3 was implemented in March 2013 in a nationally representative sample of 200 schools: 40 Ministry of Education (MOE) primary schools selected randomly from each of five sub-national regions encompassing 25 of Egypt’s 27 governorates. The 1,992 tested students were randomly selected from Grade 3 enrollment lists prior to each school visit. Five boys and five girls were tested in each school. The results are representative nationally and for each of the five regions.

The findings of this national baseline inform the design and development of MOE’s Early Grade Reading Program, now implemented in Grades 1 and 2, to expand to Grade 3. Chiefly implemented by MOE Arabic language supervisors, this EGRA strengthened ministry capacities and deepened its knowledge base of early grade reading assessment and results accountability initiated by previous EGRAs.

This assessment of Grade 3 reading skills in the formal language of primary school instruction, Modern Standard Arabic, comprised 6 subtasks: the pre-reading skills of letter sound identification and nonword reading learned in Grades 1 and 2 plus oral reading fluency, and three comprehension subtasks: reading comprehension, listening comprehension, and Maze comprehension. Comprehension is a priority reading skill for Grade 3 instruction.

Key findings of this national Grade 3 baseline include:

- ***Low reading skills across all EGRA measures (Table ESI).*** Most Grade 3 students had limited pre-reading skills. Too few students could read with sufficient fluency to comprehend the texts. Less than 20% of students performed at or above proposed benchmarks for all subtasks. The sole exception: the pre-reading skill of letter sound knowledge (30% at or above benchmark). The average student scores on nearly all subtasks were less than half the proposed benchmark. Many students whom teachers observe reading quickly and smoothly are “mechanical” readers or “word-callers” who do not comprehend well what they are reading.

Table ES1: Summary Scores for All EGRA Subtasks

Subtask	Percentage of students with zero scores	Grade 3 average score	Proposed benchmark	Percentage of students performing at or above benchmark
Letter sound identification (clpm)	18.3%	18.8	27	30%
Nonword reading (cnonwpm)	27.4%	5.9	14	11%
Oral reading fluency (cwpm)	21.6%	21.9	45	16%
Reading comprehension (max. 6)	35.4%	1.9	5	9%
Listening comprehension (max. 7)	13.3%	3.2	6	18%
Maze comprehension (max. 14)	35.4%	3.6	12	7%

- Fourteen (14%) percent of Grade 3 students are nonreaders.*** By some measures, higher percentages (20-22%) are indicated. These are functionally illiterate students, not struggling readers. Nonreaders were unable to read correctly a single letter, nonword, or familiar word on any of the letter sound identification, nonword reading, *and* oral reading fluency (passage reading) subtasks. Their reading abilities are either zero or limited to the few words whose forms they have memorized and can recognize on sight. Soon to enter Grade 4, these nonreaders are unlikely ever to learn to read without remedial instruction. Reducing the percentage of nonreaders should be as essential a measure of improved reading performance as increasing average grade-level scores on specific subtasks.
- Urban/rural differences and gender differences in reading proficiency are not great.*** On average, urban students in Grade 3 out-perform their rural peers. But the difference is not great and contributing factors are multiple. There are strong readers and nonreaders in nearly all schools, often in similar proportion. There is a wide range of reading performance in most primary schools, both urban and rural. Similarly, the average scores for girls exceed those for boys on pre-reading skills and oral reading fluency. But there is little difference in comprehension skills. Further analysis of school and regional variance in EGRA results continues for presentation at the policy workshop planned for July 2013 in Cairo and inclusion in the Final Report.
- Comparing Grade 3 baseline reading performance with previous EGRAs in Egypt reveals the impact of improved reading instruction.*** Placed alongside the previous three EGRAs in Grades 1 and 2, the results show a natural progression in average reading proficiency from Grade 1 to Grade 2 to Grade 3. As one example, in oral reading fluency the progression in average scores begins at 1 correct word per minute in Grade 1, rising to 11 correct words in

Grade 2 (2009), and 22 correct words in this Grade 3. But when the Grade 2 EGRA in 2011—conducted just six months after the start of improved reading instruction under the GILO Early Grade Reading Program—is included in this comparison, *the marked impact of improved reading instruction* is immediately clear. Average scores for this Grade 2 (2011) equal or markedly exceed the Grade 3 baseline scores of students who had an additional year of schooling. And the percentage of nonreaders dropped more than 50% (see *Conclusions and Recommendations, Figure 11*). This comparison hints at the scale of reading improvement that can be expected from the planned extension of the MOE Early Grade Reading Program to Grade 3. Building on the Program’s achievements in Grades 1 and 2, significant improvement in Grade 3 reading is possible in a single school year.

Background

USAID Support for Early Grade Reading in Egypt

USAID support for improved classroom instruction in early grade reading began with the Girls' Improved Learning Outcomes (GILO) project in 2008. The first Early Grade Reading Assessment (EGRA) in Arabic was conducted in February 2009. The assessment findings prompted the Ministry of Education (MOE) to request USAID/Egypt support for improved reading outcomes. GILO immediately began the development of a pilot Early Grade Reading Program, beginning with Grades 1 and 2. From summer 2010, the project initiated extensive training of Grades 1 and 2 teachers and Arabic language supervisors, and orientation workshops for school principals, in ninety project-supported schools to apply enhanced classroom routines and new teaching resources for Arabic instruction.

Six months from the start of pilot implementation in Grade 2 classes, GILO conducted a second EGRA in spring 2011 to measure project net results from the EGRA baseline in 2009. The marked improvement in student reading skills in just six months of pilot implementation was embraced by MOE. GILO was asked to scale-up its pilot program in 90 schools and support the launch of MOE's national Early Grade Reading Program, starting with Grades 1 and 2, within four months. The MOE Early Grade Reading Program is now in its second year of implementation in all 16,000+ primary schools across Egypt with continued GILO support for planning, training of trainer cadres, follow-up, textbook analysis, and the development, production, and national dissemination of teacher and student learning resources.

With the GILO project scheduled to close in July 2013, USAID initiated implementation of a nationally-representative, baseline EGRA for Grade 3 in spring 2013 to empirically inform development of the MOE Early Grade Reading Program for Grade 3 and establish MOE capacities to conduct future EGRAs to measure results improvement. This report presents the findings of the Grade 3 national EGRA baseline.

The MOE Early Grade Reading Program

Beginning mid-2011, the MOE launched its program of early grade reading, modeled on the GILO Early Grade Reading Program, as a "national project" for improved Arabic reading skills and student learning outcomes in primary grades. An Early Reading Unit was created in the central ministry to coordinate and follow up Program implementation in all 27 MOE *mudariyas* across in Egypt and liaise with USAID/Egypt, the GILO project, and relevant MOE technical support centers including the National Center for Examinations and Educational Evaluation (NCEEE) and the Center for Curriculum and Instructional Materials Development (CCIMD).

The national MOE Early Grade Reading Program is currently implemented in Grades 1 and 2. Beginning fall 2012, the Education Support Program of USAID/Egypt initiated technical support to MOE for remedial reading instruction in Grades 4 and 5

as a pilot initiative in a few muderiyas. This national baseline EGRA for Grade 3 launches the expansion of the national MOE reading program into Grade 3.



MOE Early Grade Reading instruction in Grade 2, Giza

Strong support for improved early grade reading is apparent at all ministry levels. Novice and experienced teachers in Grades 1 and 2 welcome the scripted classroom routines and instructional materials for the clear guidance, tools, and resources they provide for daily Arabic reading instruction. Arabic language supervisors and school principals widely report improved classroom environments with students positively engaged, less disruptive, and attentive to learning. Results appear quickly: improved reading and learning outcomes are manifest within a few months. Muderiya Planning Teams comprised of senior officials in each MOE muderiya meet nationally every 4-6 months to coordinate, plan, and report on the Program's progress and challenges in their governorates. MOE is proud of its Early Grade Reading Program and the change catalyzed in early grade classrooms.

Purpose and Design of the National EGRA Baseline for Grade 3

Why Test Early Grade Reading?

The ability to read and understand a simple text is one of the most fundamental skills a child can learn. Without basic literacy there is little chance that a child can escape the intergenerational cycle of poverty. Yet in many countries, students enrolled in school for as many as six years are unable to read and understand a simple text.

Recent evidence indicates that learning to read both *early* and at a sufficient *rate* is essential for learning to read well. A substantial body of research documents the fact that children can learn to read by the end of Grade 3, and indeed need to be able to read to be successful in school. Acquiring literacy becomes more difficult as students grow older; children who do not learn to read in the early grades (Grades 1–3) are likely to fall behind in reading and other subjects, likely to repeat grades, and eventually to drop out of school.¹

When children are first learning to read in Arabic, they must learn the letters and their forms, learn the sounds associated with each letter and diacritical marks, and apply this knowledge to decode (or “sound out”) new words that they can recognize instantly.² By the end of this first phase, children develop sufficient speed and accuracy in decoding and word recognition that they can read with fluency. When children read with fluency, they can read orally with speed and expression similar to what they use in speech. Furthermore, reading with fluency is critical for reading comprehension, because children can concentrate on the meaning of what they read rather than having to focus on decoding.^{3,4}

Previous EGRAs in Egypt

This is the first national baseline of early grade reading skills implemented in Egypt. But it is the *fourth* Arabic EGRA conducted in cooperation with the MOE with financial support from USAID. The first three EGRAs were all conducted by the GILO Project with increasing MOE participation and capacity-building:

- **EGRA 2009:** The first EGRA conducted in Egypt was a baseline assessment implemented in February 2009 in Grades 2, 3, and 4 of fifty-nine primary schools of Fayoum, Minia, and Qena governorates. The 59 schools comprised 29 GILO-supported schools and 30 control schools in different idaras that would not be supported by the GILO Early Grade Reading Program. The results of this assessment of 2,800 students: i) presented MOE with empirical data on the reading performance of early grade students that catalyzed government priority for reading improvement and the Ministry’s request for GILO support to reading, ii) informed the development and implementation of the pilot GILO reading program, and ii) provided a baseline and control to later measure post-implementation results.
- **EGRA April 2011:** The second EGRA tested 1,200 Grade 2 students in the same GILO-supported and control schools tested for the 2009 baseline. This

¹ RTI. (2009). *Early Grade Reading Assessment Toolkit*. Research Triangle Park, NC: World Bank Office of Human Development, 1.

² See E. Saiegh-Haddad. (2005). Correlates of reading fluency in Arabic: Diglossic and orthographic factors. *Reading and Writing: An Interdisciplinary Journal*, 18, 559–582. See also M. Taouk & M. Coltheart. (2004). The cognitive processes involved in learning to read in Arabic. *Reading and Writing: An Interdisciplinary Journal*, 17, 27–57.

³ S. Abu-Rabia. (2007). The role of morphology and short vowelization in reading Arabic among normal and dyslexic readers in Grades 3, 6, 9, and 12. *Journal of Psycholinguistic Research*, 36, 89–106.

⁴ G. Elbeheri, J. Everatt, A. Mahfoudhi, M. A. Al-Diyar, & N. Taibah, (2011). Orthographic processing and reading comprehension among Arabic speaking mainstream and LD children. *Dyslexia*, 17(2): 123–142. doi: 10.1002/dys.430

post-implementation EGRA measured the improvement of reading skills of the Grade 2 students who received the improved reading instruction and teacher resources developed by GILO for Grades 1 and 2 in all project-supported schools. This EGRA compared the improvement of reading skills in these Grade 2 students with same-grade students in the control schools. A *Fact Sheet* on key results of this EGRA is provided in [Annex C](#).

- **EGRA October 2011:** The third EGRA tested 1,100 Grade 1 students in sixty MOE schools of Cairo and El-Beheira governorates. The purposes of this EGRA, conducted at MOE request, were: i) to demonstrate to MOE's satisfaction the general utility and baseline results for selected metropolitan Cairo and Lower Egypt idaras of using the same EGRA tool previously applied by GILO in 2009 and April 2011 in three Upper Egyptian governorates, and ii) to establish the institutional capacity of the MOE to conduct future EGRAs. A short *Brief* of key findings of this EGRA, and their comparison with EGRA 2009 baseline results for Grade 2, is included in [Annex C](#).

Purposes and Uses of this National Baseline for Grade 3

This national baseline EGRA for Grade 3 further strengthens the capacities and deepens the knowledge base of MOE staff to implement EGRAs. All EGRA assessors and assessor team leaders were MOE staff and all planning, training, implementation, and dissemination was conducted in close collaboration with the Ministry's new Early Grade Reading Unit in a leadership role. Central ministry and MOE muderiyas actively supported the field implementation with school liaison, orientation, and enrollment lists for sample selection.



The results inform policy decisions and planning by the Government of Egypt (GOE) for improved reading instruction and student learning outcomes. The longer term objective is enhanced teacher training for improved reading proficiency by Grade 3 students. The findings of this EGRA will infuse the design and development of MOE curricula and teaching resources for enhanced reading instruction in Grade 3. This baseline will also support future MOE accountability for the improved quality of student learning expected to follow the expansion of MOE's Early Grade Reading Program to Grade 3.

What EGRA Measures

The EGRA instrument is composed of various subtasks designed to assess foundational reading skills that are crucial to becoming a fluent reader. EGRA is a method-independent approach to assessment—that is, the instrument does not reflect a particular method of reading instruction (i.e., “whole language” or “phonics-based” approach). Rather, EGRA measures basic skills that a child must have to eventually be able to read fluently and with comprehension—the ultimate goal of reading. The EGRA subtasks are based on research for a comprehensive approach to reading acquisition across languages. These foundational reading skills are described below:

- The **alphabetic principle** is considered essential for learning to read an alphabetic language. The alphabetic principle refers to the recognition and understanding that speech sounds (phonemes) are represented by units of print such as letters and diacritics (graphemes). Thus, mastery of the alphabetic principle is the understanding that there are predictable relationships between sounds and the symbols that represent them. It is necessary for mastering spelling patterns and their relationship with oral language through the letter-sound (grapheme-phoneme) correspondences.
- **Oral reading fluency** is often defined as the ability to orally read connected text with speed, accuracy, and proper expression. Reading fluency is considered critical for comprehension, because rapid, effortless word-identification processes enable the reader to focus on the text and its meaning rather than decoding, or sounding out the words.⁵
- **Reading comprehension**, considered the goal of reading, refers to the ability to actively engage with, and construct meaning from, the texts that are read.
- **Listening comprehension** refers to one’s ability to make sense of oral language in the absence of print. Listening comprehension taps many skills and sources of knowledge, such as vocabulary knowledge, facility with grammar, and general background knowledge. Assessing listening comprehension is particularly important for a diglossic language such as Arabic, because children are often not introduced to the formal dialect until after they begin formal schooling. Thus, listening comprehension assesses children’s proficiency with the formal dialect of Arabic.

EGRA measures each of the above abilities/components to assess foundational reading skills. These skills are tested in individual subtasks and presented in order of increased level of difficulty (i.e., letter sound identification, then nonword reading, etc.). In general, initial subtasks are easier than later subtasks. The listening comprehension subtask is the exception: it best follows the reading comprehension subtask in implementation but is typically easier for students. EGRA thus effectively

⁵ National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (National Institutes of Health Publication No. 00-4769). Washington, DC: U.S. Government Printing Office. See also C.A. Perfetti. (1992). The representation problem in reading acquisition. In P.B. Gough, L.C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.

measures a wide range of reading abilities for beginning readers. The specific subtasks included in the EGRA instrument developed for this national Grade 3 baseline are described below.

EGRA Measures for the National Grade 3 Baseline

The EGRA instrument is implemented one-on-one—an assessor with a single student—and requires 15-20 minutes to complete. The Arabic instrument developed for this Grade 3 assessment is presented in *Annex A*. This instrument included the following subtasks (subtasks) implemented in this order:

1. **Letter sound identification** assessed children’s automaticity in their knowledge of the sounds associated with each letter. This was a timed subtask, in which children were shown a chart containing 100 letters with diacritics arranged in 10 rows each with 10 letters. Students were asked to produce the sounds associated with each letter as quickly and accurately as they could within one minute, yielding a score of correct letters per minute (clpm).
2. **Nonword reading** assessed children’s skill at applying letter-sound correspondence rules to decode (i.e., sound out) unfamiliar words. To ensure that children were applying their knowledge of the relationships between sounds and symbols rather than reading words from memory, children were asked to read a chart of 50 pronounceable made-up words (invented or nonsense words) with diacritics arranged in 10 rows of 5 words each. Children were asked to correctly sound out as many nonwords as they could within one minute, yielding a score of correct nonwords per minute (cnonwpm).
3. **Oral passage reading** assessed children’s fluency in reading a passage of grade-level text aloud and their ability to understand what they had read. This subtask consisted of two parts:
 - a. **Oral reading fluency:** The ability to read passages fluently is considered a necessary component for reading comprehension. In this subtask, children were given a 57-word story with diacritics shown for all words and were asked to read it aloud in one minute. Before starting, each child was instructed to pay attention to the story as they read because he or she would be asked questions about the story after finishing. The oral reading fluency score was the number of correct words read per minute (cwpm).
 - b. **Reading comprehension:** After the children finished the passage, or the one minute ended, the story was removed. The assessor then asked 6 questions that required children to either recall basic facts or infer information based on the passage or the part they read. All children that read more than the first line (8 words) of the story were asked all six questions.⁶ The reading comprehension score was the number of correct

⁶ Typically, students are asked only those reading comprehension questions that can be answered by the specific narrative in the passage they read. Weak and slow readers are not asked questions for narrative they did not

answers, with a maximum possible score of 6. This subtask was untimed but students who did not reply to a specific question within ten seconds were scored as “No Reply” on that question. Each question was asked only once with no repeat.

4. ***Listening comprehension*** is considered a critical skill for reading comprehension because it shows the ability to make sense of oral language. In this subtask, the examiner read clearly and at moderate pace (approximately 0.5 seconds per word) a short narrative story of 71 words to the children. Before starting, the assessor instructed each child to listen carefully as he or she would be asked several questions about the story. After hearing the passage, each child was asked all 7 questions, always in the same order and exactly as written in Modern Standard Arabic. The listening comprehension score was the total correct answers, with a maximum possible score of 7. This subtask was untimed but students who did not reply to a specific question within ten seconds were scored as “No Reply” on that question. Each question was asked only once.
5. ***Maze comprehension*** was a second measure of reading comprehension. In the Maze subtask, children were given a passage of some 140 words, without diacritics, to read aloud. On nearly every line of the passage, a single word was replaced with a multiple-choice selection of three words. All three words in each selection shared the same grammatical category (e.g., nouns, adjectives, verbs, etc.). For each selection, students were asked to identify the word that best fit the story. The passage included 14 of these multiple-choice word selections. This was a timed subtask and children were given three minutes to read the passage and select the best word for each selection. The Maze comprehension score was the total number of correct words selected, with a maximum possible score of 14. Children who read all three words in a selection without clearly selecting the best one were scored as “No Reply” for that selection.

All written components of the EGRA were in Modern Standard Arabic, including the stories and all questions in the reading comprehension and listening comprehension subtasks. All assessors asked the subtask questions and read the listening comprehension passage exactly as written in formal dialect without variance. The oral instructions given to children for each subtask, however, were explained by assessors in the home language of Egyptian dialect. These instructions were written on the instrument in formal Arabic but presented orally by the assessor, as written, in simple, vernacular Arabic. Children were asked to confirm that they fully understood the instructions before starting each subtask. Once started, no subtask was interrupted. The only comment permitted for an assessor to make was to say “go on”

read in the one minute. The design of this subtask, however, included 1-2 later questions that referenced early narrative and inferential understanding of the passage. To ensure consistent test implementation by all assessors, it was agreed that all reading comprehension questions would be asked of all students who successfully read beyond the first line of the passage. Most children had “no reply” to questions that referenced narrative content they did not reach. It was uncommon for students to guess answers.

after three seconds to a student stalled on a specific letter or word in one of the timed subtasks.

In administering the EGRA, assessors were very attentive to making each child feel comfortable and at ease. The child's name was not recorded and assessors presented the test as a "game" that the child would enjoy and an "experimental activity" to test the instrument's utility. Assessors were very explicit that the EGRA was not an exam and students were not being graded. Participating students were told they were lucky to have been



chosen for this experiment to test the instrument. Before beginning each assessment, children were pointedly asked for their assent to participate in the assessment. Any child who declined was thanked and invited to leave. Very few children refused to participate. At the end of the assessment and regardless of how well they read, the great majority of children responded, when asked if the assessment was difficult or easy, that it was "easy."

Many children, however, were nonreaders or limited readers for whom these subtasks were difficult and might be intimidating. For these students, the letter sound, nonword, and oral reading fluency subtasks each included an "early stop" rule that required assessors to discontinue the subtask if a child did not respond correctly to any of the items on the first line (i.e., the first 10 letters, the first 5 nonwords, or the first line of 8 words of the oral reading fluency story). This rule was established to avoid frustrating children who did not understand the subtask or lacked the reading skills to respond. If a subtask was halted by the "early stop" rule, the assessor went on the next subtask. If the oral reading fluency subtask, however, was halted by the "early stop" rule, the student was not asked any of the reading comprehension questions. All subtasks halted by the "early-stop" rule were marked clearly.

The "early stop" rule was applied in two ways on the Maze comprehension test. First, all students who were "early stopped" on *all* of the letter sound, nonword, and oral reading fluency subtasks were not given the Maze subtask. Second, any student who read the first 4 multiple-choice selections in the Maze story without making a clear choice of best word (i.e., "no reply"), was also halted by the "early stop" rule. The Maze subtask was not halted, however, if the child indicated a choice of word for any of the first 4 selections, whether that choice was correct or incorrect.

The Electronic and Paper EGRA Instruments

This EGRA was largely conducted using an electronic tool. After first developing the instrument in paper format, an identical electronic version was prepared for implementation using iPads. All subtasks were implemented in the same order with the same rules. As part of this implementation, a thorough and comparative field assessment of paper versus electronic instruments for EGRA implementation was

conducted. The findings of this comparative assessment will be presented in a separate report so as not to distract from the priority purposes and findings of this reading skills assessment.

Implementing this National Baseline for Grade 3

Developing and Testing the Instrument

The EGRA instrument was developed with Arabic language and reading experts in January 2013 and then tested, using a paper instrument, with 10 students in each of four MOE primary schools in Giza and Cairo governorates. MOE selected the schools and students for this pilot implementation. The only criteria for selection was that one school in each governorate be a “weak” school of lower learning achievement and the other school a “strong” school of superior student performance. Similarly, principals in each school were asked to provide an equal mix of weaker and more capable readers from Grade 3 to be tested. The purpose of this pilot—to test that the instrument was neither too easy nor too difficult and appropriate for the range of reading abilities in Grade 3—was explained at each school and school officials were supportive and complied fully. The pilot confirmed the effectiveness and reliability of the instrument in differentiating a wide range of reading abilities. The only modification recommended by the pilot was the addition of both a final sentence of narrative and one question for the listening comprehension subtask to make it less easy for capable readers. The pilot also tested two different stories for each of the oral reading fluency and Maze subtasks. One of the oral reading fluency stories was dropped, but both Maze stories were included in the final EGRA instrument, always implemented in the same order. The purpose of including two Maze stories was to see if the nature of the story or student familiarity with the subtask affected reading performance. It did not.

Development of the electronic EGRA tool proceeded in parallel with the testing of the paper instrument. The final paper instrument is included in [Annex A](#).

Training the Assessors

The Early Grade Reading Unit identified fifty candidates from MOE *muderiya*s across Egypt to be trained as EGRA assessors. The majority of candidates were Arabic language supervisors at the primary level. All candidates were trained for five days in late February with the expectation that 35 capable assessors would be selected. The first two days introduced the instrument and trained assessors to implement the paper form. The final three days trained assessors to use the electronic (iPad) instrument. All assessors were continuously monitored for proficiency in applying the instrument. All training was conducted in Arabic. On the final day of training, some forty percent of the candidate assessors were rejected as unable to satisfactorily use the electronic tool. In nearly all instances, the rejected assessors were inexperienced with touch-screen technology and its sensitivity. An additional 3 days of training was conducted the following week on the electronic instrument. This final training included additional assessor candidates to provide the minimum of 30

required assessors. An additional 15 assessor team leaders were also selected from MOE staff and trained for their role in ensuring the satisfactory implementation of EGRA in each sample school.

Field Implementation

The population and random, stratified sample of 200 schools and Grade 3 students selected for this national EGRA baseline are presented in **Annex B** along with the sample design

The 30 assessors and 15 assessor team leaders were organized into 15 assessor teams, each with two assessors and one assessor team leader. Three teams were assigned to each of 5 regions under the direction



of a Field Implementation Coordinator. Each team assessed ten students—5 boys and 5 girls—in a school each day. With each assessor testing 5 students, the school assessment typically required two hours. The assessor team leader was responsible for gathering and confirming the identity of the randomly preselected students for each school and delivering them, one by one, to the two assessors conducting the tests in the school library or classroom vacated for their use. Assessor team leaders ensured that students and assessors were not disturbed or interrupted.

The 15 teams completed 15 schools per day: three schools each day in each of the five regions. The assessment of all 200 schools was completed in three weeks in March.

National EGRA Baseline Findings for Grade 3

Summary Scores and Levels of Student Performance

This section presents summary statistics for all subtasks of the EGRA in Egypt. First, we present average scores and proposed benchmarks for each subtask and identify nonreaders.⁷ Then we look to the share of sampled students reading well, that is, at and above the proposed benchmark levels for each subtask, and the accuracy of their reading.⁸ **Table 1** below reveals that reading skills were low across all the EGRA measures. Few students could read with sufficient fluency to enable them to comprehend the text. Further, children had limited pre-reading skills. Students in

⁷ All average scores include zero scores in their calculation.

⁸ The benchmarks indicated for these subtasks are *proposed* for MOE technical review and decision. The specific benchmarks proposed here reflect: i) the precedent of analytical breakpoints applied in previous Egypt EGRA reports, ii) expert consensus that 80-85% correct items is the appropriate benchmark for comprehension subtasks, and iii) the actual distribution of scores from this Grade 3 baseline for nonword reading. Benchmarks were proposed here to best present results to GOE policy and decision makers and promote policy dialogue on reading performance benchmarks for specific grades.

Grade 3 could identify the sounds associated with about 19 letters on average in one minute. This is well below the proposed benchmark of 27 correct letter sounds per minute. Children’s limited mastery of the letter sounds contributed to scores that were, on average, less than half the proposed benchmarks for nonword decoding and oral reading fluency. Students correctly read an average of 6 nonwords and 22 real words per minute. Consequently, children’s reading comprehension scores were also low, on average 2 correct answers in six questions. Children showed better listening comprehension performance, with an average score of 3.2 correct answers in seven questions.

Table 1: Summary of Egypt EGRA Average Scores for Grade 3

Subtask	Percentage of students with zero scores	Grade 3 average score	Proposed benchmark	Percentage of students performing at or above benchmark
Letter sound identification (clpm)	18.3%	18.8	27	30.3%
Nonword reading (cnonwpm)	27.4%	5.9	14	10.9%
Oral reading fluency (cwpm)	21.6%	21.9	45	15.5%
Reading comprehension (max. 6)	35.4%	1.9	5	8.7%
Listening comprehension (max. 7)	13.3%	3.2	6	18.3%
Maze comprehension (max. 14)	35.4%	3.6	12	6.8%

Note: clpm = correct letter sounds per minute; cnonwpm = correct nonwords per minute; cwpm = correct words per minute. The percentage zero scores for Reading Comprehension and Maze Comprehension include “early stop” students (21.7% and 14.8% respectively) who were subsequently not given these subtasks.

Please note that examining children’s performance without considering how zero scores affect the overall average may not provide a clear picture of the reading achievement of children who do learn to read. Zero scores depress the overall average, and examination of the first column of *Table 1* suggests that the large number of zero scores likely had this effect. A little more than one quarter of the children in Grade 3 were unable to correctly read a single nonword on the first line of that test; 22% of Grade 3 children could not correctly read a single word from the first line of the passage.

In addition to considering the percentage of students who were unable to complete a single item on individual subtasks, there was a subgroup of students who were **nonreaders**. Nonreaders were students who scored zero on all three of the letter-sound identification, nonword reading, and oral reading fluency tests. These students could not read correctly a single word nor correctly identify a single letter sound on the first line of each test. Overall, 297 students or **14%** of all sampled Grade 3 students were nonreaders.

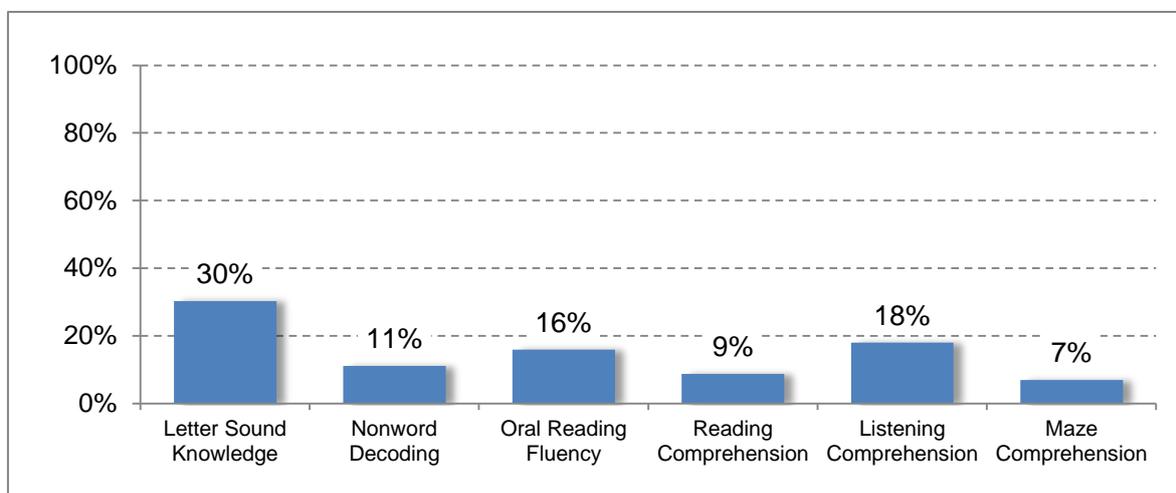
Letter sound identification and word decoding are reading skills taught in Grades 1 and 2. Reading comprehension is a priority reading skill to be learned in Grade 3. This assessment, conducted in the middle (March) of the second term of Grade 3, revealed low skills on all three tests of comprehension, i.e., the reading, listening, and Maze comprehension subtasks. One third (35.4%) of sampled students could not correctly answer a single reading comprehension question. Moreover, this percentage does not include the 21.6% of sampled students who could not correctly read a single word on the first line of the reading passage. These “zero” students were not tested for reading comprehension as they did not read the passage.

Listening comprehension results were markedly better: just 13% could not correctly answer a single listening comprehension question. On average, nearly half of the questions were answered correctly (3.2 correct of 7 questions). All sampled students, even nonreaders, were given the listening comprehension test. In an oral culture such as Egypt, it is not surprising that listening skills are stronger than reading comprehension skills. Greater concentration and understanding in listening is, however, desired.

Average performance on the Maze subtask was the lowest of all comprehension tests. This is the most difficult of the comprehension tests. But it is important to note that nonreaders were *not* given the Maze subtask. The low scores would be significantly lower if nonreaders had also been tested on Maze. Nevertheless, more than one-third of students (35.4%) who took the Maze test could not correctly answer a single item. The average score was just 3.6 of 14 total items (26%) answered correctly.

Although one may measure student achievement by average scores, it is also important to examine how many students are performing well. *Figure 1* reveals that few Grade 3 students have mastered the early reading skills. About one in ten children in Grade 3 were proficient at reading nonwords, and 16% were proficient at oral reading fluency. Answering reading comprehension questions was also problematic, as fewer than 10% of children were at or above benchmark on both the reading and Maze reading subtasks.

Figure 1: Percentage of Grade 3 Students Performing at or above Benchmark on the EGRA Subtasks



The analyses above first looked at average scores on subtasks and, second, at the percentage of students performing at or above the proposed benchmark for each subtask. A third way to analyze EGRA scores is to compare the results to the number of items *attempted* on the subtask. This is especially valuable for the analysis of reading comprehension and Maze comprehension subtasks, which were not given to “early stop” students. Moreover, most subtasks were timed and low scores also result from slow reading, even if the reading is accurate. This analysis is thus an examination of accuracy. Comparing scores to the number of items attempted on the subtask provides important insight into students’ mastery of early reading skills. Attempted scores are always higher than total scores.

Table 2 presents the average number of items attempted for each subtask and the average percentage of correct attempts. Children were most successful in letter identification and passage (familiar-word) reading. They correctly identified more than half (57%) of the letter sounds they attempted and 59% of the passage words that they read. Their success with attempted items in the other subtasks, however, was weak and low in accuracy. Even on the listening comprehension subtask, a task in which Egyptian children raised in a very oral culture might be expected to excel, fewer than half of the questions were answered correctly.

Table 2: Summary of EGRA Scores for the Number of Items Attempted

Subtask	Average number attempted	Percent correct
Letter sound identification (clpm)	26.5	57%
Nonword reading (cnonwpm)	13.2	35%
Oral reading fluency (cwpm)	27.3	59%
Reading comprehension (max. 6)	4.7	32%
Listening comprehension (max. 7)	7.0	44%
Maze comprehension (max. 14)	9.8	26%

Note: clpm = correct letters sounds per minute; cnonwpm = correct nonwords per minute; cwpm = correct words per minute.

Children had weak scores on all of the comprehension tests, accurately answering just 32% of the reading comprehension questions and 26% of the Maze questions they attempted. These results are no better than random guessing. On average, random guessing should answer these questions correctly one-third of the time.

Taken together, these results indicate that most Grade 3 students are struggling to recognize the sounds associated with each letter, to decode unfamiliar words, and to

recognize known words. Their low scores reflect both low accuracy and slow reading speed.

Subtask Analysis

In this section, we look at each subtask separately and analyze the range of student performance on each.

Letter Sound Knowledge

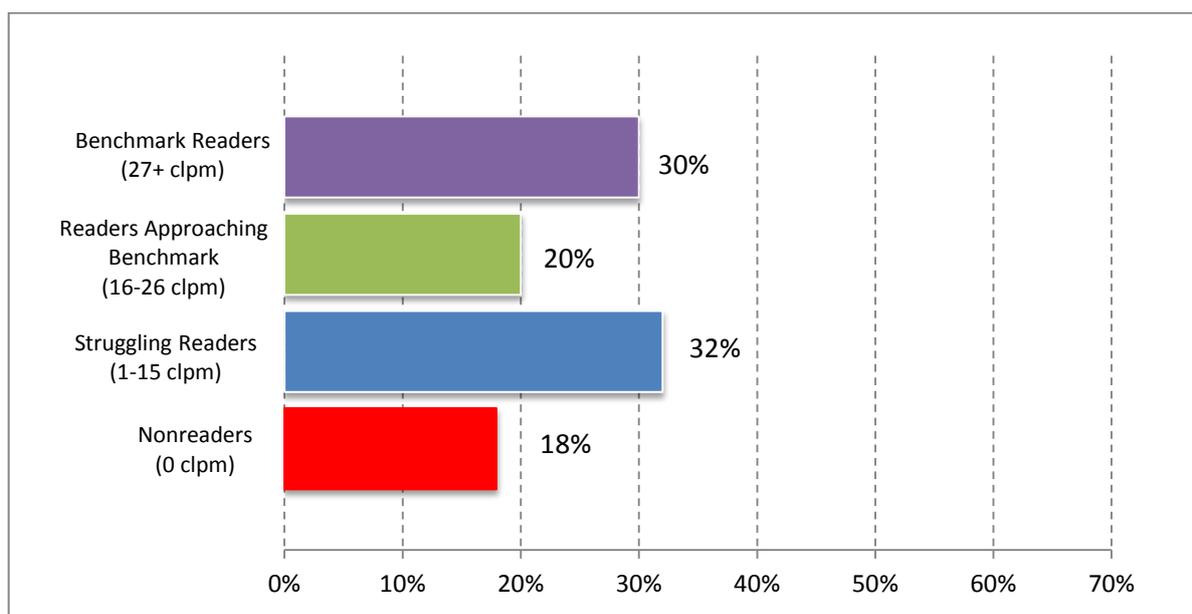
Letter sound knowledge is a basic reading skill taught in Grade 1. Letter sound knowledge, or the alphabetic principle, is considered a prerequisite skill for beginning reading and has been found to be a strong predictor of reading growth in abjads, or consonant-based alphabets, such as Arabic. Instruction in letter sounds was not, however, a systematic part of the reading curriculum for these students in their previous two years. The test of letter sound knowledge was included to appraise the skill levels of these Grade 3 students on this basic reading skill. Each student was presented with a chart of 100 random letters with diacritical marks. They were asked to pronounce the sounds associated with as many of these letters as they could within one minute. Scores for this subtask were the number of letter sounds the student could correctly pronounce within one minute (correct letters sounds per minute [clpm]).

Figure 2 presents the range of Grade 3 student performance on letter sound knowledge. Overall, 82% of tested students could identify at least one letter correctly; 18% are nonreaders and did not correctly pronounce any of the first ten letter sounds, which halted this subtask.

One third of Grade 3 students (32%) are still struggling with letter sounds. They could produce only 1 to 15 correct letter sounds in one minute. Almost as many students (30%), however, are proficient in this reading subtask, performing at or above the benchmark level of 27 correct letter sounds per minute. Between these high and low performers is a smaller share (20%) of “intermediate” readers who are approaching benchmark level. These students can correctly identify 16-26 letter sounds in one minute.

These results indicate a broad range of student performance on letter sound knowledge. There are large percentages of students at each level of skill, from proficient to nonreading. As we will see below, this wide variance in reading performance is commonly seen within the same classroom and is an acute challenge for teachers in Grade 3.

Figure 2: Distribution of Grade 3 Students on the Letter Sound Identification Subtask



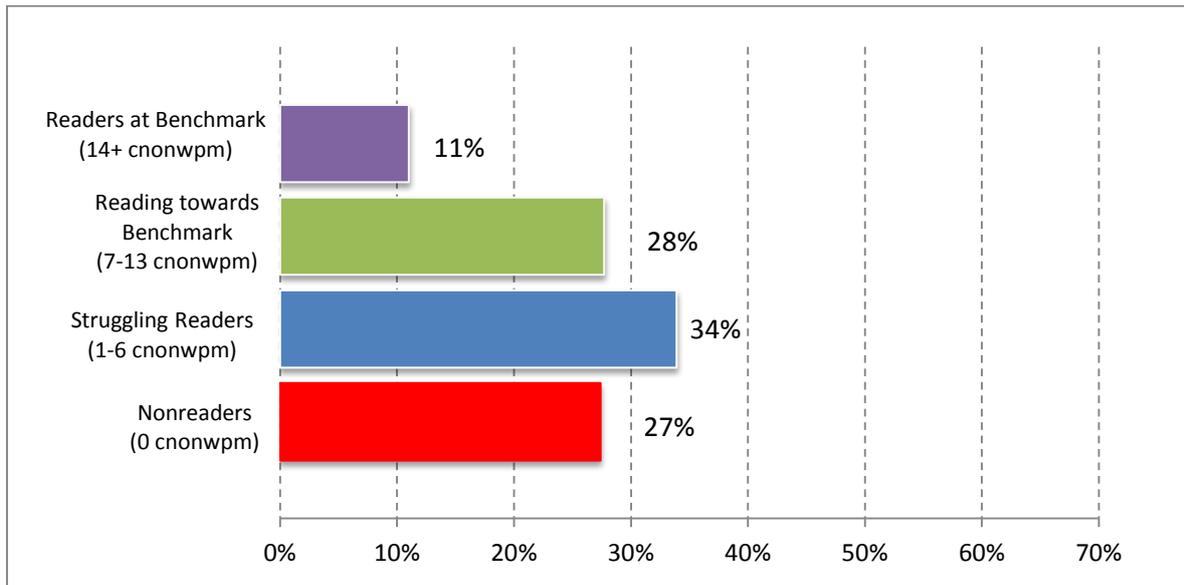
Nonword Reading

In the nonword subtask, students were presented with a chart containing 50 invented (or nonsense) words—most of them 3 or 4 letters—with diacritics and were asked to pronounce as many of the words as they possibly could within one minute. Skill in reading nonwords is a purer measure of decoding than using real words, because children cannot recognize the words by sight. Decoding is considered a self-teaching skill that enables children to read new and unfamiliar words independently.

Scores for this subtask were the number of words the student could correctly read within one minute (correct nonwords per minute or “cnonwpm”). The results presented in **Figure 3** show that reading nonsense words is considerably more difficult than reading passages with familiar words. Indeed, 27% of Grade 3 students were unable to decode a single nonword on the first line of the test. Only 11% of Grade 3 students were reading at or above the benchmark level of 14 correct nonwords per minute. But another 28% of students are approaching benchmark proficiency, correctly reading 7 to 13 nonwords per minute.

It is interesting to note that the percentage (34%) of students who struggled with this subtask—just 1-6 correct nonwords—is very similar to the percentage who struggled with letter sounds: 32%. This finding underscores the importance of enhanced instruction and classroom practice in letter sounds combined into words (and nonwords) with diacritics in Grades 1 and 2. Students completing the new MOE Early Grade Reading Program in Grades 1 and 2 this year would be expected to score significantly higher on these first two subtasks if they were tested in the spring of their third year.

Figure 3: Distribution of Grade 3 Students on the Nonword Reading Subtask



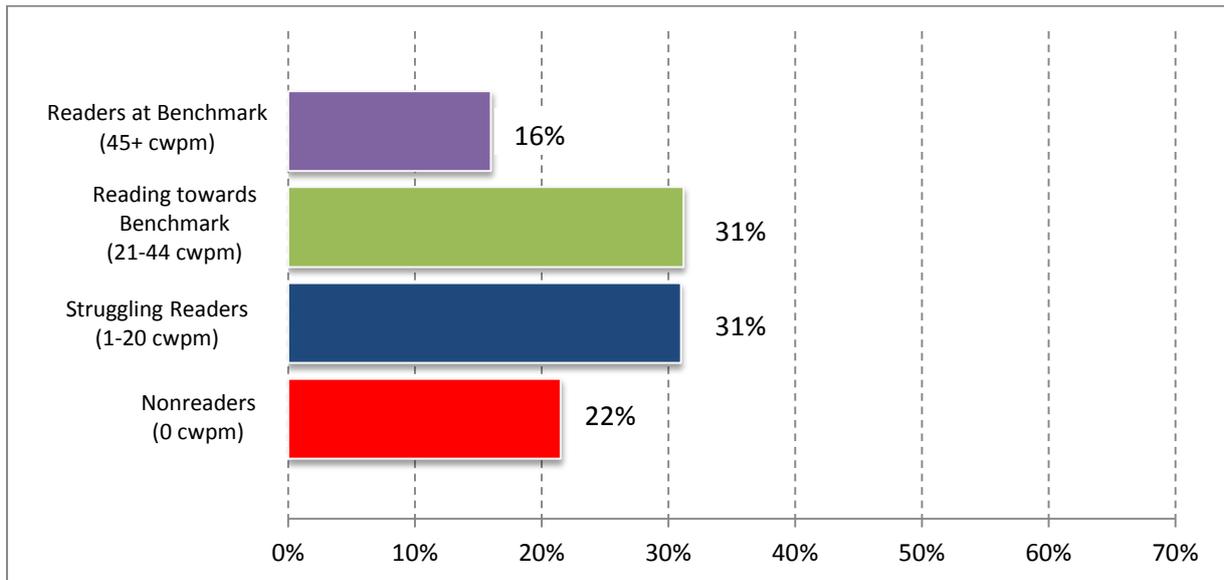
Oral Reading Fluency

In the oral reading fluency subtask, children were asked to read a narrative passage of local relevance within one minute. Oral reading fluency is a core index of reading competence, as it measures the skill and speed with which children translate letters into sounds, decode unfamiliar words, recognize known words, and simultaneously make sense of the text’s meaning. Weakness in any one of these processes can slow or disrupt children’s reading fluency. The score for this subtask was the number of words from the passage that children could correctly read in one minute (cwpm).

Figure 4 shows that 22% of the students in Grade 3 could not correctly read a single word of the first eight words of the passage. This is a large percentage of nonreaders of known and familiar words. Sixteen (16%) percent of sampled students were reading at or above the benchmark of 45+ correct words per minute, with equal shares of students in the intermediate categories between high and no proficiency.

Correlational analyses suggest that weak performance at oral reading may be due to their limited knowledge of letter sounds (with a moderate effect size of $r^2 = .58$) and weak decoding skills, as measured by the invented reading subtask (with a moderate to large effect size of $r^2 = .64$). Taken together, these findings again recommend that children’s limited mastery of the letter sounds and weak decoding skills be first addressed to improve their oral reading fluency.

Figure 4: Distribution of Grade 3 Students on the Oral Reading Fluency Subtask

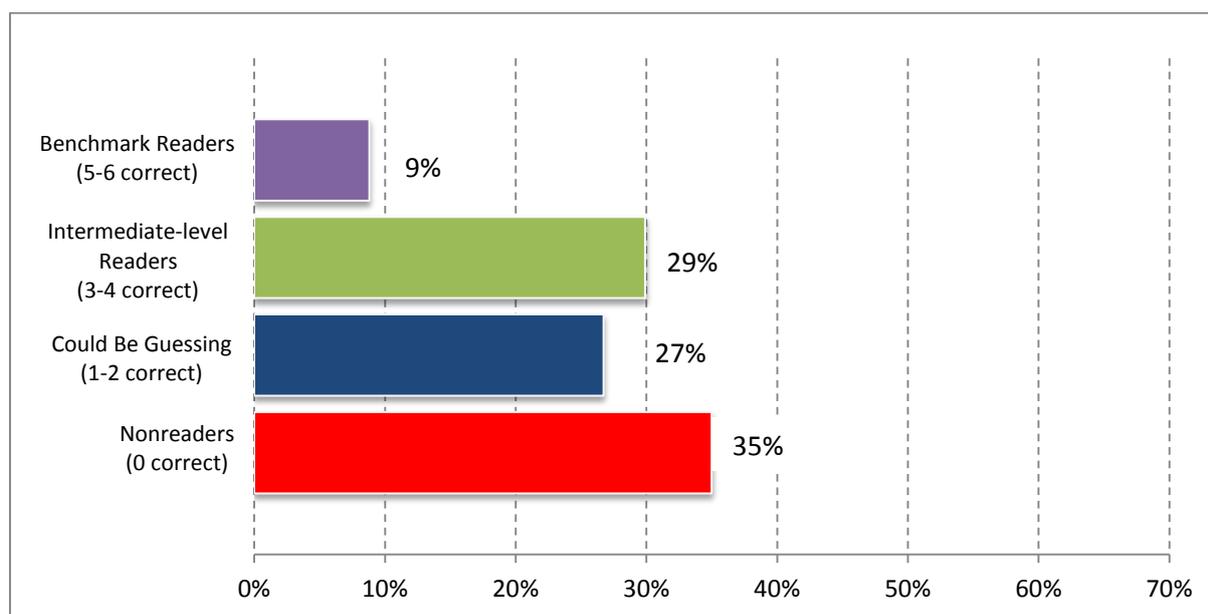


Reading Comprehension

After children had read the short reading passage (57 words) for one minute, *all those who were able to read at least one word correctly* were asked six questions about the story. Students who scored zero on oral reading frequency were not asked questions. The questions were both literal, requiring students to directly recall information from the story, and inferential, requiring students to combine information from the story with their background knowledge to derive a correct answer. Children’s reading comprehension scores were recorded as the number of correct responses to the six questions.

Overall, children had weak reading comprehension scores. More than one-third (35%) of students could not correctly answer a single question and another 27% answered just 1-2 questions correctly (**Figure 5**). Fewer than 10% of the children were performing at the benchmark level of 5-6 correct answers. This result for reading comprehension is significantly below the 16% of students who were reading the passage at benchmark proficiency. It confirms that many students, even strong readers who visibly impress their teachers by reading at speed with correct pronunciation, are reading “mechanically”—or “word calling”—with little or no understanding or retention of what they are reading. They have not been adequately taught to think about and comprehend the meaning of the passage they read.

Figure 5: Distribution of Grade 3 Students on the Reading Comprehension Subtask



Numerous large-scale studies and meta-analyses have reported robust correlations between oral reading fluency and reading comprehension.⁹ In other alphabetic languages, the relationship between decoding speed and reading comprehension is particularly strong among beginning readers because their word recognition skills still require conscious control.¹⁰ This was supported by the correlation ($r^2 = .58$) between students' scores in oral reading fluency and reading comprehension. These findings confirm that fluent oral reading is a critical component for reading comprehension, but that it is not sufficient for reading comprehension. In other words, addressing children's word recognition and decoding skills is critical for improving children's reading comprehension, but is not the only step required.

Listening Comprehension

In the EGRA listening comprehension subtask, the assessor read a short narrative story (71 words) to the child, followed by seven questions about that story. This was purely a listening subtask: the child was not given a copy of the story to follow along or refer to when answering the questions. All students, including the nonreaders who

⁹ See Abu-Rabia (2007); and also:

M.C. Daane, J.R. Campbell, W.S. Grigg, M. J. Goodman, & A. Oranje. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading (NCES 2006-469)*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.
G.S. Pinnell, J.J. Pikulski, K.K. Wixson, J.R. Campbell, P.B. Gough, & A.S. Beatt. (1995). *Listening to children read aloud: Data from NAEP's Integrated Reading Performance Record (IRPR) at grade 4 (NCES 95-726)*. Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

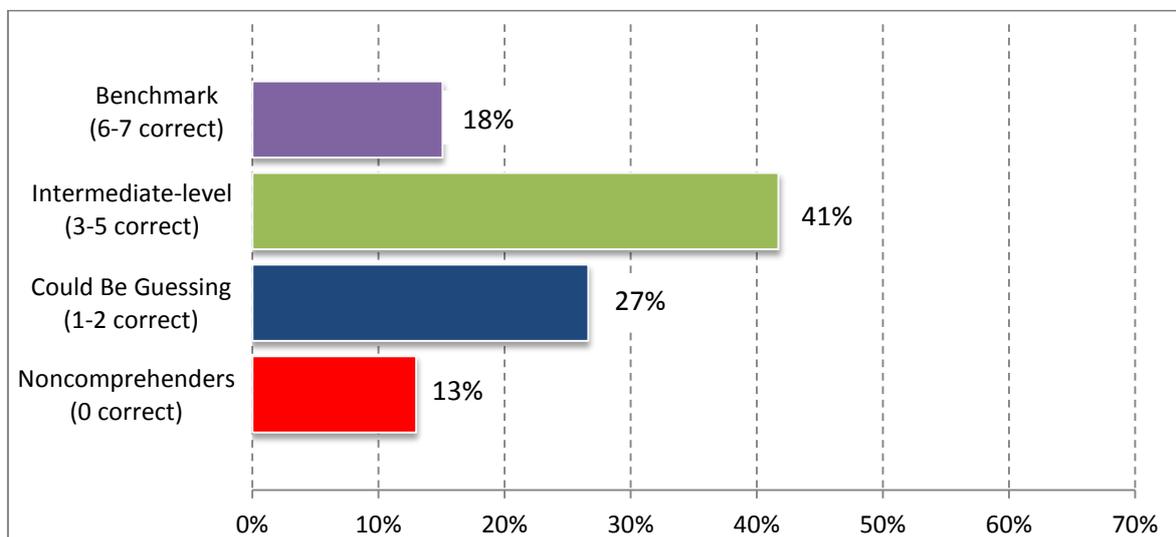
¹⁰ W.A. Hoover & P.B. Gough. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127–160.

did not correctly read a single item in the letter sounds, nonword decoding, and oral reading fluency subtasks, were tested for listening comprehension.

Although the listening comprehension subtask typically assesses a range of language and skills, such as attention, vocabulary knowledge, comprehension strategies, processing of oral language, and generation of appropriate replies, for Egyptian children, it also assessed their proficiency in the formal dialect of Arabic. Modern Standard Arabic differs substantially from the vernacular dialect used in their homes. Comparing children’s comprehension in these two modalities is important, because it allows determination of whether poor reading comprehension can be attributed to limited reading skills or to more general difficulties in comprehending the formal Arabic dialect used in schools.

In general, the listening comprehension subtask proved to be challenging (**Figure 6**). Although children’s listening comprehension scores were stronger than their reading comprehension scores, their overall performance was still weak. Thirteen (13%) percent of the children were unable to correctly answer any listening comprehension questions and just 18% could correctly answer six (or 86%) of the seven listening comprehension questions correctly—the benchmark for this subtask. These findings emphasize the often underestimated challenge that children schooled in Modern Standard Arabic face: proficiency in the vernacular, home dialect does not prepare students for the linguistic demands of Modern Standard Arabic taught in schools.

Figure 6: Distribution of Grade 3 Students on the Listening Comprehension Subtask



The research team also examined the relationship between listening comprehension and reading comprehension. Whereas oral reading fluency shared a large amount of variance with reading comprehension ($r^2 = .58$), listening comprehension’s correlation with reading comprehension was less ($r^2 = .42$). Thus, it appears that in addition to children’s decoding skills, children’s reading comprehension also reflected their difficulties in comprehending oral stories in the formal dialect.

Children would benefit not only from instruction that would build their decoding and word recognition skills, but also from instruction that would help them further develop their proficiency in Modern Standard Arabic.

Maze Reading Comprehension

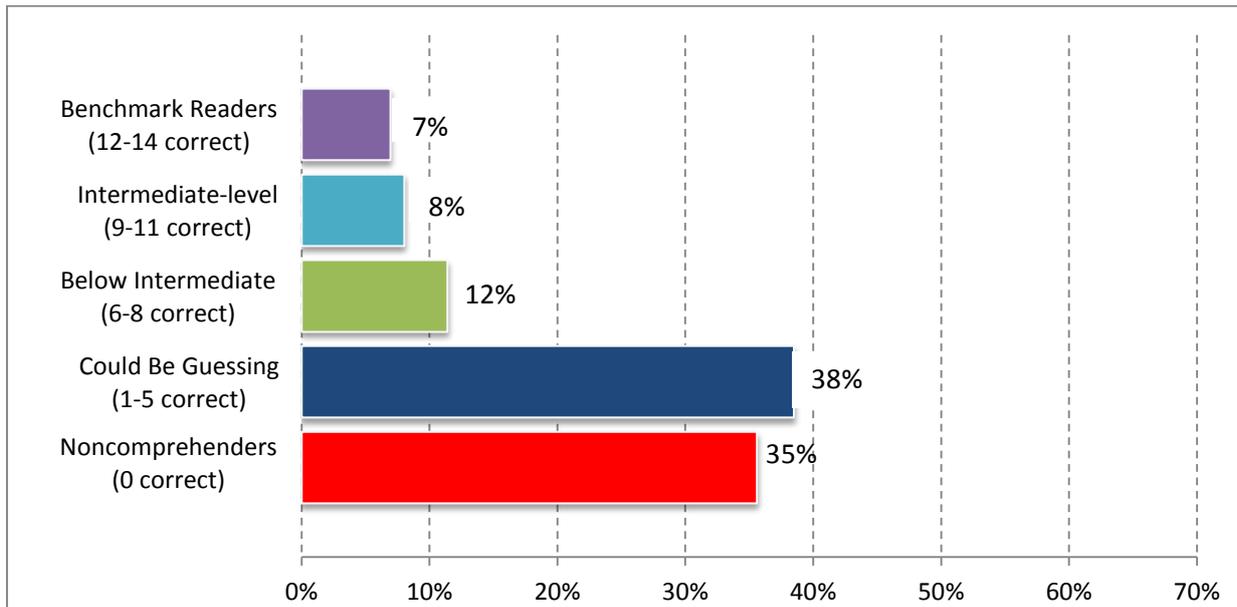
In addition to measures of reading and listening comprehension, we assessed children’s reading comprehension using the multiple-choice format of a Maze test. This was the most difficult of the three comprehension tests. In the Maze test, children were asked to read a narrative story of some 140 words and, as they read, identify which of three words in each of 14 selections in the story was the best word for the story. To correctly choose the best word in each selection, the children must comprehend the story as they read. Each child had three minutes to complete the story and all 14 selections. As this was an unfamiliar format to students, assessors carefully rehearsed the test for 1-2 minutes with each child, using a short practice paragraph with three items of multiple-choice selection that the assessor and student practiced together before implementing the test. Children were also asked to confirm that they understood the exercise before the actual test was started.

Even using the multiple-choice format of the Maze test, children had weak reading comprehension scores. As shown in **Figure 7**, over one-third (35%) of students were noncomprehenders. This category includes two groups of sampled students: the 295 nonreaders who did not correctly select a single item in the letter sound, nonword, and oral reading fluency subtasks and were not given the Maze test *in addition to* the 448 students who took the Maze subtask but did not make any selection in the first four choices. These latter students (22% of the total sample) read all three words in each of the first 4 multiple-choice selections but chose none and continued reading the story. After four consecutive non-selections, the test was halted for these “noncomprehenders.”

The largest share of Grade 3 students (38%) made just 1-5 correct selections. For a multiple-choice format with just 3 choices, students who simply guess or randomly choose one of the selections will, on average, be correct one-third of the time. These 38% of students scored no better than if they had been guessing. The low number of correct selections is partly due to the fact that many students read too slowly to complete the story, even in the three minutes allotted. Unfamiliarity with the test format might also have lowered performance some—but only partially. In fact, the students taking the Maze test had the lowest percentage of correct answers (26%) *per items attempted* (see **Table 2**).

These low scores on the Maze subtask recommend greater instruction by teachers and classroom practice by students using this format that requires children to think along as they are reading in order to choose the “best” word for each selection in the story.

Figure 7: Distribution of Grade 3 Student on the Maze Comprehension Subtask

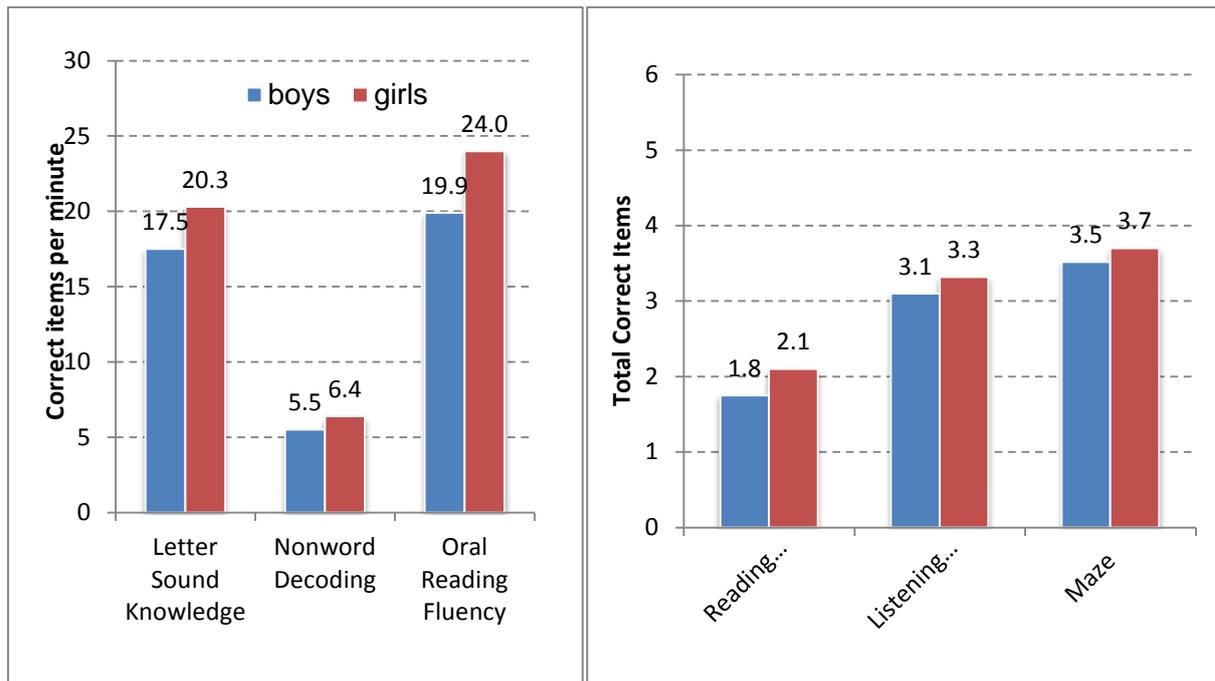


Finally, the Maze measure of reading comprehension correlated strongly with oral reading fluency ($r^2 = .68$). This is higher than the relationship between traditional measures of reading comprehension and oral reading fluency. Similarly, the Maze subtask shared a moderate correlation with the reading comprehension subtask ($r^2 = .51$). However, the Maze measure of reading comprehension's relationship with listening comprehension was much smaller ($r^2 = .26$).

Gender Differences in Reading Performance

The findings point to gender differences in reading proficiency in Grade 3 (**Figure 8**). Although boys and girls showed comparable performance in the listening comprehension and Maze comprehension subtasks, girls out-performed boys in their knowledge of letter sounds, reading nonwords, oral reading fluency, and reading comprehension. These differences are statistically significant. These Grade 3 results are consistent with the findings of previous Egypt EGRAs in Grades 1 and 2.

Figure 8: Literacy Achievement for Boys and Girls



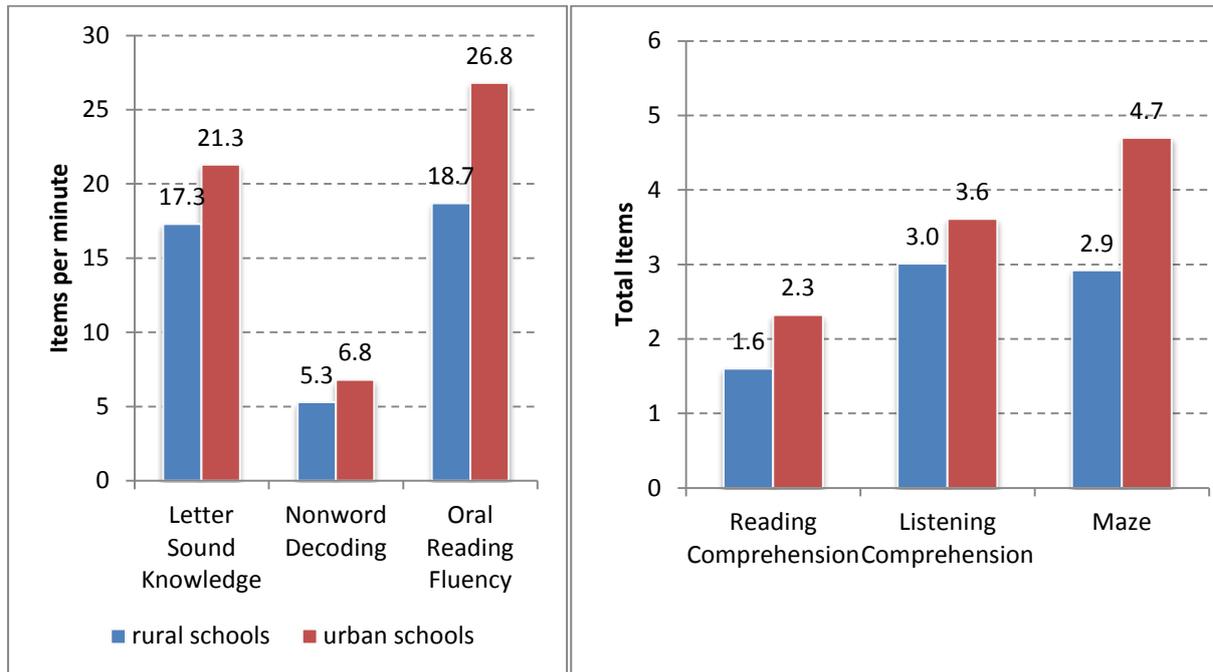
Note: Separate scales were used for the two parts of Figure 8. The graph on the left shows student performance on the timed tasks and uses items/minute as the unit of measurement; whereas the graph on the right shows student performance on the comprehension tasks, including the timed Maze test.

Geographical Differences in Reading Performance

Rural/Urban Differences in Reading Performance

Children who attended urban schools performed better on some subtasks (**Figure 9**). That is, children who attended urban schools were more successful at producing the sounds associated with each letter, in their decoding skills, and in their oral reading fluency. Attending school in urban centers was also associated with stronger performance in both listening and reading comprehension.

Figure 9: Literacy Achievement for Students Who Attended Urban and Rural Schools



Note: Separate scales were used for the two parts of Figure 10. The graph on the left shows student performance on the timed tasks and uses items/minute as the unit of measurement, whereas, the graph on the right shows student performance on the timed and untimed comprehension tasks.

This result may partially reflect better quality and less-interrupted instruction in urban schools. Teachers in urban schools are more likely to live close to their work and, thus, less likely to be absent. But in many rural schools, especially schools in small and remote villages, teachers commute from their homes elsewhere, often at considerable distance. Our sample included small and remote schools. Local transportation to these villages is often difficult and time-consuming. Absenteeism is typically greater among rural teachers in these locales.

But the better performance of urban students partially reflects the higher educational status, on average, of their parents and the greater value and attention that better educated parents give to the educational success of their children. When both of these factors are considered, the resulting difference in the average reading performance of village and urban children is not considerable. Except for the oral reading fluency and Maze results, the mean scores of urban students on most subtasks are not significantly higher than rural students. Low reading proficiency is not confined to rural schools but is prevalent also in many urban schools.

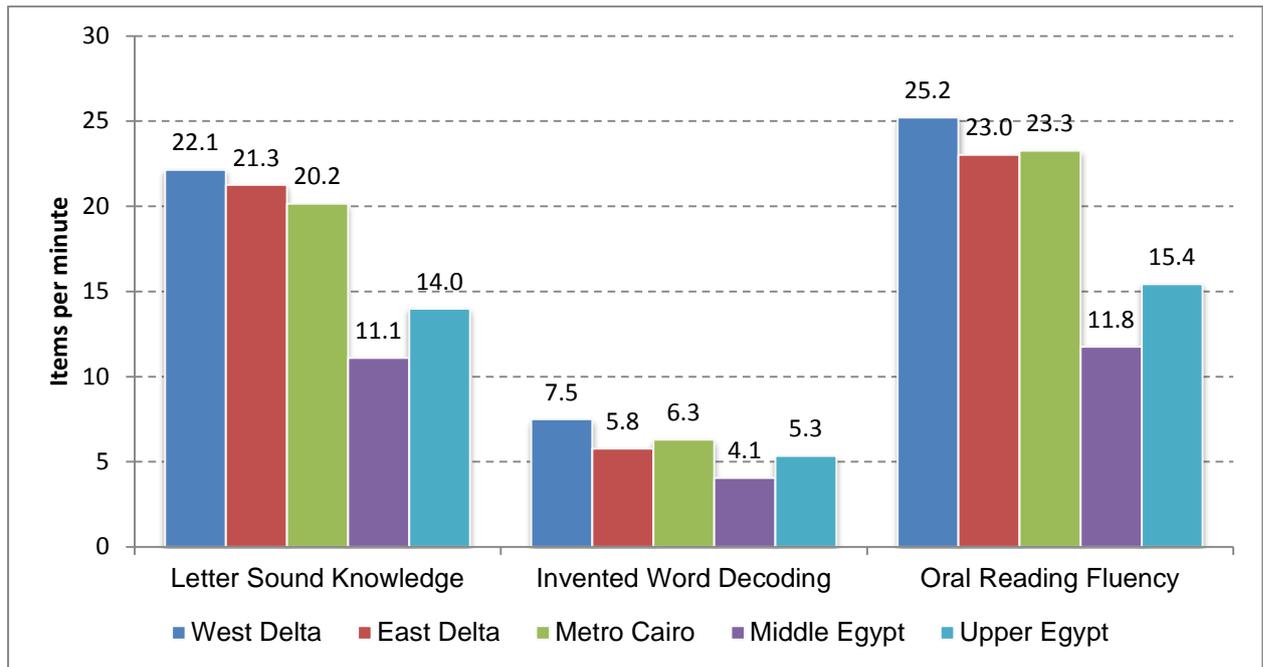
Reading Performance across Regions

The findings also illustrate regional differences in educational opportunities for children (please see **Figure 11**). Overall, children from West Delta schools were more successful in their mastery of print (that is, in producing the sounds associated with each letter, in their decoding skills, and in their oral reading fluency) and in their reading comprehension than their peers attending Mid Egypt and Upper Egypt

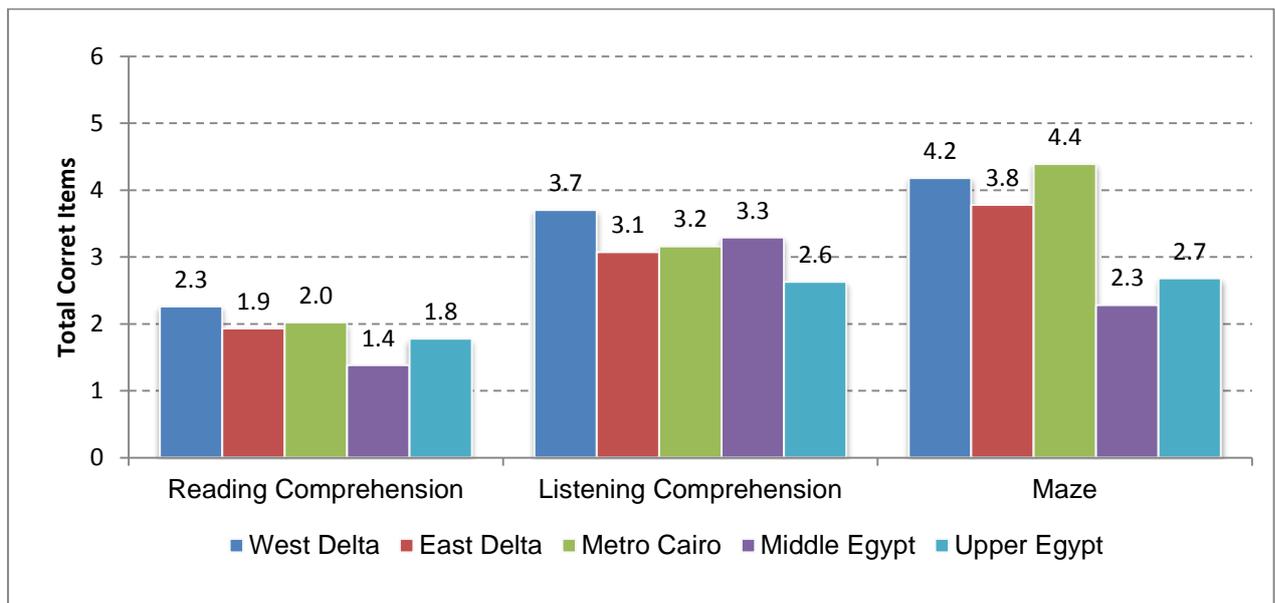
schools. West Delta students decoded more nonwords on average than did their peers in East Delta schools. Finally, West Delta students showed stronger comprehension of orally presented stories than their peers attending schools in the East Delta, Metro Cairo, and in Upper Egypt.

Further analysis of the regional differences and school-level variance in this national Grade 3 baseline will be added after initial comments to this draft report.

Figure 10: Regional Differences in Literacy Achievement for Students



Note: Separate scales were used for the two parts of Figure 11. The graph above shows student performance on the timed tasks and uses items/minute as the unit of measurement. The graph below shows student performance on the timed and untimed comprehension subtasks.



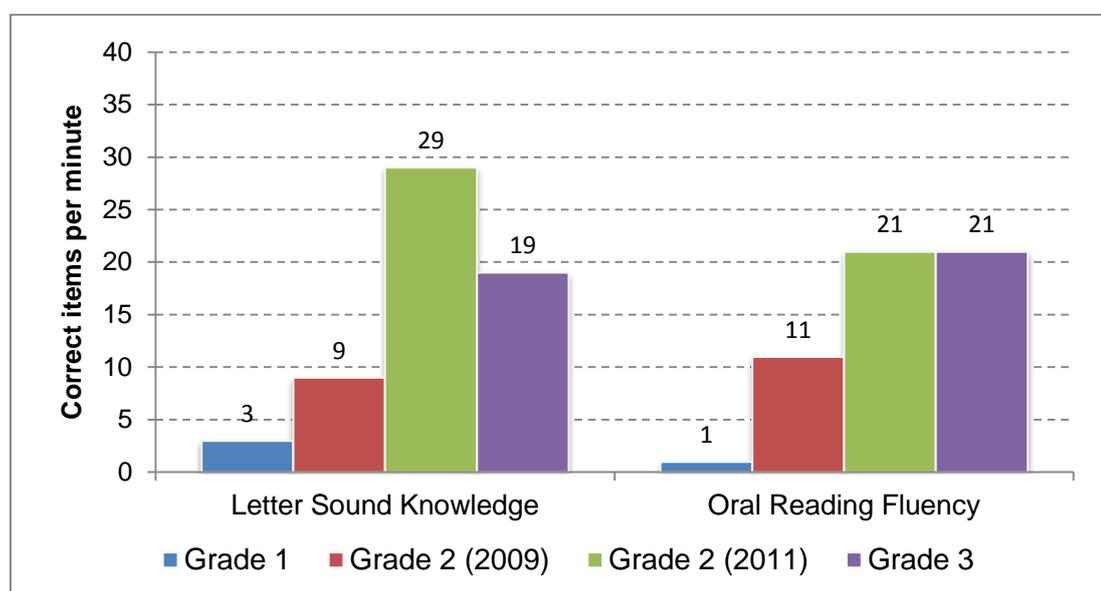
Conclusions and Recommendations

The baseline results make clear the weak average reading proficiency of current Grade 3 students. This student cohort did not benefit in 2011/12 or 2012/13 from the MOE Early Grade Reading Program in Grades 1 and 2. Average proficiency in each of the tested reading skills is significantly below proposed benchmarks. For all subtasks except letter sounds knowledge, fewer than 20% of students performed at or above the benchmark scores.¹¹

Performance Comparison with Previous Egypt EGRAs

On two subtasks—letter sounds knowledge and oral reading fluency—we can loosely compare the average baseline scores for this Grade 3 with the results of previous Egypt EGRAs conducted in Grades 1 and 2 (*Figure 11*).¹² Two key points stand out from this rough comparison. First, there is a natural progression in average proficiency from Grade 1 to Grade 2 to Grade 3. Average scores on letter sounds knowledge rise from 3 correct letter sounds per minute in Grade 1 to 9 correct letter sounds in Grade 2 (2009) and 19 correct letter sounds in Grade 3. In oral reading fluency, the progression of average scores in reading a simple passage begins at **1** correct word per minute in Grade 1, rising to **11** correct words in Grade 2 (2009) and then **22** correct words in Grade 3. This was the progression and general trend in reading skills acquisition *before* implementation of the MOE Early Grade Reading Program.

Figure 11: Average Scores on Comparable Subtasks for all Egypt EGRAs



¹¹ Thirty (30%) percent of tested students achieved the benchmark score for letter sounds knowledge.

¹² Comparison of other subtasks is not possible either because the subtasks were significantly different (e.g., nonword decoding) or were not included (reading comprehension, listening comprehension, Maze comprehension) in all four EGRAs. This is a “loose” and not strict comparison as previous Egypt EGRAs were not national assessments or nationally representative. The point of the comparison is to reveal general trends and not exact figures.

Second, and especially important, this comparison reveals the impact of improved reading instruction. The average Grade 2 scores in 2011 on these two subtasks are post-implementation of the GILO Early Grade Reading Program. These average scores for Grade 2 in 2011 *meet or significantly exceed* the average Grade 3 baseline scores of students who completed an additional year of schooling. On letter sound knowledge, the *average* Grade 2 score in 2011—29 correct letter sounds per minute—surpassed the provisional Grade 3 benchmark for this subtask (27 clpm). This comparison hints at the scale of reading improvement that can be expected from planned extension of the MOE Early Grade Reading Program to Grade 3.

Nonreaders

This national baseline highlights the percentage of **nonreaders** in this Grade 3. Over 14% of tested children could not correctly answer any item in the first three subtasks: letter sounds, nonwords, and oral reading fluency (passage reading). These are illiterate students, not struggling readers. Their reading abilities are either zero or limited to the few words whose forms they have memorized and can recognize on sight. The higher percentage of zero scores (22%) on oral reading fluency hint that the true share of functional illiterates is higher than 14%.¹³ One in every 5 to 7 students in this Grade 3 cohort have zero reading skills. Soon to enter Grade 4, these nonreaders are unlikely ever to learn to read without remedial instruction.

How much reduction in nonreaders is possible? The answer to this question might best be indicated in comparing zero scores for these same two subtasks—letter sounds knowledge and oral reading fluency—in Grade 2, between the 2009 baseline and the 2011 EGRA after six months of the GILO Early Grade Reading Program. In 2009, *nearly half* of all Grade 2 students had zero scores on these two subtasks: 48% for letter sounds knowledge and 44% for oral reading fluency. In 2011, those percentages—for Grade 2 children from the same schools—were 11% and 21%. The percentage of nonreaders had fallen dramatically. The improved reading performance achieved by the Early Grade Reading Program equals nearly a full year of additional schooling.

Nonreaders markedly lower average grade scores. And they profoundly challenge teachers seeking to “pitch” reading instruction to the average abilities of students in their classes. Lowering the percentage of nonreaders (zero scores) is as important a benchmark for measuring grade reading performance as the average scores on specific subtasks. Comparing average scores alone masks the scope of illiteracy within schools. Setting national benchmarks for *both* average scores on specific EGRA subtasks *and* reducing the percentage of nonreaders in each grade are key recommendations of this study.

¹³ Another 7% of students were barely able to read. They read fewer than 5 correct words on the oral reading fluency subtask in one minute.

Promote Technical Standards for EGRAs in Modern Standard Arabic for Specific Grades

It is less than five years since the first EGRA instrument was prepared in Modern Standard Arabic. Since then, multiple EGRAs in Modern Standard Arabic have been implemented in Egypt, Iraq, Jordan, Morocco, and Yemen. These EGRAs often include the same subtasks. But the grammatical content and level of difficulty of these subtasks can vary significantly. As professional experience with EGRAs in



Modern Standard Arabic grows, the value added of setting technical standards for subtask design for specific grades becomes manifest. Standardizing Arabic EGRAs will promote “best practice” in reading instruction on individual skills and greater sharing of professional experience of “what works best” in the classroom. Standardizing the design of EGRA subtasks for different grades will better inform the content of

grade textbooks, the design of routines and classroom activities for effective reading instruction, and the motivation of constructive competition in accelerating progress towards reading proficiency. It will exponentially broaden the audience of Arabic teachers and professionals exchanging tips and crowd-sourcing of new activities and effective approaches for sustained improvement of student reading performance. Learning to read Modern Standard Arabic in school challenges tens of millions of young children. Steepening their learning curve through improved teaching practice in reading skills, measured through the comparing and sharing of outcomes, is a timely recommendation for future EGRAs.

Recommendations and Next Steps: Outcomes of the Policy Dialogue Workshop

To be added following the Policy Dialogue Workshop scheduled for July 2-4, 2013, in Cairo, Egypt.

Annex A: The National EGRA Baseline Instrument for Grade 3

أداة تقييم مهارات القراءة في الصفوف الأولى: نموذج تعليمات الفاحص ٢٠١٣

تعليمات عامة

من المهم أن تضفي جواً من المرح مع الطفل الذي سيخضع للتقييم بحيث تبدأ معه بمحادثة بسيطة حول مواضيع تهمة (انظر المثال أدناه). أشعره بأن هذا التقييم هو تقريبا بمثابة لعبة سيستمتع بها وليست بالمهمة الصعبة. من المهم جداً أن تقرأ محتوى المربعات فقط، بصوت عالٍ وبوضوح وتمهل.

صباح الخير. اسمي _____ أسكن في _____. أريد أن أتكلم معك عن نفسي، عندي من الأطفال، عمرهم؛ عندي في البيت.....، الرياضة التي أمارسها، الخ] ١. تكلم لي عن نفسك وعن عائلتك؟ [انتظر الجواب؛ إذا كان التلميذ غير متحمس للكلام، وجه إليه السؤال رقم ٢. إذا تكلم بارتياح، انتقل لفقرة الموافقة الشفهية]. ٢. ما اللعبة التي تحبها؟

الموافقة الشفهية

اسمح لي أن أقول لك لماذا أنا معك اليوم. أنا أعمل في وزارة التربية والتعليم وأحاول أن أفهم كيف يتعلم الأطفال القراءة. لقد تم اختيارك للقيام بهذا الاختبار بشكل عشوائي. أحب أن تتعاون معي في هذه العملية. ولكن إذا لم ترد المشاركة، فلك ذلك. سنلعب لعبة القراءة حيث سأطلب منك أن تقرأ بعض الحروف وبعض الكلمات وقصة قصيرة بصوت عالٍ. سأستعمل هذه الساعة لأحسب الوقت الذي تحتاجه في القراءة. هذا ليس امتحاناً وليس له أي تأثير على علامتك المدرسية. سأسألك بعض الأسئلة الأخرى عن عائلتك. لن أكتب اسمك على ورقة الاختبار. لن يرى أي أحد إجاباتك عليه. مرة أخرى، أنت غير ملزم بالمشاركة إذا لم تكن ترغب في ذلك. وإذا بدأتنا ولم ترد الجواب عن أي سؤال، فلا مشكلة في ذلك. هل لديك سؤال؟ هل أنت مستعد؟

إذا حصلت على الموافقة الشفهية للطفل ضع علامة (x) في هذا المربع نعم (إذا لم تحصل على الموافقة، اشكر الطفل وانتقل للطفل الذي بعده واستعمل نفس الاستمارة)

١. تاريخ التقييم مثال: ١٥ مارس ٢٠١٣ = ٢٠١٣-٠٣-١٥	اليوم: _____ الشهر: _____ السنة: _____
٢. المديرية:	
٣. الإدارة:	
٤. اسم المدرسة	
٥. كود المدرسة	
٦. فترات الدراسة بالمدرسة	<input type="radio"/> فترة واحدة <input type="radio"/> فترة صباحية <input type="radio"/> فترة مسائية
٧. اسم الفاحص	
٨. الصف:	<input type="radio"/> الثالث (٣)
٩. تاريخ ميلاد الطفل:	الشهر: _____ السنة: _____
١٠. جنس الطفل:	<input type="radio"/> طفل <input type="radio"/> طفلة
١١. وقت البدء بالاختبار:	<input type="checkbox"/> صباحاً (اختر واحدة منها) <input type="checkbox"/> مساءً

القسم ١- التعرف على أصوات المقاطع

٦٠ ثانية

بعد مرور ٦٠ ثانية،
ستقول للطفل
'توقف'.

إذا تردد الطفل في
قراءة الحرف لمدة
تزيد عن ٣ ثوانٍ،
أشّر للحرف الذي
يليه وقل: "انكمل
من فضلك".

قاعدة التوقف
المبكر: إذا وضعت
علامة (/) على
جميع الأوجية في
السطر الأول على
أنها خاطئة ولم
يصحح الطفل أي
خطأ من أخطائه،
قل "شكراً" وأوقف
التمرين. ضع
علامة (x) في
المربع الموجود في
أسفل الصفحة
وانتقل للتمرين
الذي بعده.

هذه ورقة تضم حروفاً مضبوطة بحركات، اقرأ قدر ما تستطيع منها (اقرأ صوت الحرف مع الحركة، وليس اسمه).
مثلاً، صوت هذا الحرف مع الحركة [أشّر إلى الحرف "س"] هو "س" كما في كلمة "سَمَك".

و الآن لنقم بهذا التمرين: قل لي صوت هذا الحرف [وأشّر إلى الحرف "ض"]:
إذا كان جواب الطفل صحيحاً، قل: جيد، الصوت هو "ض"
إذا كان جواب الطفل غير صحيح، قل: الصوت هو "ض"
لتجرب مثلاً آخر: قل لي صوت هذا الحرف مع الحركة [أشّر إلى الحرف "ج"]:
إذا كان جواب الطفل صحيحاً، قل: أحسنت، الصوت هو "ج" إذا كان جواب الطفل غير صحيح، قل: الصوت هو "ج"
هل فهمت المطلوب منك؟

عندما أقول لك "لنبدأ"، اقرأ الأصوات بدقة وبأكبر سرعة ممكنة. سنبدأ من هنا ونكمل بهذه الطريقة [أشّر إلى الصوت الأول في السطر الأول، وتنبع معه بأصبعك على الحروف الموجودة في السطر الأول بأكمله]. هل أنت مستعد؟ لنبدأ

ضع بوضوح علامة (/) على أي خطأ يرتكبه الطفل.

في حالة قيام الطفل بتصحيح نفسه، قم بوضع دائرة (o) حول علامة (/) التي وضعتها مسبقاً له.

ضع العلامة () على آخر حرف قرأه الطفل.

مثال: سَ ضَ جَ

١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
غُ	و	بِ	طَ	ضِ	ظُ	ثُ	فِ	نِ	خُ
مِ	ثُ	هَ	طُ	أُ	حَ	دُ	ذُ	لِ	قُ
جُ	زُ	سَ	نَ	شُ	دُ	وُ	تِ	سِ	جُ
صُ	دَ	بِ	لُ	زُ	شُ	صُ	أُ	مُ	عَ
بِ	وُ	قُ	رِ	يَ	خِ	عِ	حَ	قَ	رِ
تِ	قَ	جُ	سِ	وُ	هَ	رِ	غُ	و	يَ
شُ	دُ	حَ	لُ	زُ	شُ	عَ	طَ	ضِ	ظُ
زُ	سَ	نَ	رِ	يَ	خِ	مُ	ثُ	فِ	نِ
لِ	قَ	جُ	دَ	بِ	أُ	خُ	مِ	ثُ	
حَ	دُ	بِ	وُ	قُ	صُ	عِ	طُ	أُ	

الوقت المتبقي من وقت التمرين (عدد الثواني):

ضع علامة (x) في هذا المربع في حال أوقفت هذا الجزء من التقييم لأن الطفل لم يقرأ أي حرف في السطر الأول بشكل صحيح.



بعد مرور ٦٠ ثانية، ستقول توقف!

إذا تردد الطفل في قراءة كلمة لمدة تزيد عن ٣ ثوانٍ. أشر للكلمة التالية وقل: "النكمل من فضلك".

قاعدة التوقف الميكر:
إذا وضعت علامة (/) على جميع الأوجه في السطر الأول لأنها خاطئة ولم يصحح الطفل أي خطأ من أخطائه، قل "شكراً" وأوقف التمرين. ضع علامة (x) في المربع الموجود في أسفل الصفحة وانتقل للتمرين الذي بعده.

هذه بعض الكلمات ليس لها معنى. اقرأ بشكل صحيح أكبر عدد ممكن. لا تقرأ صوتاً بصوت بل اقرأ الكلمة بالكامل. مثلاً هذه الكلمة ليس لها معنى هي "ماصيح".

الآن اقرأ الكلمة التالية: [أشر إلى كلمة بيلاد]:

إذا قال الطفل "بيلاد"، قل له: أحسنت، "بيلاد"

إذا لم يقل الطفل "بيلاد" بشكل صحيح، قل: هذه الكلمة ليس لها معنى هي "بيلاد"

لنجرّب الآن كلمة أخرى: اقرأ هذه الكلمة [أشر إلى كلمة "ناسيب"]:

إذا قال الطفل "ناسيب"، قل: جيد جداً، "ناسيب"

إذا لم يقل الطفل "ناسيب" بشكل صحيح، قل: هذه الكلمة ليس لها معنى، هي

"ناسيب". عندما أقول لك "ابدأ"، اقرأ الكلمات بدقة وبأكبر عدد ممكن. سنبدأ من هنا

ونكمل بهذه الطريقة [أشر إلى الكلمة الأولى في السطر الأول، وتتبع معه بأصبعك

الكلمات في السطر الأول بأكمله]. هل أنت مستعد؟ لنبدأ

ضع بوضوح علامة (/) على أي خطأ يرتكبه الطفل.

في حالة قيام الطفل بتصحيح نفسه، قم بوضع دائرة حول علامة (/) التي وضعتها مسبقاً له

ضع العلامة (x) على آخر كلمة قرأها الطفل.

مثل: ماصيح بيلاد ناسيب

	٥	٤	٣	٢	١
(٥)	شُجِبُ	نَلَر	تُقِيرُونَ	أزري	ضا
(١٠)	أشِي	مائي	قُدْفُنْ	را	ظلو
(١٥)	كَلْبَةٌ	عِيه	بي	هَلِيب	زَوْن
(٢٠)	صَخْت	خَيْرَم	سا	عَكِير	أكا
(٢٥)	نِيها	أَنر	رَقْلَمْ	فَع	صلوذ
(٣٠)	تُخْم	تُداساً	تُكَلِجِي	ظاري	دَف
(٣٥)	دَاطِ	وملسي	أَنْفَصَر	دَلْعَبْ	بَلُحْ
(٤٠)	سَمَةٌ	رِكاه	داف	مَحْبُ	قَمْبُ
(٤٥)	يَمُغْ	مَرَك	سَمَةٌ	أَفْبُ	سَربمة
(٥٠)	قَبِيحَةٌ	عَقِل	أَبْتَسُنْ	لِناة	أَقِي

الوقت المتبقي من وقت التمرين (عدد الثواني):



ضع علامة (x) في هذا المربع في حال أوقفت هذا الجزء من التقييم لأن الطفل لم يقرأ أيّاً من الكلمات في السطر الأول بشكل صحيح.

القسم ٣ - أ.:		القسم ٣ - ب.:		القسم ٣ - ج.:	
قراءة نص شفيهاً		قراءة نص مكتوباً		قراءة نص مكتوباً	
<p>هذه قصة قصيرة، ركز جيداً وقرأها بشكل صحيح وبصوت عالٍ وبسرعة. حين تنتهي، سأسألك بعض الأسئلة حول ما قرأته. هل فهمت المطلوب منك؟ حين أقول لك "أبداً"، ابدأ بالقراءة. مستعد؟ أبداً.</p>		<p>اسحب نص القصة من أمام الطفل ووجه إليه الأسئلة أدناه.</p> <p>ترك للطفل ١٠ ثانية على الأكثر كي يجب عن كل سؤال، ووجه السؤال المقابل لكل سطر قرأه الطفل حتى تصل إلى السطر الذي يحتوي العلامة (I) والتي تشير إلى مكان توقف الطفل عن القراءة.</p> <p>ساوجه إليك الآن بعض الأسئلة حول القصة التي قرأت. أجب عن الأسئلة بشكل صحيح.</p>		<p>ضع علامة في الخانة التي تتناسب مع إجابة الطفل، و من ثم انتقل إلى السؤال الذي يليه.</p>	
٦٠ ثانية	بعد مرور ٦٠ ثانية، ستقول 'توقف'.	لا إجابة	صحيحة	غير صحيحة	صحيحة
<p>حين يتردد الطفل لمدة تزيد على ٣ ثوانٍ في قراءة الكلمة. أشر إلى الكلمة التالية وقل: <u>"أكمل من فضلك"</u></p> <p>قاعدة التوقف المبكر: إذا وضعت علامة (/) على جميع الكلمات في السطر الأول على أنها خطأ ولم يصحح الطفل أي خطأ من أخطائه، قل "شكراً" وأوقف التمرين. ضع علامة (X) في المربع الموجود في أسفل الصفحة وانتقل إلى التمرين الذي بعده.</p>	<p>ضع <u>بوضوح</u> علامة (/) على أي خطأ يرتكبه الطفل أثناء القراءة. ضع العلامة (I) على آخر كلمة قرأها الطفل.</p>	<p>ماذا تحب أماتي؟ (القراءة) (البحث) (البحث عن المعرفة) (المعرفة)</p> <p>إلى أين ذهبت أماتي مع والديها؟ (المكتبة العامة) (المكتبة)</p> <p>ماذا لاحظت أماتي عندما دخلت المكتبة؟ (الجميع يقرأون في جو هادئ) (الجميع يقرأون) (الجو الهادئ في المكتبة) (كثف قائلت أمينة المكتبة أماتي ووالديها؟ (رحبت بهما) (أهلاً بكما). (وعرضت عليهما المساعدة)</p> <p>كيف يتكلم أمينة المكتبة إن تساعدها أماتي؟ (تقدم لها القصص المفيدة) (والكتب الممتعة) (تساعدها في الحصول على القصص) (والكتب التي تحبها) (أو تحب قراءتها)</p> <p>ماذا قالت أمينة المكتبة لأماتي؟ (فضلي يا أماتي)؛ (في مكتبتنا كل ما تحتاجون إليه)</p>	<p>٨ أماني تلميذة ذكية تحب القراءة والبحث عن المعرفة،</p> <p>١٩ في إجازة يصف العام ذهبت أماني مع أبيها إلى المكتبة العامة؛</p> <p>٢٦ حين وجد الجميع يقرأون في جو هادئ،</p> <p>٣٦ استغاثتهما أمينة المكتبة قيلة. أهلاً بكما، كيف يمكنك أن تساعدهما؟</p> <p>٤٥ قال الأب: أماني تهوى قراءة القصص المفيدة، والكتب الممتعة،</p> <p>٥٧ قالت أمينة المكتبة: فضلي يا أماتي؛ في مكتبتنا كل ما تحتاجون إليه.</p>	<p>ضع الوقت المتبقي من وقت التمرين (عدد الثواني):</p> <p>ضع علامة (X) في هذا المربع في حال أوقفت هذا الجزء من التقييم لأن الطفل لم يقرأ أي كلمة في السطر الأول بشكل صحيح.</p>	

القسم ٤ - فهم المسموع

يقرأ الفالحص بصوت عال النص التالي ولمرة واحدة فقط ويتلأ (كلمة كل ثانية تقريباً).

اترك للطفل ١٠ ثانية على الأكثر كي يجيب عن كل سؤال.

قل للطفل: ساقرا عليك قصة قصيرة بصوت عال، مرة واحدة فقط. و بعد ذلك سأوجه إليك بعض الأسئلة. اسمع جيداً من فضلك وأجب عنها بشكل صحيح. هل فهمت المطلوب منك؟

"سارَة طفلة ذكيّة تعيش مع أسرتها في الريف، وفي يوم الجمعة ذهبت سارَة مع أمها إلى سوق القرية، وفي طريق عودتهما توقفتا عند دكان عم أحمد؛ لتشتريا اللبن، وبعد أن رحب بهما عم أحمد اعتذر لهما بأن اللبن لم يأت اليوم، قالت الأم: لا تحزني يا سارَة؛ سوف تشتري اللبن من دكان عم سعيد بجوار محطة القطار، ونسيت الأم أن عم سعيد لا يفتح دكانه في يوم الإجازة الأسبوعية"

صحيحة	غير صحيحة	لا إجابة	
			١- أين تعيش سارَة؟ في الريف (القرية) / مع أسرتها (عائلتها)
			٢- إلى أين ذهبت سارَة مع أمها؟ إلى سوق القرية / إلى السوق
			٣- لماذا توقفت سارَة عند دكان عم أحمد؟ لتشتري اللبن
			٤- كيف قابل عم أحمد سارَة ووالديها؟ رحب بهما / اعتذر لهما بأن اللبن لم يأت اليوم
			٥- لماذا كانت سارَة حزينة؟ لأنها لم تجد اللبن عند عم أحمد
			٦- غلام ماذا اتفقت سارَة وأمها؟ اتفقتا على شراء اللبن من دكان عم سعيد (بجوار محطة القطار).
			٧- في النهاية؛ لماذا رجعت سارَة إلى بيتها دون أن تشتري اللبن؟ لأن دكان عم سعيد كان مغلقاً / لا يفتح دكانه في يوم الإجازة (الجمعة)

القسم ٥ - متاهة الفهم

٣ دقائق (١٨٠ ثانية)



بعد مرور ٣ دقائق،
ستقول "توقف".



حين يتردد الطفل لمدة
تزيد على ٣ ثواني في
قراءة الكلمة. أشر إلى
الكلمة التالية وقل:
"لنكمل من فضلك"



قاعدة التوقف المبكر:
إذا أخطأ التلميذ في
الاختيارات الأربعة
الأولى؛ قل "شكراً"
وأوقف التمرين. ضع
علامة (x) في المربع
الموجود في أسفل
الصفحة.



اترك للطفل ١٠ ثواني
على الأكثر كي يجيب
عن كل اختيار.

حين يتردد الطفل لمدة
تزيد على ١٠ ثواني
في الاختيار، أشر إلى
الكلمة التالية وقل:
"لنكمل من فضلك"

قدم للتلميذ/ة الصفحة الأولى من صفحات تقييم المتاهة، وقل:

في هذا النشاط، سوف تقرأ قصة من نوع خاص؛ فقد تم استبدال بعض الكلمات في القصة بمجموعة من ثلاث كلمات، وعليك أن تختار أنت الكلمة الأكثر مناسبة للقصة من مجموعة الكلمات الثلاث، وأشر إلى الكلمة التي تعتقد أنها الأفضل.

سنبدأ بالتدريب معاً على ذلك هنا (أشر إلى الجملة)، سوف تقوم بقراءة الجملة الأولى لنفسك بينما أقوم أنا بقراءتها بصوت عال.

كان أَشْرَفُ يَقِفُ عَلَى جَانِبِ الطَّرِيقِ،

الآن سوف أقرأ الجملة التالية:

وكانت (السُّنُنُ - الطُّيُورُ - السَّيَّاراتُ) تَسِيرُ بِسُرْعَةٍ.

كلمة "السيارات" تتناسب مع بقية القصة؛ أشر إلى الكلمة "السيارات". (تأكد من أن التلميذ/ة قد أشر إلى الكلمة الصحيحة).

لنقرأ الجملة التالي:

حاول أَشْرَفُ (العُيُورَ - الثَّوَمَ - اللُّعِبَ) إِلَى الْجَانِبِ الأَخْرِ،

ما الكلمة المناسبة هنا في هذه الجملة؟ (استمع إلى التلميذ)؛ (فإن كان مصيباً؛ قل: نعم، كلمة "العُيُورَ" تناسب القصة تماماً؛ أشر إلى الكلمة "العُيُورَ". (وإن كانت الإجابة خطأ؛ قل: لا، في الواقع إن كلمة "العُيُورَ" تناسب الجملة بصورة أفضل؛ أشر إلى الكلمة "العُيُورَ" وأنطق الكلمة.

أما فيما يتعلق بالجملة الأخيرة؛ فأريد منك أن تقرأ بنفسك وتشر إلى الكلمة التي تناسب الجملة بصورة أفضل. اعمل بسرعة، ولكن ليس بهذه السرعة الكبيرة التي تجعلك تقع في الخطأ. الآن اقرأ الجملة. (اعط التلميذ وقتاً كي يقرأ الجملة ويضع دائرة حول الكلمة).

كلمة "تتوقف" صحيحة؛ فإذا كنت قد أشر إلى كلمة أخرى؛ ضع خطأ عليها، وأشر إلى كلمة "تتوقف".

أطلب من التلميذ أن يقلب الصفحة. عندما أقول "ابدأ" اقرأ الكلمات بسرعة ودقة بقدر ما تستطيع، وكلما رأيت مجموعة من الكلمات؛ أنطق الكلمة الصحيحة. سوف أصمت أنا وأستمع إليك. هل تفهم ما سوف تقوم بعمله؟ هل أنت مستعد؟ ابدأ.

الوقت المتبقي من وقت التمرين (عدد الثواني):

ضع علامة (x) في هذا المربع في حال أوقفت هذا الجزء من التقييم لأن الطفل أخطأ في الاختيارات الأربعة الأولى.

بَيْنَمَا كَانَ التُّغْلُبُ يَسْتَرِيحُ تَحْتَ ظِلِّ شَجَرَةٍ رَأَى غُرَابًا (يَطِيرُ - يَغُومُ -
يَقْرَأُ) وَفِي مَنْفَارِهِ قِطْعَةً مِنَ الْجُبْنِ، ثُمَّ حَطَّ (عَن - عَلَى - فِي) غُصْنِ
الشَّجَرَةِ الَّتِي يَجْلِسُ تَحْتَهَا؛ قَالَ التُّغْلُبُ لِنَفْسِهِ: لَا بُدَّ أَنْ (أَنَامَ - أَحْصَلَ -
أَجْلَسَ) عَلَى قِطْعَةِ الْجُبْنِ هَذِهِ؛ فَأَنَا (عَطْشَانٌ - سَعِيدٌ - جَائِعٌ) جِدًّا، وَلَمْ
أَكُلْ شَيْئًا مُنْذُ (الْبَحْرِ - الصَّبَاحِ - الطَّرِيقِ).

وَقَفَ التُّغْلُبُ وَنَظَرَ إِلَى أَعْلَى (النَّهْرِ - الشَّجَرَةِ - الْبَيْتِ) قَائِلًا: صَبَاحُ
الْخَيْرِ يَا سَيِّدِي (الْبَطَّةَ - الْغُرَابَ - الشَّجَرَةَ)، مَا أَجْمَلَ لَوْنَ رِيَشِكَ الْيَوْمَ،
أَمَّا عَيْنَاكَ (فَهُمَا - فَهْمٌ - فَهْوٌ) لِامْتِعَانِ كُنُجُومِ السَّمَاءِ، لَا بُدَّ أَنْ صَوْتُكَ
الْجَمِيلَ لَا مَثِيلَ لَهُ بَيْنَ (النَّاسِ - الطُّيُورِ - الْحَيَوَانَاتِ) الْآخَرَى، دَعْنِي
أَسْتَمِعَ إِلَى لَحْنِ عَذْبِ (مِنْكُمْ - مِنْكُمَا - مِنْكَ) يَا مَلِكِ الطُّيُورِ.

كَانَ الْغُرَابُ (حَزِينًا - غَاضِبًا - سَعِيدًا) بِكَلَامِ التُّغْلُبِ فَرَفَعَ رَأْسَهُ وَبَدَأَ
بِالنَّعِيقِ بِأَعْلَى صَوْتِهِ؛ فَسَقَطَتْ قِطْعَةُ (اللَّحْمِ - الْجُبْنِ - السَّمَكِ) إِلَى
أَسْفَلِ الشَّجَرَةِ؛ فَأَمْسَكَ بِهَا (الْكَلْبُ - التُّغْلُبُ - الْقِطُّ) بِسُرْعَةٍ وَقَالَ: هَذَا
كُلُّ مَا أَرَدْتُ أَنْ (أَحْصَلَ - أَرْجِعَ - أَسْمَعَ) عَلَيْهِ.

Annex B: Sample Design and Weighting

This annex presents additional details about the sample design for the Egypt 2013 EGRA study.

Stage 1: The Selection and Weighting of Sample Schools

The Ministry of Education's Information Center provided a list of all MOE primary schools in the nation, totaling 15,349 schools. Of these, 234 schools were removed from the list because they were located in North Sinai, deemed too insecure for the assessment teams to conduct assessments. A total of 15,115 primary schools remained in the final population, from which a study sample was drawn. The 15,115 schools included a total enrollment of 1,654,157 Grade 3 students.

Before drawing the random sample of schools to be included in the study, the 15,115 schools were stratified by 5 regions (West Delta, East Delta, Metro Cairo, Middle Egypt, and Upper Egypt). From each of the five regions, 40 schools were randomly selected, to allow for maximum statistical power within each region. This gave an initial sample size of 200 schools.

Within each region, schools were sorted by *muderiya*, *idara*, and Grade 3 enrollment. Schools were then selected with probability proportional to their Grade 3 enrollment. For each selected school, two replacement schools were selected, to be used if the sampled school could not be visited. A total of three schools (1.5% of sample) were replaced when the team of assessors at the school discovered that the test had been leaked to the school in advance of their field visit. Early grade students who have been coached and previously exposed to a specific EGRA form are always discerned. The final sample was 199 tested schools. One school was lost from the sample due to technical problems during test implementation.

To make the sample representative of the national population of all MOE primary schools, school weights were calculated as the inverse of the selection probability of the school (*Weight1*, Stage 1 selection) and then scaled to the total number of schools for each region. **Table B1** shows that the weighted counts and percentages of the sampled schools in each region are, in fact, representative of the population.

$$\text{Weight}_{School} = \text{Weight1} \cdot \text{Scaled School Weight}$$

Where: $\text{Weight1}(s, i)$ represents the weight of the first stage of selection.

$$\text{Weight1}(s, i) = \frac{[\text{Total Number of Grade 3}] \text{ in Region}(r)}{[\text{Number Grade 3 in Selected School}(i)]}$$

$$\text{Scaled School Weight}(s) = \frac{[\text{Sum of Weight1 of All Schools}] \text{ in Region}(r)}{[\text{Total Schools in Region}(r)]}$$

$$r = 1 \text{ to } 5 \text{ Regions}$$

$$i = 1 \text{ to } 199 \text{ Selected Schools}$$

Table B1: Regional Distribution of Schools in the Total Population and Sample

Regions	Population		Sample		
	Total number of schools	Percentage of schools (%)	Sampled number of schools	Weighted number of schools	Weighted percentage of sampled schools (%)
West Delta	4,246	28.1	40	4,246	28.1
East Delta	3,499	23.1	40	3,499	23.1
Metro Cairo	2,075	13.7	40	2,075	13.7
Mid Egypt	2,914	19.3	39	2,914	19.3
Upper Egypt	2,381	15.8	40	2,381	15.8
Total	15,115	100	199	15,115	100

*Population counts are based on MOE data after removing schools in North Sinai.

Stage 2: The Selection and Weighting of Sample Students

The second stage of sample selection was the random stratified selection of students to be tested in each sample school. Grade 3 students were stratified by gender prior to selection and were selected with equal probability. For each sample school, the Assessment Team obtained complete lists of all enrolled Grade 3 students prior to the field visits. The names of 10 boys and 10 girls were randomly selected from the list for each school. Arriving at the school on the day of the assessment, each team randomly selected 5 boys and 5 girls from this list to assess, for a total of ten 10 assessments in each school. Students absent that day were replaced by another student of the same gender from the pre-selected random sample for that school. In single-gender schools, 10 students of that gender were tested. The random sample included 5 single-gender primary schools (all of them girls' schools).

The student weights were calculated by multiplying the school weight by the probability of selecting the student in the given school. This was then multiplied by the student scaled weights to guarantee that the sampled students were representative of the population at the national level.

$$\text{Student Weight}(g, s) = \text{Weight} * \text{Weight2}(j) \cdot \text{Student Scaled Weight}(t)$$

Where: Weight2 (g,s) represents the weight of the second stage of selection: student by gender (g) within the selected school (s)

$$\text{Weight2}(g, s) = \frac{\text{Total Number of Students by Gender}(g, s)}{\text{Sampled Number of Students by Gender}(g, s)}$$

Scaled Student Weight(r, g)

$$= \frac{[\text{Sum of Weights of Selected Students}] \text{by Region}(r), \text{Gender}(g)}{[\text{Total Number of Students}] \text{by Region}(r), \text{Gender}(g)}$$

The gender distribution of the population and sample are shown in **Table B2**. In the total population of enrolled Grade 3 students in the 2012/13 school year, boys outnumber girls (51.9% to 48.1%). But girls were more numerous in our sample, owing to the random inclusion of 5 all-girl primary schools in the school sample. Consequently, males were weighed slightly higher (52%) than the females (48%) to balance the gender distribution in the sample with the actual Grade 3 r population.

Grade 3 male representation by region can be seen in **Table B3** and Grade 3 female representation by can be seen in **Table B4**.

Table B2: Gender Distribution of Grade 3 Students in the Population and Sample

Gender	Population		Sample		
	Total Grade 3 Students	Percentage of Grade 3 Students (%)	Sampled number of Grade 3 Students	Weighted number of Grade 3 students	Weighted Grade 3 students (%)
Males	857,712	51.9	976	857,712	51.9
Females	796,445	48.1	1,016	796,445	48.1
Total	1,654,157	100	1,992	1,654,157	100

*Population counts are based on MOE data after removing schools in North Sinai.

Table B3: Distribution of Grade 3 Boys in the Regional Populations and Samples

Region	Population		Sample		
	Total Grade 3 Males	Percentage of Grade 3 Males (%)	Sampled number of Grade 3 Males	Weighted number of Grade 3 Males	Weighted Grade 3 Males (%)
West Delta	215,956	25.2	189	215,956	25.2
East Delta	156,657	18.2	196	156,657	18.2
Metro Cairo	222,820	26.0	194	222,820	26.0
Mid Egypt	162,794	19.0	191	162,794	19.0
Upper Egypt	99,485	11.6	206	99,485	11.6
Total	857,712	100	976	857,712	100

*Population counts are based on MOE data after removing schools in North Sinai.

Table B4: Distribution of Grade 3 Girls in the Regional Populations and Samples

Region	Population		Sample		
	Total Grade 3 Females	Percentage of Grade 3 Females (%)	Sampled number of Grade 3 Females	Weighted number of Grade 3 Females	Weighted Grade 3 Females (%)
West Delta	203,459	25.5	189	203,459	25.5
East Delta	149,011	18.7	196	149,011	18.7
Metro Cairo	208,478	26.2	194	208,478	26.2
Mid Egypt	143,754	18	191	143,754	18
Upper Egypt	91,743	11.5	206	91,743	11.5
Total	796,445	100	976	796,445	100

*Population counts are based on MOE data after removing schools in North Sinai.

Note about Precision Estimates:

The sample's overall proportion of the population is not relevant in a study with a large population; for instance, in this study where our population is 1,654,157 Grade 3 students and our sample size is 1,992 (0.12%) of the population. The sample size compared to the population is not relevant because regardless of how large our population is, a sample size of 1,992 students provides us with extremely high statistical precision. For example, a 95% confidence band width of ± 3.5 is considered

an acceptable precision for oral reading fluency (ORF), so with a mean ORF score of 21.9, we would say that a 95% confidence interval of (18.4, 25.4) is acceptable.

Figure B1 compares the accepted 95% confidence interval with the actual 95% confidence interval (20.2, 23.5) for the mean ORF score of 21.9. As we can see in the figure, the actual 95% confidence interval is smaller (or “tighter”) than the acceptable 95% confidence interval; thus a sample size of 1,992 students provides more precise estimates than the acceptable precision level.

Figure B1: Accepted and Actual 95% Confidence Interval, Oral Reading Fluency Subtask

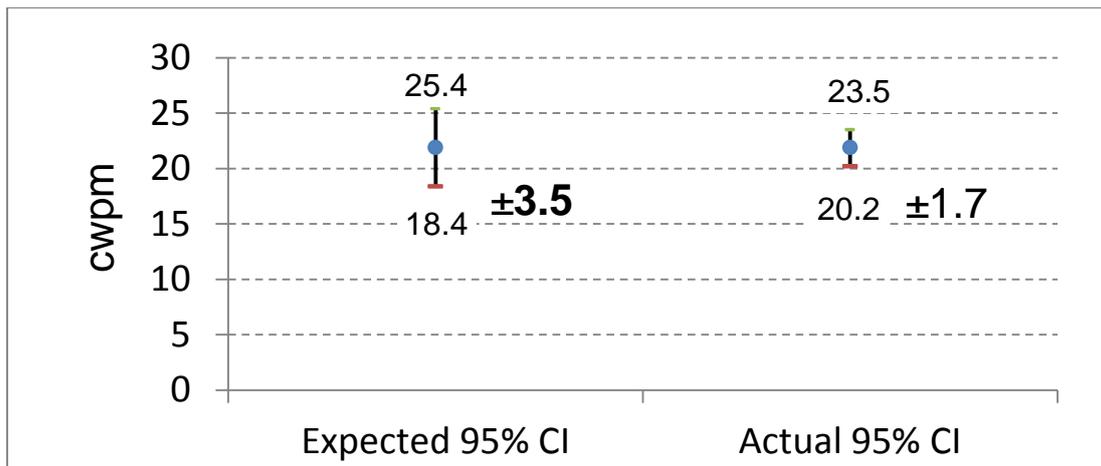


Table B5: EGRA Means and 95% Confidence Intervals

EGRA Subtask	n	Mean	95% Confidence Interval
Correct Letters Sounds Per Minute	1,983	18.8	(17.5, 20.2)
Correct Nonwords Per Minute	1,992	5.9	(5.4, 6.4)
Oral Reading Fluency	1,976	21.9	(20.2, 23.5)
Total correct Reading Comprehension questions	1,985	1.9 (of 6)	(1.8, 2.0)
Total correct Listening Comprehension questions	1,992	3.2 (of 7)	(3.1, 3.4)
Total correct Maze Comprehension selections	1,992	3.6 (of 14)	(3.2, 3.9)

Annex C: Brief Reports of Previous Egypt EGRAs



Early Grade Reading Assessment (EGRA): Egypt

Progress Report – September 2011

- **Project Name:** Girls' Improved Learning Outcomes (GILO)
- **Start Date to Projected End Date:** February 5, 2008 to September 30, 2012
- **Contractors:** Prime: RTI International; Subcontractors: World Education, CID Consulting, Keys to Effective Learning, Infonex
- **Implementation Vehicle:** U.S. Agency for International Development (USAID)/Egypt
- **Coverage in Egypt:** 166 schools in Governorates of Beni Suef, El-Fayoum, El-Minia, and Qena
- **Description of Purpose:** To support the Egypt Ministry of Education to improve the quality of primary education and the teaching of reading in Arabic in the early grades. To demonstrate the impact of highly-focused, data-based approaches to reading instruction in the early grades. The approach is based on evidence on best practice from around the world.
- **Study design:** The project used a randomized controlled design, with non-overlapping cohorts of Grade 2 students. The first (2009) cohort included 444 students from the intervention schools and 465 from the control schools. The second (2011) cohort comprised 574 students from the same intervention schools and 635 from the same control schools. Teachers in intervention schools received an average of 42 hours of professional development and implemented the Early Grade Reading Program for 6 months (October – January, March – April) in the 2010/11 school year.
- **Progress through September 2011:** Student outcomes on three key measures of the Early Grade Reading Assessment showed very significant improvement in intervention schools. On *Syllable Reading*, the mean number of syllables correctly read

in one minute increased from 9.76 to 28.47 syllables (+192%) after six months of program implementation. The change in control schools on this same Syllable Reading measure was +18% -- from a mean score of 8.55 correct syllables per minute in 2009 to 10.10 correct syllables in 2011. This improvement in control schools is consistent with the delayed implementation of EGRA in 2011 (April-May) versus 2009 (January-February).

In *Word Reading*, Grade 2 students from the intervention schools showed a 111% improvement in the mean number of words read correctly in one minute: from 7.35 correct words in 2009 to 15.50 correct words in 2011. The comparable improvement in control schools was 34% -- from the low mean score of 5.56 correct words in 2009 to 7.45 correct words in 2011.

These results, and the improved mean scores in *Oral Reading Fluency*, are presented in the following table:

EGRA Measures	Mean Scores, Intervention Schools		% Change
	2009	2011	
Syllable Reading	9.76	28.47	+ 192%
Word Reading	7.35	15.50	+ 111%
Oral Reading Fluency	11.09	21.14	+ 91%



• **How the results were achieved:** The project EGRP Team, supported by the Ministry of Education Working Group comprised of Arabic Language Specialists, Curriculum Specialists, and renowned senior linguists, developed an Early Grade Reading Strategy and Package for teacher training. The Strategy and Package included the following essential elements:

- **Established time for EGRP:** The Ministry of Education declared 20 minutes each day for teachers to implement the Early Grade Reading Package.
- **Explicit and direct lessons:** Teachers used a manual with instructional routines for the 28 Arabic letters, student work sheets and sequence.
- **Phonics resources:** Teachers infused an Alphabet Flip Book into lessons for demonstrating letter sounds of at the beginning, middle and end of a word.
- **E-learning resources:** Teachers were provided with several E-learning products demonstrating Active Learning Strategies, Sounds of the Alphabet, and Big Book. Each product was carefully researched, developed, piloted and revised.



- **Gender sensitive environment:** Teachers practiced strategies to balance positive learning experiences for both boys and girls

through: taking turns, giving 10 seconds to answer a question, seating girls at the front of the room, and selecting resources with girls and boys in similar roles.

- **Intense training and supervision:** Grade 2 teachers were provided with 12 days of face-to-face training and regular in-school visits that included coaching and supervision. Teachers met once a week to share teaching experiences, lessons learned and challenges.
- **School-Based Training & Evaluation Unit support:** Senior teachers and heads of the school-based Training & Evaluation Units organized teacher training in the school, providing resources, mentoring and coaching.
- **Summer Reading Program:** A number of intervention schools conducted short Summer Reading Programs lead by teachers and parents. Students practiced reading skills through a variety of genres, storytelling, role play, puppetry, games, songs, and arts and crafts.
- **A supportive USAID/Egypt Mission:** The USAID Mission in Egypt was supportive of a highly focused strategy based on measurement and empirical research and facilitated field visits and opportunities for the Ministry of Education, international visitors and USAID colleagues to observe the achievement in classrooms.

Next steps: The Ministry of Education requested GILo support for an expansion of the EGRP to all Grade 1 classrooms in the 4 Project Governorates from September 2011. GILo will support the implementation of an Early Grade Reading Assessment in Grade One in two other governorates – El-Beheira and Cairo -- in October 2011. At Ministry of Education request, GILo will support the national roll-out of EGRP in Grade 1 in all remaining governorates in FY2012.

For more information, contact:

Hala ElSerafy, COTR GILo Project
USAID Egypt
helserafy@usaid.gov

تحسين الأداء التعليمي للبنات

Girls' Improved Learning Outcomes (GILO)



A Reading Snapshot:

Results of the Early Grade Reading Assessment in El-Beheira and Cairo Governorates

June 2012



USAID | EGYPT
FROM THE AMERICAN PEOPLE



جمهورية مصر العربية

وزارة التربية والتعليم

Table of Contents

Introduction	45
Implementation Overview	46
Results of the Two EGRAs	46
Mean Scores of EGRA Sub-Tests:	46
Levels of Performance on EGRA Sub-Tests:.....	49
Conclusion	50
Annex 1: Implementing the October 2011 EGRA and Key Milestones of EGRA in Egypt	52

A Reading Snapshot: Results of the Early Grade Reading Assessment in El-Beheira and Cairo Governorates

Introduction

At the request of the Ministry of Education (MOE), the Girls' Improved Learning Outcomes (GILO) project of USAID Egypt provided assessor training, technical and funding support for a baseline implementation of early grade reading skills among Primary 1 students from 60 schools in El-Beheira and Cairo governorates. This baseline assessment, conducted in October 2011 at the start of the 2011/12 school year, had two purposes: i) to demonstrate to MOE satisfaction the general utility and baseline results for selected metropolitan Cairo and Lower Egypt idaras of using the same Early Grade Reading Assessment (EGRA) tool applied by GILO in 2009 and spring 2011 in three Upper Egyptian governorates, and ii) to establish the institutional capacity of the MOE to conduct early grade reading assessments.

This report presents the key findings of this October 2011 reading assessment in Primary 1. These findings are compared with Primary 2 results of the February 2009 baseline implementation of EGRA conducted by GILO in El-Fayoum, El-Minia and Qena governorates. Strictly speaking, the results are not comparable: the two assessments were conducted with different grades and at different times of the school year. Yet both are baseline assessments conducted before implementation of an enhanced MOE program for improved early grade reading. And both assessments similarly applied the same EGRA tool. It is useful that MOE and USAID Egypt appraise the Primary 1 results from sample idaras in El-Beheira and Cairo alongside the Primary 2 results from provincial urban and rural schools in these Upper Egyptian governorates.

The two idaras sampled in each of El-Beheira and Cairo governorates were selected by MOE. This sample is not representative of either governorate. The results are not generalizable to all Primary 1 students in Cairo and El-Beheira.¹⁴ The findings are, however, a valid “snapshot” of Primary 1 reading skills across a sufficiently large number of schools to satisfy the purposes of this EGRA. The findings also validate the results of the EGRA 2009 baseline assessment conducted in Upper Egypt.

¹⁴ EGRA 2009 was likewise not strictly representative of all Primary 2 students in El-Fayoum, El-Minia and Qena governorates. The sampled idaras and schools in February 2009 were not rigorously selected as a random, representative sample of all schools in these idaras and governorates. The students were, however, a representative sample of all Primary 2 students in their schools, as were the sampled Primary 1 students in the October 2011 EGRA.

Implementation Overview

MOE conducted this October 2011 implementation of the Arabized Early Grade Reading Assessment (EGRA) tool with a stratified random sample of 1098 Primary 1 students selected from two idaras in each of El-Beheira and Cairo governorates. The sample was stratified by gender: approximately equal numbers of boys (536) and girls (562) were randomly selected and tested. The sample was drawn from 60 primary schools: 10 schools from each of Masr al-Qadima and Sharq Medinet Nasr idaras in Cairo and 20 schools in each of Kafr al-Dawar markaz and Abu Homs idaras in El-Beheira.

Each student was assessed individually by trained MOE assessors using identical EGRA forms – the same form applied in the EGRA 2009 baseline. A detailed description of the October 2011 implementation of EGRA and general overview of the history of EGRA in Egypt are provided in Annex 1.

Results of the Two EGRAs

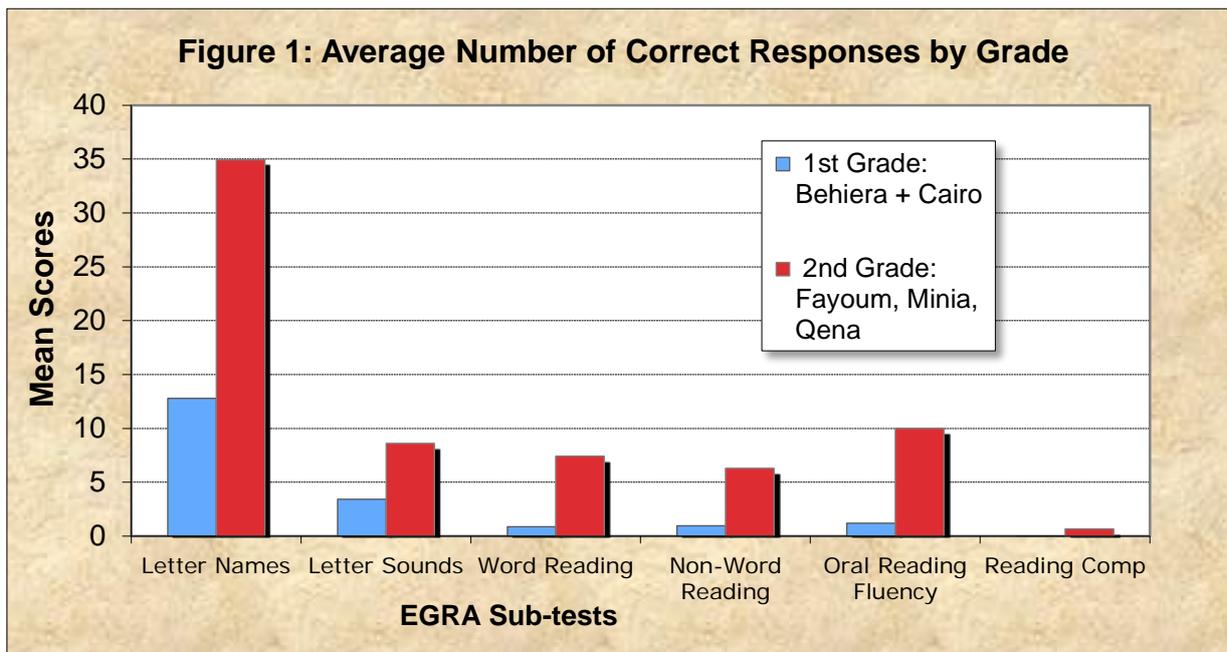
The results of the October 2011 EGRA with Primary 1 students in El-Beheira and Cairo governorates are most meaningful when placed alongside the February 2009 EGRA results for Primary 2 students in El-Fayoum, El-Minia and Qena governorates. Both are baseline assessments, conducted prior to implementing an enhanced program of instruction in early grade reading in these grades. Their broadly similar results belie the significance of the assessments being conducted at times of the school year and different grades.

Mean Scores of EGRA Sub-Tests:

Table 1 and Figure 1 below present the average (mean) scores of Primary 1 and Primary 2 pupils on each of six (6) EGRA sub-tests. The sub-tests in both EGRAs were identical and both assessments were administered similarly. The mean scores are the average number of correct answers on each of these timed (one minute) sub-tests.

Table 1: Mean Scores for EGRA Sub-tests

EGRA	Letter Name Fluency	Letter Sound Fluency	Word Reading Fluency	Non-Word Reading Fluency	Oral Reading Fluency	Reading Comp
Primary 1 Beheira + Cairo	12.8	3.4	0.9	0.9	1.2	0.0
Primary 2 Fayoum, Minia + Qena	34.9	9.1	6.4	5.5	10.0	0.7



The results of these two EGRAs are broadly similar in their *pattern of results*. Both Primary 1 and Primary 2 students did best on the Letter Names sub-test with significantly less success on the other sub-tests. As expected, Primary 2 students out-performed Primary 1 students. Let's consider each of these sub-tests in turn:

- Letter Name Fluency:** In October 2011, this sample of Primary 1 students could – on average – correctly read 12.8 letter names in one minute. At mid-year, the average Primary 2 student in the February 2009 sample performed significantly better, correctly reading 34.9 letter names in one minute. This is an average improvement of 22 correct letter names from Primary 1 to Primary 2. That both grades performed significantly better on this sub-test than any other EGRA measure confirms the priority given to letter name instruction in reading instruction in the MOE curriculum *before* implementing the enhanced program of Arabic reading instruction: the Early Grade Reading Program (EGRP).
- Letter Sound (Syllable Reading) Fluency:** Table 1 and Figure 1 results on this sub-test are clear and consistent for both grades. Average performance in reading letter *sounds* was *markedly less* for both Primary 1 and Primary 2 students than their average performance in reading letter names. Indeed, the average performance of Primary 2 students in letter sound (syllable reading) fluency at mid-year (9.1 correct letter sounds) was not greatly better than Primary 1 students at the start of their school year (3.8 correct letter sounds). This is an average improvement of just 5.7 correct letter sounds from Primary 1 to Primary 2 – much less than the average improvement of 22 correct letter names in one minute (*see above*).

What is most striking about this result is that the results of other EGRA sub-tests more closely mirror the lagging performance of both grades on letter sounds than the significantly improved performance of Primary 2 students in letter name fluency. Primary 2 students markedly out-performed Primary 1 students in letter name fluency. But they were not reading significantly better than Primary 1 students. This is clear from the other EGRA sub-tests.

- **Familiar Word Reading:** This EGRA sub-test measures students' reading of familiar words in one minute. On average, Primary 1 students could correctly read less than 1 word. At the start of their first year of school, this result might be expected. Primary 2 students performed better on this sub-test, correctly reading an average of 6.4 words in a minute. The Primary 2 result is, however, a weak performance. The target average score for this EGRA sub-test in Primary 2 should be in the range of 16-24 correct words per minute.
- **Non-Familiar Word Reading:** This sub-test measures students' ability to read "words" that they have not previously heard or seen. The words included in this sub-test are non-words that do not actually exist in Arabic. The performance of both Primary 1 and Primary 2 students on this test mirrored their performance in reading familiar words (*see above*). Primary 1 students correctly read 1 non-word in a minute. Primary 2 students read 5.5 non-words correctly.
- **Oral Reading Fluency:** For this EGRA sub-test, students orally read a short passage and were scored on the number of words correctly read in one minute. The average score for Primary 1 students – 1.2 correct words – mirrored their performance in familiar and non-familiar word reading. The average performance of the sample of Primary 2 students – 10.0 correct words per minute – was superior to their performance on either of the word sub-tests. Higher average scores in oral reading fluency in Primary 2 are expected. In a passage, some words (e.g. pronouns) will be very familiar and easy to read and other words may be more readily identified by the passage context. But the target average score for this EGRA sub-test in Primary 2 should also be in the range of 16-24 correct words per minute.
- **Reading Comprehension:** Very few Primary 1 students could correctly answer any of the five comprehension questions after the passage reading. Almost 98% of Primary 1 students could not answer any of the comprehension questions. And the 2% of Primary 1 students who performed well on the other EGRA tests could answer only 1 comprehension question correctly.

The performance of Primary 2 students in reading comprehension was also weak: their mean score for reading comprehension was just 0.65 questions answered correctly. Over 70% of Primary 2 students could not correctly answer any of the five comprehension questions. Many of these students can mechanically read words but without comprehension. *See also page 6.*

Like the Primary 2 results from EGRA 2009 in Upper Egypt, the sample Primary 1 girls from El-Beheira and Cairo idaras out-performed boys on all sub-tests of the October 2011 EGRA. The gender difference in reading performance is statistically significant.

Levels of Performance on EGRA Sub-Tests:

Average scores are a useful measure of central tendency in student performance. But it is also useful to examine the *range* or *different levels* of performance of students on the first four (4) EGRA sub-tests. The tables below compare the levels of performance of Primary 1 and Primary 2 students from the October 2011 and February 2009 EGRAs respectively.

Table 2: Comparing Primary 1 and Primary 2 Performance on Letter Names and Letter Sounds

	Correct Letter Names per minute		Correct Letter Sounds per minute	
	Primary 1	Primary 2	Primary 1	Primary 2
27+ Correct Letters	16.9%	58.1%	2.8%	12.4%
16-26 Correct Letters	14.6%	13.4%	5.1%	10.6%
1-15 Correct Letters	37.0%	20.9%	19.7%	26.8%
<u>NO</u> Correct Letters	31.5%	7.6%	72.4%	50.1%
Mean Number of Correct Letters	12.8	34.9	3.4	9.1

- Fluency in Letter Names and Letter Sounds (Table 2): Many Primary 1 students were familiar with the names of different letters before starting primary school. As early as October, nearly one-third of sample Primary 1 students (31.5%) in El-Beheira and Cairo could correctly name 16 or more letters in one minute.¹⁵ The identical percentage, however, could not name *any* letters correctly. *Overall, this sample of Primary 1 students demonstrated a wide range of competency in reading letter names at the start of their first year of primary school.*

Fluency in letter *sounds*, however, was generally low. Only 8% of Primary 1 students could correctly pronounce 16 or more letter sounds. Almost three-quarters (72.4%) of Primary 1 students did not know any letter sounds. The sample of Primary 2 students at mid-year was not markedly better: half of Primary 2 students still did not know letter sounds (Table 2). Only 23% of Primary 2 students could correctly pronounce 16 or more letter sounds.

These low results on letter sound fluency, and the lagging performance of Primary 1 and Primary 2 pupils on the other EGRA sub-tests, have strongly recommended [enhanced reading instruction in phonics](#).

- Fluency in Word and Non-Word Reading (Table 3): The results in Table 3 below confirm that there is little variation in Primary 1 fluency in both word and non-word reading. The large majority of sample students could not read any of the first five words correctly in either test, and the test was stopped. One might speculate that the 13-14% of Primary 1 students who correctly read one or more words (or non-words) attended pre-school education and/or learned to read at home before Primary 1.

¹⁵ Almost 17% of sample Primary 1 students could read letter names quickly and correctly, at the highest level of 27+ correct letters in a minute.

Table 3: Comparing Primary 1 and Primary 2 Performance on Word and Non-Word Reading

	Correct Words per minute		Correct Non-Words per minute	
	Primary 1	Primary 2	Primary 1	Primary 2
25+ Correct Words	0.3%	8.1%	0.3%	4.5%
16-24 Correct Words	1.1%	7.6%	0.5%	8.0%
1-15 Correct Words	13.0%	32.2%	12.1%	30.0%
<u>NO</u> Correct Words	85.6%	52.1%	87.1%	57.5%
Mean Number of Correct Words	0.9	6.4	0.9	5.5

Sample students at the mid-year of Primary 2 were still struggling to read words and non-words. More than half of Primary 2 students still could not read any of the first five words (or non-words) on the EGRA test. The words are not difficult. The lagging performance of Primary 2 students in word and non-word reading is the direct outcome of their lagging performance in letter sound (syllable reading) fluency.

- Fluency in Passage Reading: Primary 1 performance on the oral reading sub-test mirrored their performance on the word and non-word sub-tests reported above. Almost 87% of Primary 1 students could not read any of the first 9 words of the reading passage. The same 13% of students who performed well on other sub-tests could read 1-15 correct words.
- Reading Comprehension: As noted above (page 4), 98% of the sample Primary 1 students could not correctly answer any of the reading comprehension questions. *This includes most of the 12-13% of Primary 1 students who performed well on all of the other EGRA sub-tests.* At mid-year, 70% of Primary 2 students also could not answer any of the comprehension questions. *The mechanical reading fluency of even the better students did not improve their reading comprehension.* Enhanced instruction in reading comprehension is clearly a priority for improved learning outcomes from reading.

Conclusion

The chief finding and conclusion of this analysis is clear. The results of the October 2011 EGRA conducted with a sample of Primary 1 students from El-Beheira and Cairo governorates are fully consistent with and validate the results of the February 2009 EGRA conducted with a sample of Primary 2 students in El-Fayoum, El-Minia and Qena governorates. Both assessments produced the same *pattern of results*. If the Primary 1 students from El-Beheira and Cairo were again tested at mid-year of Primary 2, without the benefit of the Early Grade Reading Program, we would expect their performance to be nearly identical to the Primary 2 results from Upper Egypt.

From this general conclusion, four (4) results are clearly indicated:

1. The EGRA tool is a valid and useful tool for both system diagnosis by MOE of the performance and instructional needs of a representative cross-section of early grade students and periodic “snapshots” of reading skills performance for any large, random sample of early grade students. These two baseline assessments produced similar results from different grades and regions of Egypt. These results validate both the tool and the reality of reading competency among early grade students of MOE public schools. The October 2011 EGRA achieved its first purpose.
2. The consistent pattern of results from these two EGRAs also validates the performance of MOE in implementing the October 2011 EGRA. MOE has established the capacity to implement EGRA, achieving the second purpose of this assessment.
3. There is a strong priority for greater attention to phonics instruction in Primary 1. Attention to teaching letter sounds and active classroom practice in reading words and non-words should equal or exceed the current focus on teaching letter names.
4. Reading comprehension should be the next priority to enhance both reading instruction and learning outcomes from reading, starting in Primary 2. Improved mechanical reading fluency alone will not improve reading comprehension. Without comprehension, readers are not learning from what they read.

Annex 1: Implementing the October 2011 EGRA and Key Milestones of EGRA in Egypt

Implementing the October 2011 EGRA

Implementing the October 2011 EGRA in a sample of Primary 1 schools from El-Beheira and Cairo governorates was the responsibility of MOE. GILO supported MOE by training assessors and assessor team leaders, providing technical support in the orientation, assessor selection, planning, and logistics required for EGRA implementation, funding implementation expenses, performing data entry and data analysis from the original EGRA forms, and preparing this report.

The EGRA form used by MOE for the October 2011 assessment of Primary 1 students in El-Beheira and Cairo was the same EGRA form used by GILO in 2009 and 2011 in Upper Egypt. The form had not circulated and was not available or known to MOE staff or schools in Cairo and El-Beheira prior to the assessor training.

The three (3) days of training provided to 65 MOE assessors and assessor team leaders (52 men, 13 women) for the October 2011 EGRA was led by the same trainer (Dr. Sylvia Linan-Thompson of RTI) who trained Egyptian assessors for EGRA 2009 and EGRA 2011. Assessor training for all three assessments was closely similar. Assessors were trained using the same EGRA form that they applied in schools. The assessor trainees in September 2011 were a mix of MOE teachers, senior teachers and supervisors of Arabic language at primary level. Smaller numbers of senior supervisors and idara administrators were trained as assessor team leaders. The criteria for selecting assessor trainees included strong experience in teaching Primary 1 Arabic, facilitating training of trainers or assessment, and knowledge of good teaching practices and training skills. Assessors were selected from the same two governorates – El-Beheira and Cairo – sampled for this assessment. All assessors conducted assessments in their own governorates. Each governorate organized 4 teams of 6-7 assessors with each team completing assessments in one school per day.

- *Selecting the Idaras and Schools:* Neither the 4 idaras nor the 60 primary schools – 20 schools in Cairo and 40 schools in El-Beheira – selected in these two governorates were chosen in a strictly random manner. The idaras were selected by MOE to represent different levels of general educational status in these governorates. The schools were selected as proximate clusters to facilitate the logistics of in-school assessment.
- *Selecting the Student Sample:* Repeating the design of GILO 2009 and GILO 2011, a stratified random sampling method was used to select participating Primary 1 students. First the enrollment list of students in each school was divided by gender. Then equal numbers of boys and girls were selected randomly from these lists to comprise the sample. If the selected student was absent on the day of the assessment, a substitute student was chosen from a backup list of additional students in that school. The large majority of tested

students (at least 94%) were randomly selected from the student lists.¹⁶ The final sample included 536 boys and 562 girls.

- *Comparing the Cairo and El-Beheira Samples of Students:* The October 2011 EGRA included 501 Primary 1 students from Cairo and 597 students from El-Beheira. The difference in the total number of tested students in each governorate may, however, be less significant than the difference between governorates in the average number of students tested per school. Most of the 40 schools in El-Beheira tested 15 students each, with 9 schools sampling less than 15 students.¹⁷ In contrast, two-thirds of the sampled Cairo schools tested 25-29 students each; one-third of schools tested 20-24 students. Smaller numbers of students per class could bias the sample in favor of better-performing students. One cannot say definitively that more capable readers were over-represented in the El-Beheira sample of students, or that the schools or idaras selected in El-Beheira were generally stronger than the Cairo sample. *But the sample of Primary 1 students from El-Beheira significantly outperformed the Primary 1 sample from Cairo governorate.* The partial results of one sub-test alone are sufficient to demonstrate this difference between governorates:

- Percentage of Sampled Students Unable to Read Any Letter Names Correctly:

Cairo: **42.3%** El-Beheira: **22.4%**

This finding should correct any presumption that the Cairo sample outperformed El-Beheira students. The opposite is true and may well be valid and representative of the sampled idaras.

The October 2011 EGRA was implemented similarly to EGRA 2009 and EGRA 2011. Each assessment was conducted individually. For each of the timed sub-tests (letters, words, non-words and passage reading), if the student did not correctly read *any* of the letters or words on the first line of that sub-test, the sub-test was stopped and scored as “zero”. Most Grade 1 assessments were completed in 15 minutes.

Key Milestones of EGRA in Egypt

Research Triangle International (RTI) introduced the Early Grade Reading Assessment to Egypt as a pilot activity of the Girls’ Improved Learning Outcomes (GILO) project, for which RTI is the prime contractor. With technical support from Arabic language experts in UAE and Egypt, the EGRA form was adapted into Arabic and piloted with 90+ students in two Cairo schools in summer 2008.

¹⁶ Of the 1098 Primary 1 students assessed for the October 2011 EGRA, 43 were outside the sample lists. The forms for an additional 26 students were missing this information, i.e. they did not indicate if the students were in or outside the sample.

¹⁷ In three El-Beheira schools, small samples of 10, 9 and 4 students were tested.

EGRA 2009: Successful implementation of this pilot prompted USAID to approve a baseline implementation of EGRA in 58 schools in three (3) Upper Egypt governorates – El-Fayoum, El-Minia and Qena – receiving GILO training, technical support, and educational furnishings and equipment. This baseline assessment, conducted in February 2009, included 28 schools directly supported by the GILO project and 30 “control” schools from other idaras of these same governorates. EGRA 2009 comprised stratified random samples (stratified by gender) of students from Primary 2, 3 and 4 grades. The same EGRA form was used for all 3 grades and equal numbers of students were sampled in each grade. The results were analyzed and reported for all schools together (GILO-supported plus control schools) by grade.

GILO’s pilot Early Grade Reading Program (EGRP): GILO presented the results of EGRA 2009 to USAID Egypt and MOE in October 2009. Following that presentation, MOE and USAID requested that GILO design and implement a program of enhanced instruction in reading for Grade 2 in project-supported schools. From early 2010, GILO was supported by the MOE Working Group for Early Grade Literacy to design and prepare a teacher’s manual and supplemental learning materials for enhanced Arabic instruction in phonics for Primary 2. GILO began teacher and supervisor training, coaching and distribution of prototype training materials for EGRP in September 2010. Active coaching support, follow-up and additional training in EGRP continued for Primary 2 teachers of selected GILO-supported schools through April 2011. With additional expert input and direct implementation experience, the prototype EGRP manuals were improved and retargeted to Primary 1 students.

EGRA 2011: GILO implemented EGRA for the second time in April-May 2011. This EGRA included all 28 schools in which GILO had implemented both EGRA 2009 and the Early Grade Reading Program from September 2010 to April 2011. EGRA 2011 also included Primary 2 students from all 30 control schools that participated in EGRA 2009. The purpose of EGRA 2011 was to measure the reading outcomes after seven months’ initial implementation of the Early Grade Reading Program among Primary 2 students. The results showed very significant improvements in Primary 2 reading in GILO-supported schools versus the control schools.

MOE’s national Early Grade Reading Program (EGRP): Immediately following the positive results of EGRA 2011, MOE requested GILO support for nationwide implementation of the Early Grade Reading Program in Primary 1 of all MOE public schools in all governorates during the 2011/12 school year. MOE also requested GILO support to implement the **October 2011 EGRA** as a future benchmark against which MOE might subsequently measure progress in improved reading outcomes in Grade 1. A separate GILO report of the national roll-out of the Early Grade Reading Program in Grade 1 is planned in fall 2012.