# Promoting Literacy Through Cognitive Science: Experimental Evidence from the United Arab Emirates

Endline Report

Anett John and Todd Pugatch[1]

December 15, 2025

**Research Associates:**
Mennatallah Badr
Marc Nabil Georgy

[1]Author emails: John: a.john@bham.ac.uk; Pugatch: tpugatch@buffalo.edu. This experiment is registered at the AEA RCT Registry (AEARCTR-0014676). We thank the Al Qasimi Foundation for Policy Research, particularly Gehad Al Najr, Munirah Eskander, Jessica Maluch, and Reham Zahran for their support and guidance. We thank Helen Abadzi, Mei-Ching Lien, and Serena Zadoorian for many helpful conversations about this project. J-PAL MENA coordinated data collection and analysis for this evaluation. We thank Malak Koura, Mennatallah Badr, and Marc Nabil Georgy for excellent research assistance. Nayera Adly Husseiny contributed to sections of this report. Maha Balbaa and Mariam George provided project support.

# Contents

# 1 Introduction

A significant challenge for Arabic-speaking students, particularly those from disadvantaged backgrounds, lies in the stark differences between their home dialects and Modern Standard Arabic (MSA), the language of instruction. This phenomenon, known as Arabic diglossia, often leads students in the Arab world to perceive MSA as a foreign language (Eviatar and Ibrahim 2014), diminishing their motivation and engagement in learning. The Iqra curriculum, named after the Arabic word for "read," addresses this issue by incorporating insights from cognitive science into language acquisition. It. Iqra is designed to leverage perceptual and statistical learning, helping students detect visual patterns and probabilistic structures in MSA that they would not naturally acquire at home. Through practice and systematic repetition of predictable patterns, the program helps transfer knowledge to implicit memory, enabling the development of automaticity, a critical skill for fluent reading (Abadzi and Martelli 2014, Abadzi 2017). Initial pilot studies in the United Arab Emirates (UAE) demonstrated promising results, with students showing improved literacy skills and reduced errors in reading (Eckert, Wilson, Abadzi and Jeon 2020).

We evaluate a randomized controlled trial (RCT) conducted in Ras Al Khaimah, UAE, in academic year 2024-2025 to measure the impact of Iqra on literacy outcomes. We also examine the cognitive processes hypothesized in its theory of change. This evaluation aims to generate rigorous evidence to inform a possible regional scale-up of the program. Aligned with the UAE Ministry of Education's ongoing efforts to strengthen Arabic reading fluency for early-grade students, Al Sheikh Saud bin Saqr Al Qasimi Foundation for Policy Research (AQF) has supported the development and pilot implementation of the Iqra program.

Our main findings are:

1. **The Iqra program made a large and significant difference to teaching practices in the classroom**, resulting in the use of distinct pedagogical techniques and materials in Iqra classrooms compared to the control classrooms. Program implementation went largely as planned.

2. **The program significantly improved Arabic literacy across key domains, including letter identification, word reading, and reading fluency.** The program's impact compares well with other literacy and education interventions evaluated rigorously in a diverse set of global contexts (Evans and Yuan 2022).

3. **No significant effects were observed on secondary mechanisms - neither positive nor negative.** In particular, Iqra neither crowds out nor complements performance in math, nor does it inherently build higher attention capacity. Despite

some parental engagement components, Iqra also did not meaningfully improve children's home literacy environment.

# 2 Methodology

## 2.1 Description of the Iqra Curriculum

The Iqra program is an early-grade curriculum developed in 2017 by Dr. Helen Abadzi in collaboration with AQF. Designed to enhance students' reading speed and accuracy, the curriculum is grounded in contemporary linguistics and cognitive science. The curriculum has aspired to address the challenges that both native Arabic-speaking and non-native children face when learning to read Arabic. The Arabic language has a complex script and grammar, which often contributes to slower reading rates. For over a decade, international assessments such as the Progress in International Reading Literacy Study (PIRLS) have consistently revealed that Arabic-speaking students perform notably lower in reading, mathematics, and science than their global counterparts. All participating Arabic-speaking countries demonstrate scores below the global average. Country-level data further highlights this gap; for instance, the 2018 Programme for International Student Assessment (PISA) results in the UAE showed significant disparities in math, science, and reading scores between students in English-medium and Arabic-medium schools. A major contributing factor is a deficiency in oral reading fluency, which is critical for mastering all subjects. These early reading deficits exacerbate the achievement gap over time.

Iqra employs a systematic and evidence-based approach to teaching Arabic reading skills. The curriculum provides structured, incremental instruction, prioritizing automaticity and rapid decoding to ensure students can process Arabic text effortlessly. Extensive practice, systematic feedback, and clear progress benchmarks are integral to the program. The program integrates active pedagogy techniques such as paired, echo, and choral reading. Assessments track student progress and inform instructional adjustments. By the end of grade 1, students are expected to achieve a reading speed of 40 words per minute.

The Iqra textbook is distinguished by its systematic, letter-by-letter instruction methodology. Each lesson introduces a single letter, beginning with simpler, non-connecting letters and progressively advancing to more complex ones. This structured progression is designed to enhance visual clarity and facilitate pattern recognition. Each lesson builds on previously introduced content, incorporating short vowels (*harakat*) and letter combinations to reinforce learning. To further support fluency, the textbook employs large font sizes and ample spacing between letters, enabling students to develop automatic recognition of letters and words. This design helps streamline the decoding process, which is

critical for improving reading proficiency. The program is delivered through 40 lessons, each lasting approximately 45 minutes. It is designed to be completed over one academic year, making it adaptable to existing school schedules (Eckert et al. 2020).

## 2.2 Research Design

The evaluation employs a cluster randomized controlled trial (RCT) design, with classrooms as the unit of randomization. AQF partnered with 26 schools in Ras Al Khaimah to implement the Iqra program during the 2024–2025 academic year in their first-grade or Kindergarten 2 (KG2) classes. The sample includes schools from diverse educational settings, comprising 7 public and 19 private schools offering curricula spanning Emirati, British, American, Indian and Pakistani systems. Partner schools provided class lists and student names for the relevant grades. While most schools shared lists for all their classes in the selected grades, a few others provided lists for only a portion of their classes to be included in the study. A total of 83 classrooms participated in the study. From these classrooms, we selected a random sample of students to participate in the study.

Following baseline data collection, we randomly assigned the 83 participating classrooms to either the control group or the Iqra program, stratified by school. Within each school, half of the classrooms were allocated to the control group (42 classrooms) and half to Iqra (41 classrooms). Students in the control group continued to follow the standard, compulsory Arabic curriculum provided by the Ministry of Education (MoE). Official guidelines specify MoE Arabic classes should be taught 7 periods per week, 40 to 45 minutes per period, or 4.5 to 5.25 hours of weekly instruction time.

Treated classrooms were assigned to the MoE curriculum, plus the Iqra curriculum. Treated classrooms were assigned Iqra for two class periods per week, or 1 hour 20 minutes weekly instruction time. Iqra content was integrated into a portion of the Arabic class time in the majority of schools, while in some schools, Iqra was taught during extra-curricular time.

Following the randomization, teachers assigned to Iqra participated in training sessions organized by AQF. During these sessions, they were provided with comprehensive materials and guidance to implement the Iqra program in their respective classrooms.

## 2.3 Description of Instruments

We assessed student literacy in Arabic twice: at baseline in September–October 2024, prior to random assignment and teacher training; and at endline in May–June 2025, at the end of the academic year. We randomly sampled up to 15 students from class rosters at baseline. Average sample size per classroom was 10.7, given survey time constraints during school hours, small class sizes, or lack of parental consent. At endline, we re-assessed students from the baseline and randomly sampled additional students to reach a

target of 15 students per classroom. All students included in the sample received parental consent to participate.

Additional data collection instruments included student surveys, teacher surveys, and classroom observations. The next subsections provide details of all instruments.

### 2.3.1 Baseline instruments

The baseline student survey was developed in collaboration with AQF to ensure its suitability for the local context of Ras Al Khaimah. Recognizing the diverse linguistic and cultural backgrounds of the involved students, the survey was made available in three languages: English, Modern Standard Arabic (MSA), and Emirati Arabic dialect. The survey was structured into three main parts: the first part involved obtaining assent from the children. Before administering the Early Grade Reading Assessment (EGRA), enumerators explained the purpose of the study, emphasizing that the assessment was anonymous, would not affect students' grades and aimed to improve literacy learning for children. Each child was given the choice to orally assent to participate or to decline without any consequences. If a child opted not to participate, they were thanked, and the enumerator moved on to the next child.

The second part of the survey focused on measuring literacy skills using the EGRA instrument. We decided to begin with the literacy assessment to ensure that we could capitalize on the children's full attention and focus, recognizing that younger learners tend to have shorter attention spans. Eight different versions of the assessment were created to ensure variation and robustness. Each student was given a handout corresponding to their assigned version. The baseline assessment consisted of four subtasks:

1. **Initial Sound Identification:** This task measured phonemic awareness by presenting students with 10 words orally. The students were asked to pronounce only the first sound of each word. This subtask was standardized across all models, with students receiving the same 10 words in the same order.

2. **Letter Sound Identification:** This subtask assessed alphabet knowledge and decoding skills. Students were presented with a sheet of paper containing 100 letters organized in 10 rows, each containing 10 items arranged randomly according to their assigned version. Students were then asked to articulate the sound of each letter.

3. **Nonword Reading:** This subtask measured decoding skills using a set of 48 nonwords. Nonwords are nonsense words, used to test decoding skills because students cannot use prior familiarity with the word. The sets of nonwords were split into four tables per model, with each table containing four rows of three nonwords. The

formatting of the tables was varied to test the influence of font size and spacing on performance, using the following configurations:

(a) Small font size (similar to a textbook) and standard spacing

(b) Small font size with enlarged spacing, using kashidas (a technique to lengthen and connect letters)

(c) Large font size (26) and standard spacing

(d) Large font size and enlarged spacing using kashidas

Each version presented the tables in a different order, ensuring maximum variation across the dimensions, with only one feature altered in the next table. Additionally, the nonwords in each table were randomly selected and varied across versions.

4. **Oral Reading Fluency:** This subtask assessed reading fluency by presenting students with six sentences, totaling 50 words. If a student's responses for all words in any one sentence were incorrect, the administration of the subtest was discontinued, making it the only subtask with an "early-stop rule." This ensured that students did not continue with the task if they demonstrated significant difficulty early on. The subtask was standardized across all models.

Each subtask was timed for 120 seconds to assess the automaticity of the skill(s) being tested. Successful readers must develop automatic basic reading competencies. Fluency measures not only knowledge but also how well that knowledge is integrated and quickly accessed. Time constraints enable accurate comparisons of fluency between slow and fast readers who might achieve similar scores over different durations. Marking was based on recording the student's response for each attempted item in the subtest, whether correct or incorrect. This approach ensures that each response is treated as a binary variable, capturing either success or error for every attempt.

The third part of the survey was a questionnaire that relied on the self-reporting of students. It included 12 questions, covering their educational background, home literacy activities, and socioeconomic status. Notably, no skip patterns or conditional questions were included, even where it would be natural to do so. For example, if a child reports having no books at home, they are still asked whether they have Arabic books. This approach was intentional, serving as an accuracy check, as mistakes are common even with adult respondents. By avoiding conditionality, we preserved valuable data that might otherwise have been lost. However, this required careful training of enumerators, who had to adapt their framing and tone to gently reframe questions when needed, turning the next question into a follow-up or double-checking the previous one.

### 2.3.2 Endline Instruments

***Student Assessment and Survey***

The endline student instrument closely mirrored the structure and administration of the baseline survey. As in the baseline, enumerators first obtained verbal assent from each child in their preferred language (English, Modern Standard Arabic, or Emirati Arabic dialect). Children were reminded that participation was voluntary, that their answers would remain anonymous, and that the activity would not affect their school grades. If a child declined to participate, they were thanked and the enumerator moved on to the next student.

The literacy assessment at endline followed the same Early Grade Reading Assessment (EGRA) subtasks described in the Baseline Instruments section, with one intentional modification: the *Initial Sound Identification* subtask was removed, reflecting the expected developmental shift from phonemic awareness to more advanced decoding and fluency skills. All other EGRA subtasks were administered exactly as at baseline, with the same timing rules, scoring procedures, and versioning structure. In addition, two literacy subtasks were *added* at endline to capture more advanced reading competencies:

- **Familiar Word Reading:** Students were presented with a grid of high-frequency Arabic words and asked to read each word aloud. This subtask aimed to capture automatic recognition of common vocabulary items and progression from decoding toward sight-word fluency.
- **Reading Comprehension:** Following the oral reading passage, students were asked a series of timed comprehension questions assessing literal understanding of the text. Enumerators recorded each response as correct, incorrect, or "does not know," and continued asking questions until the time limit expired.

To measure children's executive function—specifically attentional capacity and sustained attention—the endline instrument incorporated a visual search task administered twice, once at the start of the assessment and once at the end. This task, closely following Brown, Kaur, Kingdon and Schofield (2025), required students to identify a target shape (a square missing its left border) from a large matrix of similar distractor shapes within one minute. The number of correctly identified target shapes in the first round provides a measure of *attentional capacity*, while the change in performance between the first and second rounds captures *cognitive endurance*, or the ability to sustain focus after approximately fifteen minutes of effortful cognitive processing.

The endline assessment also included a numeracy module to test whether the literacy intervention substituted for or complemented the development of other academic skills. Students were orally administered a timed set of single- and double-digit addition and

subtraction items, adapted from the Level 1 Addition/Subtraction subtask of the Early Grade Mathematics Assessment (EGMA). Students answered as many items as possible in one minute.

Finally, similar to the baseline, the assessment concluded with a student questionnaire covering educational background, home literacy practices, and socioeconomic characteristics. The questionnaire maintained the same design principle of avoiding skip patterns or conditional questions, even when a previous answer logically rendered a subsequent question inapplicable. This approach ensured data completeness and allowed for internal consistency checks. Enumerators were trained to administer these items in a child-sensitive and non-leading manner, reframing questions when necessary while preserving the full structure of the instrument. To save time, students who participated in the baseline survey were asked only the home literacy practices questions; all other items should not have changed or would be unaffected by the intervention.

***Classroom Observations***

Classroom observations were conducted at endline by independent enumerators who were not involved in program delivery. The purpose of the observations was to capture instructional practices, use of materials, and classroom dynamics in Arabic language lessons. All classrooms participating in the study were asked to participate in an observation. Enumerators visited both treatment (Iqra) and control (MoE) classes and documented pedagogical behaviors and materials use through a detailed checklist aligned with the Iqra pedagogical framework and the Ministry of Education curriculum.

The observation tool began with basic descriptive information about the class, including the curriculum being taught, the duration of the session, class size by gender, the presence of a teacher's aide, and the topic of the lesson. Enumerators also recorded whether the teacher and students were using any curricular materials—specifically, Iqra textbooks or Ministry of Education textbooks—and whether the teacher appeared to follow the corresponding textbook in a systematic fashion.

The main body of the instrument consisted of a structured set of binary indicators capturing the presence or absence of specific instructional practices during the observed lesson. These included whether the teacher reviewed the previous lesson, wrote or stated learning objectives, used visual aids (e.g., letter cards, slides, or projected textbook pages), displayed letters, words, and sentences in large and legible font, read aloud with students, demonstrated correct pronunciation, and instructed students to read targeted materials individually at least twice. Additional items assessed whether the teacher presented letter combinations or sentences using the target letter or vowel, provided corrective feedback, supported weaker students, encouraged focus, maintained classroom discipline, grouped students by ability, and integrated Arabic writing. The tool also documented the use of less Iqra-aligned practices—including songs, pictures, stories, and

multimedia—allowing for systematic comparison across pedagogical approaches.

In addition to teacher practices, enumerators recorded student engagement with materials by noting the proportion of students who used Iqra or MoE textbooks during the lesson. The instrument also included open-ended questions that allowed observers to comment on what went well during the lesson and to provide recommendations for strengthening instructional quality.

Data collected through the classroom observation instrument serve as the basis for several implementation outcomes. Our **observed Iqra pedagogy index** is constructed as the mean of binary indicators capturing key Iqra-aligned instructional behaviors observed during the class, such as reviewing the previous lesson, focusing on one new letter or vowel, presenting letter combinations and sentences using the target letter, reading aloud with students, encouraging repeated reading, providing immediate feedback, and supporting weaker students. The index measures the extent to which teachers implement the pedagogical approach promoted by the Iqra curriculum in both treatment and control classrooms.

The **observed Iqra materials index** summarizes the use of Iqra-aligned instructional resources. It includes the proportion of students who used Iqra textbooks during the lesson, a binary indicator of whether the teacher used Iqra curriculum materials, and indicators of effective use of visual aids displaying letters, word combinations, or sentences in large font. This index captures the material environment supporting Iqra-style instruction.

Finally, the **observed MoE pedagogy index** is constructed for classrooms following the Ministry of Education curriculum. It aggregates the proportion of students with MoE textbooks and binary indicators for teacher use of MoE materials, presentation of lesson objectives, and incorporation of songs, pictures, stories, multimedia, and Arabic writing practice. Although designed primarily for MoE classes, the same practices may also appear in Iqra classrooms, enabling comparisons across pedagogical approaches.

***Teacher Survey***

The endline teacher survey was administered by AQF staff during the second half of the academic year. The teacher survey was designed to capture teachers' backgrounds, instructional practices, and perceptions related to Arabic literacy instruction. Because teachers often taught multiple classes within a school, the survey asked separately about each classroom included in the study. Accordingly, the unit of observation for all teacher-survey-based measures is the classroom. For each class taught, teachers reported curriculum usage, instructional practices, pacing, home–school engagement, and exposure to Iqra-related trainings or materials.

The teacher survey began with a background module on personal and professional

characteristics, including gender, nationality, native language status, educational attainment, and years of experience teaching Arabic generally and at the current school. Teachers were also asked whether they had received training on the Iqra program during the academic year.

The main body of the survey captured detailed classroom-level information. For each class the teacher taught, the survey collected enrollment figures, class composition by Arabic language proficiency, timetable information, frequency and duration of lessons, use of teacher aides, and whether the teacher had been teaching the class since the start of the academic year. Teachers also identified the principal and supplementary curricular materials used in each class and the proportion of instructional time allocated to each material. This information was collected separately for two types of classes: (i) Ministry of Education (MoE) Arabic curriculum classes, and (ii) Iqra curriculum classes. When classes followed similar instructional approaches, teachers were allowed to link them to avoid unnecessary repetition.

For each class, teachers were asked whether they introduced letters one by one, the order of letter introduction, the lesson number reached at the time of the survey, and whether an information session with parents had been conducted. In addition, the survey gathered rich information on classroom practices through frequency-based questions covering a wide range of pedagogical strategies. These included reviewing previous lessons before introducing new content, teaching only one new letter or syllable per session, using visual aids and large-font materials, reading aloud with students, presenting letter combinations, encouraging individual reading, restricting reading to previously taught letters or syllables, providing immediate feedback, conducting formative assessments, grouping students by ability, and engaging parents through home-learning activities or tailored support for struggling learners. Questions on less Iqra-aligned practices (such as the use of songs, pictures, stories, or multimedia) were also included.

Finally, the survey included a perceptions module in which teachers reported how important they considered a variety of instructional practices for supporting Arabic literacy. Teachers were asked to prioritise among competing pedagogical approaches, drawing on their professional experience. This module captured perceptions of both Iqra-aligned practices (e.g., reviewing previous lessons, using large-font materials, grouping by ability, providing feedback) and less aligned practices (e.g., incorporating songs, pictures, or multimedia).

The teacher survey provides the foundation for several exploratory implementation outcomes. The **Iqra pedagogy index** aggregates self-reported use of key Iqra-aligned practices, including the order of letter introduction and the frequency of practices such as reviewing previous lessons, teaching a single new letter per session, reading aloud with students, using controlled texts, providing immediate feedback, grouping students

by ability, and engaging parents. The **Iqra materials index** captures use of Iqra curricular resources, including whether the Iqra textbook served as the primary instructional material and the use of visual aids or large-font texts. The **Iqra perceptions index** aggregates teachers' stated views on the importance of practices aligned with the Iqra approach. Together, these indices allow us to construct a structured view of teachers' self-reported instructional practices, curricular materials, and pedagogical priorities across treatment and control classrooms.

## 2.4 Methodology

We follow an analysis plan submitted to the AEA RCT Registry (John and Pugatch 2024). For each learning outcome, we compare students whose classrooms were assigned to treatment versus control using the following ordinary least squares (OLS) regression:

$$y_{ics1} = \alpha_s + \beta Iqra_{cs} + \delta y_{ics0} + \lambda_i + X_i\theta + \varepsilon_{ics} \tag{1}$$

where $i$, $c$, and $s$ index students, classrooms, and schools, respectively; $y_1$ is an outcome measured at endline; $\alpha$ is a school fixed effect; $Iqra$ is an indicator variable for the classroom being assigned to Iqra; $y_0$ is the outcome measured at baseline (where available); $\lambda$ is a vector of fixed effects for assessment version; $X$ is a vector of student characteristics; and $\varepsilon$ is an error term. Because treatment was stratified by school, the school fixed effect isolates random variation in treatment assignment. The student characteristics $X$ include an indicator for Arabic non-nativity, female, and age in months. Including these characteristics controls for any spurious correlations with treatment assignment and increases precision.

Our coefficient of interest is $\beta$, which measures the effect (intent to treat) of assignment to Iqra on the outcome. To account for correlation in outcomes within classrooms, we cluster standard errors at the classroom level.

For classroom-level outcomes, we estimate a regression analogous to equation (1):

$$y_{cs} = \alpha_s + \beta Iqra_{cs} + \varepsilon_{cs} \tag{2}$$

Equation (1) differs from equation (2) in that the unit of analysis is the classroom.

# 3 Data Collection

## 3.1 Baseline Data Collection

A data collection company was contracted to manage the overall baseline data collection process. Their responsibilities included recruiting enumerators who met the required qualifications, ensuring the distribution of materials such as handouts, stickers, and tablets,

coordinating the scheduling of data collection sessions and communicating any issues or concerns that arose during the data collection process, ensuring that we were promptly made aware of any potential challenges. Following the recruitment and preparation of materials, training sessions were conducted for the enumerators to ensure they were fully equipped with the knowledge and skills needed for the survey implementation.

After finalizing the survey and conducting the training sessions, a pilot was carried out in a school that expressed willingness to participate in the program but was ineligible due to having only one first-grade class. The school agreed to host the pilot, which was conducted on September 27, 2024. This pilot was crucial for testing the enumerators' readiness and identifying necessary adjustments to the survey. Based on the findings, the survey was refined, and feedback was provided to enumerators to address areas where improvement was needed, ensuring better performance during the actual data collection phase.

Baseline data collection at schools in the study took place between Monday, September 30 and Wednesday, October 23, 2024. Over this four-week period, teams of enumerators and supervisors visited the sample schools for survey administration. Prior arrangements were made with the schools to ensure that at least two convenient rooms were available for the surveys, and school staff were assigned to escort children to and from their classes. The survey was administered through face-to-face interviews between an enumerator and the student. The student was given a handout corresponding to their assigned model, while the enumerator used their tablet to mark responses and record data. The interview lasted, on average, 20 minutes, from the start of the greeting and assent process to the completion of the student questionnaire. Upon completion of the interview, students were given stickers as a token of appreciation for their time.

### 3.1.1 Challenges

A successful data collection process relies heavily on daily monitoring of field activities to promptly identify and address challenges, ensuring the integrity of the collected data and the validity of results. Several challenges arose during the day-to-day monitoring of field activities. Initially, the data collection process was hindered by limited preparation time, which reduced enumerator training to just two days. Compounding this issue, some trained enumerators dropped out or were excluded due to other commitments, dissatisfaction with pay, or limited proficiency in Arabic, necessitating additional training sessions for replacements. Further disruptions occurred due to unexpected school events, such as trips or inspections, which led to last-minute cancellations of planned data collection days. Similarly, instances of enumerators being unable to attend on scheduled days resulted in cancellations at one or more schools, straining the already limited resources. Consequently, the data collection timeline had to be extended by two weeks to ensure

sufficient coverage for all schools. However, the data collection firm was unwilling to proceed into the third week, necessitating intervention from J-PAL to resolve the issue and ensure continuity. We were able to reschedule and proceed with data collection in most cases, but due to the inability of some schools to accommodate us for rescheduled visits, we were unable to collect the required sample in a few classes.

A recurring challenge was the failure of enumerators to adhere to the randomly generated student lists provided by J-PAL and the unequal allocation of survey time between classes. In one school, the coordinator prioritized bringing students they considered of good caliber rather than following the provided list. Enumerators sometimes mistakenly adhered to the schools' own student lists, while in other instances, schools did not permit enumerators to follow the randomized student lists, citing parental non-consent as a reason. This deviation from the planned selection process disrupts the randomly selected sample of students. In addition, enumerators were not provided with designated rooms for assessments, leading to frequent interruptions. In some instances, surveys were conducted in spaces like the computer lab and clinic, which were shared with other students. The classrooms were sometimes located far from the assessment areas, which added to the time required to bring students. Supervisors often had to personally escort students to the assessment rooms, which diverted their attention from overseeing the quality of data collection. Alternatively, enumerators themselves took on the responsibility of escorting students to and from their classes, which added to their workload. Also, many school coordinators were bringing in students before the previous child had completed their assessment, causing the new students to wait outside the assessment room, leading to unnecessary apprehension.

Nonetheless, these challenges do not threaten the internal validity of the study. At the time of baseline data collection, classrooms had not yet been assigned to the Iqra program or control, making it impossible for schools to systematically select (or avoid) particular students for the Iqra or control samples. Our estimates of program impact remain valid for the sampled students.

### 3.1.2 Accomplishments

Despite the numerous challenges we encountered throughout the baseline data collection process, we were able to swiftly react and overcome obstacles, ensuring that we collected the necessary sample of students. While we faced setbacks such as limited preparation time, enumerator dropouts and logistical constraints, we adapted quickly and managed to conduct new training sessions for new enumerators, ensuring they were prepared to continue the work without significant delays. Despite disruptions like unexpected school events and issues with school coordination, we managed to reschedule and continue data collection in most instances.

When the data collection company expressed reluctance to extend the timeline into a fourth week, we took the necessary steps to ensure that we concluded data collection successfully. This intervention allowed us to push forward and complete the data collection as planned, meeting our sample size goals despite the challenges. Our ability to adapt to changing circumstances, along with the collective effort and determination of everyone involved, was key to successfully completing the data collection process.

## 3.2 Endline Data Collection

While baseline data collection consisted solely of student surveys and assessments, the endline incorporated three complementary data sources: student surveys and assessments, classroom observations, and a teacher survey. Expanding the data collection instruments at endline allowed us to capture not only student learning outcomes but also instructional practices and teacher perceptions, providing a more comprehensive assessment of program implementation and effects.

### 3.2.1 Endline Student Survey

A total of 874 students were surveyed at baseline.[2] Of these, 65 students could not be reached at endline and were classified as attriters. At baseline, 70 of the 83 participating classes had fewer than 15 surveyed students; however, all classes met our self-imposed minimum requirement of at least 8 surveyed students.

To strengthen statistical power, the data collection team expanded the sample at endline and successfully recruited an additional 499 students. As a result, only six classes at endline had fewer than 15 surveyed students.

The final sample sizes are as follows:

- Baseline observations: 874
- Endline observations: 1,309
- Students surveyed in *both* baseline and endline: 809

The same data collection company managed the endline fieldwork. Following the finalization of the endline instrument and enumerator training, a pilot was conducted on April 30, 2025, in a school that was willing to participate but ineligible for the intervention due to having only one Grade 1 classroom. The pilot was essential for testing enumerator readiness, especially given the addition of new sections to the endline survey. Based on

[2]An additional 18 students were surveyed. However, these students were part of the single eligible classroom of non-native Arabic speakers in their school, leaving no counterpart classroom for comparison. We therefore dropped these students from the study.

pilot results, revisions were made to the instrument, and enumerators received targeted feedback.

Endline data collection took place from Thursday, May 1 to Friday, May 30, 2025. During this four-week period, field teams—consisting of enumerators, quality controllers, and supervisors—visited participating schools to administer the student surveys. Interviews were conducted face-to-face. Each student received a printed handout aligned with the survey model to which they were assigned, while enumerators recorded answers electronically on tablets. On average, each interview lasted approximately 30 minutes, including assent and introductory procedures. Students received stickers as a token of appreciation upon completing the survey.

Several challenges emerged during fieldwork. Schools frequently faced coordination and communication difficulties, leading to situations where staff were unprepared or unavailable, requiring rescheduling or additional visits. Limited room availability, school event schedules, and examination periods created access constraints and delayed data collection. Student absenteeism and challenges in obtaining parental consent reduced the ability to reach all baseline students and required persistent follow-up.

Miscommunications regarding permissions, schedules, and student lists further complicated the process, and some schools required intervention from supervisors to secure cooperation. In addition, many students at Asian schools had returned to their home countries by the endline, contributing to higher attrition and reducing continuity between baseline and endline samples. Logistical constraints, including the absence of dedicated assessment rooms, meant that enumerators often conducted interviews in suboptimal environments with frequent interruptions and student movement.

Despite these obstacles, the field teams successfully adapted to on-site challenges, adjusted schedules as needed, and ensured that the required sample size was achieved. The teams completed data collection as planned, reaching the targeted number of surveyed students and maintaining the quality and completeness of the endline dataset.

### 3.2.2 Classroom Observations

Classroom observations served as an additional data source to assess instructional practices and implementation fidelity. Schools were asked in advance to provide weekly schedules for Arabic language classes, indicating whether each class followed the Ministry of Education (MoE) curriculum or the Iqra curriculum. Two independent enumerators—distinct from those who administered the student assessments—were assigned to observe all classrooms across the study to ensure consistency in observation procedures and scoring.

Although we initially planned to observe 83 classrooms, several contextual and logistical challenges during endline fieldwork resulted in six missing classroom observations.

These cases reflect common constraints encountered in school-based data collection:

- **Curriculum completion timing:** In some instances, teachers had already completed both the MoE and Iqra curricula prior to the scheduled visit, leaving no Arabic lessons available for observation.
- **Staffing changes:** In a few cases, teacher resignations occurred shortly before or during the data collection period, making it impossible to observe the planned classes.
- **Administrative constraints:** Some schools declined or restricted observation requests due to internal scheduling pressures, or limited availability of instructional time.
- **Access and coordination challenges:** In certain situations, schools were unable to accommodate observers on the planned dates, and rescheduling was not feasible within the endline data collection window.

These factors illustrate how operational realities—such as staffing turnover, administrative decisions, and curriculum pacing—can affect the completeness of classroom observation data despite careful planning and coordination.

### 3.2.3 Teacher Survey

Data collection for the teacher survey took place between February and July 2025. A staff member of AQF conducted the teacher surveys. The survey included all teachers who had been assigned to the study classrooms at any point during the academic year, even if they had changed classes by the time of the survey. As a result, some classrooms in the data appear to have more than one Arabic teacher recorded under the same curriculum, reflecting routine teacher rotations and mid-year reassignments.

To capture the full structure of teaching assignments, the analysis is conducted at the level of *teacher–curriculum–classroom* combinations, yielding a total of 138 such units. This allows us to account for the fact that a single teacher may teach multiple curricula or multiple classrooms.

The final teacher sample consists of:

- 57 teachers in total,
- 39 teachers who teach Iqra, including 13 who teach only Iqra,
- 44 teachers who teach the MoE Arabic curriculum, including 18 who teach only MoE,
- 26 teachers who teach both curricula.

At the classroom level:

- 39 classes are taught Iqra, although only one receives Iqra as the sole Arabic curriculum,
- 77 classes are taught the MoE curriculum, including 39 classes that receive MoE only.

This distribution reflects the curricular structure in participating schools: Iqra is almost always delivered alongside the MoE curriculum, whereas the MoE curriculum is implemented both independently, in the case of control classrooms, and jointly with Iqra, in the case of treatment classrooms.

Table 1 provides a concise summary of these patterns. It highlights that Iqra rarely appears on its own, only one case, while MoE is widely taught both alone and in combination with Iqra. The table also illustrates how these overlaps generate a substantially larger number of teacher–class–curriculum combinations than the number of individual teachers or classrooms, motivating our choice of unit of analysis later on.

Table 1: Teachers and Classes by Curriculum

| | **Curriculum** | | **Total** |
|---|---|---|---|
| | **Iqra** | **MoE** | |
| Teachers teaching only one curriculum | 13 | 18 | 31 |
| Total teachers teaching each curriculum (not necessarily only one) | 39 | 44 | 57 |
| Classes taught in only one curriculum | 1 | 39 | 40 |
| Total classes taught in each curriculum (not necessarily only one) | 39 | 77 | 78 |
| Total unique teacher–class–curriculum combinations | 49 | 89 | 138 |

# 4 Sample Characteristics and Balance

## 4.1 Sample Characteristics

Table 2 presents descriptive statistics for the endline sample. The sample is gender-balanced, with 50.4% female students. The majority of students (85.5%) are in Grade 1, with the remainder in KG2. Nearly all students (87.8%) are Muslim, reflecting the religious composition of the study region. Approximately two-thirds of students (67.2%) are native Arabic speakers, with the remainder speaking other languages at home, primarily Urdu and other South Asian languages. These characteristics are representative of the diverse student population in Lebanese public schools.

Table 2: Descriptive Statistics

| | Percentage | N |
|---|---|---|
| Female | 50.4 | 1308 |
| Grade 1 | 85.5 | 1308 |
| Muslim | 87.8 | 1308 |
| Native Arabic speaker | 67.2 | 1308 |

*Notes:* This table presents descriptive statistics for the endline sample (N=1,308 students). Percentages represent the proportion of students with each characteristic. The sample includes students from both treatment and control groups who participated in the endline assessment.

## 4.2 Balance Tests

This section presents the results of balance tests of the comparability of treatment and control groups prior to intervention. Table 3, Panel A presents balance tests for baseline EGRA scores, using the sample of students who participated in both the baseline and endline. The overall baseline EGRA score is 0.11 standard deviations higher in the control group than the treatment group, significant at 10%. This difference is driven by the letter identification subtask, for which the control group correctly identifies two more letters, on average, than the treatment group (significant at 5%). This imbalance arose due to chance, because treatment was not assigned until after baseline data collection ended, preventing schools from choosing better (or worse) students from control classrooms to participate. To mitigate the potential for this imbalance to confound endline results, we control for baseline EGRA score (overall and for letter identification) in all analyses of student outcomes. Moreover, any remaining bias from imbalanced baseline EGRA scores will work against finding a positive effect of the program, making our treatment effect estimates a lower bound.

Table 3, Panel B presents balance tests for student and household characteristics from the student survey. These tests include all students who participated in the endline, regardless of their baseline participation. We find two characteristics with marginally significant imbalances (dummies for female and father present at home), close to what one would expect to arise by chance. To further mitigate the potential for confounding, we also include a vector of characteristics (female, native Arabic speaker, father at home, age in months, a dummy for baseline participation) in all analyses of student outcomes.

Table 3: Balance Test

**Panel A: Early Grade Reading Assessment**

| Variable | (1) Control Mean/(SE) | (2) Treatment Mean/(SE) | (1)-(2) Pairwise t-test Mean diff |
|---|---|---|---|
| Baseline EGRA overall score | 0.00<br>(0.07) | -0.10<br>(0.07) | 0.11* |
| Baseline initial sound score | 4.76<br>(0.44) | 4.30<br>(0.52) | 0.46 |
| Baseline letter identification score | 0.12<br>(1.27) | -1.94<br>(1.12) | 2.06** |
| Baseline nonword reading score | -6.00<br>(0.52) | -6.66<br>(0.53) | 0.66 |
| Baseline reading fluency score | -6.44<br>(0.64) | -7.18<br>(0.63) | 0.74 |
| **N** | 400 | 409 | 809 |
| **Clusters** | 42 | 41 | 83 |

**Panel B: Demographics & Household Characteristics**

| Variable | (1) Control Mean/(SE) | (2) Treatment Mean/(SE) | (1)-(2) Pairwise t-test Mean difference |
|---|---|---|---|
| Age in months | 87.36<br>(0.85) | 87.13<br>(0.85) | 0.23 |
| Female | 0.52<br>(0.03) | 0.48<br>(0.02) | 0.04* |
| Emirati | 0.41<br>(0.07) | 0.46<br>(0.07) | -0.05 |
| Native | 0.68<br>(0.06) | 0.66<br>(0.06) | 0.01 |
| Muslim | 0.88<br>(0.03) | 0.87<br>(0.03) | 0.01 |
| Attendance at pre-school or nursery (Yes=1, No=0) | 0.90<br>(0.02) | 0.87<br>(0.03) | 0.03 |
| Has books at home (Yes=1, No=0) | 0.79<br>(0.02) | 0.79<br>(0.02) | -0.00 |
| Has arabic books at home (Yes=1, No=0) | 0.70<br>(0.03) | 0.71<br>(0.03) | -0.00 |
| Last read book (Knows=1, Doesn't know=0) | 0.49<br>(0.03) | 0.50<br>(0.03) | -0.00 |
| Mother present at home (1=Yes, 0=No) | 0.99<br>(0.01) | 0.99<br>(0.01) | 0.00 |
| BL: Father present at home (Yes=1) | 0.58<br>(0.03) | 0.60<br>(0.03) | -0.02 |
| Family has driver (1=Yes, 0=No) | 0.22<br>(0.03) | 0.23<br>(0.03) | -0.00 |
| Mother works outside home (1=Yes, 0=No) | 0.39<br>(0.03) | 0.39<br>(0.03) | 0.00 |
| Has own room (1=Yes, 0=No) | 0.15<br>(0.02) | 0.17<br>(0.02) | -0.02 |
| **N** | 661 | 647 | 1308 |
| **Clusters** | 42 | 41 | 83 |

*Notes:* This table presents balance tests comparing baseline characteristics between the treatment and control groups. Standard errors are reported in parentheses and clustered at the classroom level. All regressions include school fixed effects. Panel A regressions additionally control for EGRA exam model fixed effects, while Panel B regressions include only school fixed effects. **Columns (1) and (2):** Report mean values and standard errors (in parentheses) for the control and treatment groups, respectively. N indicates the number of observations and Clusters indicates the number of classroom clusters. **Pairwise t-test [Column (1)-(2)]:** Reports the difference in means between treatment and control groups (Treatment - Control). Statistical significance is based on t-tests from regressions of each baseline variable on the treatment indicator, controlling for school fixed effects and EGRA exam model fixed effects in Panel A, and school fixed effects only in Panel B. Standard errors are clustered at the classroom level in all specifications. **Panel A - Early Grade Reading Assessment:** All EGRA scores are standardized and penalized for incorrect responses. These measures capture students' reading proficiency across multiple dimensions including overall reading ability, phonemic awareness (initial sound), letter recognition, decoding skills (nonword reading), and reading speed (fluency). Regressions control for school and EGRA exam model fixed effects. **Panel B - Demographics & Household Characteristics:** Includes student demographic information and household socioeconomic indicators. Binary variables are coded as 1 = Yes and 0 = No. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

# 5 Attrition

A total of 874 students were surveyed at baseline. Of these, 65 students were not reached at endline, leaving 809 students successfully surveyed again. This corresponds to a retention rate of 92.6%. At endline, the total number of surveyed students was 1,308, which includes both the reinterviewed baseline students and additional students recruited to increase the overall sample size.

Table 4 presents the differential attrition analysis, where the dependent variable equals one if a student in the analytical baseline sample is missing at endline. Column (1) shows that attrition is low and statistically balanced between treatment and control groups. Column (2) adds baseline covariates and their interactions with treatment to assess potential selective attrition. None of these interaction terms are statistically significant, indicating that there is no evidence of differential attrition across observable characteristics.

Table 4: Differential Attrition Analysis

| | (1) | (2) |
|---|---|---|
| | Treatment Only | With Baseline Characteristics |
| Treatment | -0.001 | -0.096 |
| | (0.015) | (0.209) |
| Female | | 0.012 |
| | | (0.025) |
| Native | | -0.038 |
| | | (0.044) |
| Age in months | | 0.001 |
| | | (0.002) |
| Baseline EGRA overall score | | -0.008 |
| | | (0.014) |
| Treatment × Female | | -0.041 |
| | | (0.034) |
| Treatment × Native | | 0.025 |
| | | (0.040) |
| Treatment × Age in months | | 0.001 |
| | | (0.002) |
| Treatment × Baseline EGRA overall score | | 0.003 |
| | | (0.022) |
| DOB month missing dummy | | 0.105 |
| | | (0.111) |
| Observations | 874 | 872 |
| Control Mean | 0.076 | 0.076 |

*Notes:* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. This table examines differential attrition between treatment and control groups. The dependent variable is a binary indicator equal to 1 if the student attrited (was not present at endline), and 0 otherwise. Column (1) presents a simple regression of attrition on treatment status only. Column (2) includes baseline student characteristics and their interactions with treatment to test whether attrition patterns differ by student type. **Treatment** represents the main effect of treatment assignment on attrition probability. The interaction terms (Treatment × Female, Treatment × Native, Treatment × Age, Treatment × Baseline EGRA) test whether the treatment effect on attrition varies by student characteristics. **Control Mean** shows the attrition rate in the control group. The lack of significant treatment effects indicates no evidence of differential attrition between treatment and control groups, supporting the validity of the impact estimates. Standard errors are clustered at the school level.

Overall, the results indicate that (i) attrition is low in absolute terms; (ii) attrition rates do not differ significantly between treatment and control groups; and (iii) there is no evidence of selective attrition based on observable characteristics. Therefore, the

internal validity of the study's impact estimates is unlikely to be threatened by attrition bias.

# 6 Academic Outcomes

Academic outcomes in this study are organized as primary, secondary, and exploratory outcomes. The primary outcome is Arabic literacy, measured by the overall Early Grade Reading Assessment (EGRA) score, which averages normalized scores from all reading subtasks. Secondary outcomes include performance on individual EGRA subtasks. As an exploratory outcome, we assess numeracy using a timed addition and subtraction test, with scores normalized to the control group. All analysis controls for school fixed effects, to isolate the random assignment of classrooms to Iqra and the control group within a given school.

## 6.1 Arabic Literacy

In this section, we assess Arabic literacy using the Early Grade Reading Assessment (EGRA) scores. For each scoring approach, we present treatment effect regressions on both the overall EGRA score and the individual subtasks.

The five EGRA subtasks included in the endline survey are:

1. Letter sound identification
2. Nonword reading
3. Familiar word reading
4. Oral reading fluency
5. Reading comprehension

All regressions control for baseline scores of the same subtask, where available (all except reading comprehension). Additionally, all regressions control for letter sound identification score, given a slight imbalance at baseline. These controls allow us to interpret the treatment coefficient as the effect of Iqra after adjusting for a student's initial literacy level.

We find that random assignment of a classroom to Iqra increased overall EGRA scores by 0.09 standard deviations (sd), on average, relative to the control group (Table 5, column 1). This effect is precisely estimated, with statistical significance at the 1% level. This effect implies that the median student in a control classroom would move from the 50th to the 54th percentile of Arabic literacy if exposed to the program.

Further estimates suggest that Iqra leads to reading improvements equivalent to an additional 25% of a school year, relative to the progress made by students in the control group under the business-as-usual approach. The program's impact also compares well with other literacy and education interventions evaluated rigorously in a diverse set of global contexts (Evans and Yuan 2022).

The remaining columns of Table 5 report the effect of Iqra on specific EGRA subtasks. The program increased students' ability to read words, non-words (nonsense words that test the ability to decode sounds), identify letters, and read more fluently. The program was particularly effective at improving word reading, with an effect of 0.19 standard deviations (equivalent to moving the median control student from the 50th to the 58th percentile). The only task on which Iqra students did not improve significantly was reading comprehension, which was not explicitly taught by the program and tends to develop at later stages of literacy, once more fundamental competencies have been acquired.

Table 5: Impact on Student Academic Outcomes

| | (1) EGRA Overall Score | (2) Letter Identification Score | (3) Word Reading Score | (4) Non-word Reading Score | (5) Reading Fluency Score | (6) Reading Comprehension Score |
|---|---|---|---|---|---|---|
| Treatment | 0.0916*** | 0.102*** | 0.194*** | 0.0890** | 0.0960** | -0.0142 |
| | (0.0314) | (0.0379) | (0.0409) | (0.0438) | (0.0388) | (0.0187) |
| Observations | 1308 | 1308 | 1308 | 1308 | 1308 | 1308 |
| Control Mean | 0.000 | 0.000 | 0.000 | 0.000 | -0.000 | 0.306 |

*Notes:* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Arabic literacy outcomes are based on penalised EGRA endline scores. Incorrect responses were deducted when calculating the raw scores for all EGRA subtasks. **(1) EGRA Overall Score** is the average across all subtask scores, each normalized to the control group mean and standard deviation. **(2) Letter Identification** measures alphabetic knowledge as the normalized number of correct letter sounds identified per minute (CSPM) at endline. **(3) Word Reading Score** measures familiar word reading as the normalized number of correct words identified per minute (CWPM), capturing decoding skills. **(4) Non-Word Reading** captures decoding skills as the normalized number of correct non-words identified per minute (CNWPM) at endline. **(5) Reading Fluency** measures oral reading fluency as the normalized number of correct words read per minute (CWPM) from a short passage at endline. **(6) Reading Comprehension** measures comprehension through the number of correct responses to questions about an oral passage within 90 seconds. All scores are normalized to the control group mean and standard deviation. *Controls:* Regressions (1–4) include the following baseline controls: **Baseline EGRA Overall Score** — average across all baseline subtask scores (initial sound identification, letter sound identification, non-word reading, and oral reading fluency), normalized to the control group mean and standard deviation; **Baseline Letter Identification Score** — normalized number of correct letter sounds identified per minute (CSPM) at baseline; **Baseline Participation** — dummy variable equal to 1 if the student was included in the baseline sample; **Native** — dummy variable equal to 1 if the student is a native Arabic speaker; **Female** — dummy variable equal to 1 if the student is female; **Age in Months** — student's age in months; **Date-of-Birth Month Missing Dummy** — indicator for missing Date of birth month date; **Father Present at Baseline** and **Father Present at Endline** — dummy variables equal to 1 if the father was present in the household at baseline or endline, respectively. Regression (4) additionally controls for the **Baseline Non-Word Reading Score** (normalized number of correct non-words identified per minute at baseline). Regression (5) additionally controls for the **Baseline Reading Fluency Score** (normalized number of correct words read per minute at baseline).

Given the focus on Arabic diglossia, did native Arabic speakers respond differently to the program than non-native Arabic speakers? To test this possibility, we add an interaction term between assignment to Iqra and an indicator for whether the student is a native Arabic speaker. Table 7 reports results. Not surprisingly, the main effect of na-

tive is positive, reflecting greater Arabic literacy among native speakers. The interaction term is not statistically significant for overall literacy or for most subtasks, indicating no differential effect of Iqra between native and non-native Arabic speakers. However, the interaction term is negative and significant (at 10%) for letter identification, meaning native Arabic speakers gained slightly less from the program than non-native speakers on this task. One potential explanation is that non-native speakers, who were less familiar with Arabic letters at the outset, experienced particularly large benefits from the repetition of letters in Iqra lessons.

The relatively strong performance of the program for non-native speakers is a welcome development. Nonetheless, it is somewhat surprising, given the program's focus on improving literacy in MSA in the presence of Arabic diglossia. They suggest that Iqra may be universally beneficial for teaching early-grade Arabic literacy, even where diglossia is not present. These results present further motivation to understand the underlying mechanisms.

Table 7: Heterogeneity Analysis: Native vs Non-Native

| | (1) EGRA Overall Score | (2) Letter Identification Score | (3) Word Reading Score | (4) Non-word Reading Score | (5) Reading Fluency Score | (6) Reading Comprehension Score |
|---|---|---|---|---|---|---|
| Treatment | 0.128** | 0.186*** | 0.189** | 0.0931 | 0.177** | 0.0118 |
| | (0.0584) | (0.0531) | (0.0823) | (0.0849) | (0.0869) | (0.0292) |
| Native | 0.303*** | 0.337*** | 0.242** | 0.226** | 0.453*** | 0.113** |
| | (0.0829) | (0.0964) | (0.101) | (0.105) | (0.104) | (0.0551) |
| Treatment × Native | -0.0540 | -0.125* | 0.00635 | -0.00617 | -0.121 | -0.0389 |
| | (0.0726) | (0.0746) | (0.0987) | (0.0986) | (0.101) | (0.0393) |
| Observations | 1308 | 1308 | 1308 | 1308 | 1308 | 1308 |
| Control Mean | 0.000 | 0.000 | 0.000 | 0.000 | -0.000 | 0.306 |

*Notes:* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. This table presents heterogeneity analysis examining differential treatment effects by native language status. **Treatment** represents the main effect of the intervention for non-native Arabic speakers (the reference group). **Native** is a dummy variable equal to 1 if the student is a native Arabic speaker, capturing baseline differences between native and non-native speakers in the control group. **Treatment × Native** is the interaction term testing whether treatment effects differ for native Arabic speakers compared to non-native speakers. All outcome measures are normalized to the control group mean and standard deviation. All regressions include school fixed effects and the same set of baseline controls as Table 5: baseline EGRA overall score, baseline letter identification score, baseline participation dummy, native speaker status, gender, age in months, date-of-birth month missing dummy, and father presence at baseline and endline. Regressions (4) and (5) additionally control for baseline non-word reading score and baseline reading fluency score, respectively.

Of the 26 schools included in the study, 13 are Emirati schools following the MOE curriculum, 7 cater to South Asian students following Indian or Pakistani curricula, and 6 are international schools following International Baccalaureate (IB), British or American curricula. These school types differ along multiple dimensions including curriculum, student socio-economic background and school resources. In particular, South Asian schools tend to have students from more disadvantaged backgrounds, have fewer resources, and are less well-equipped than Emirati schools.

The Iqra program has a positive and significant effect on the Word Reading score for Emirati schools (0.215 sd, p¡0.01), suggesting substantial improvement in students' word reading skills. For the other outcomes, the effects for Emirati schools are mostly non-significant. The interaction terms indicate that, for students in South Asian and International schools, there are generally no significant differences in the program's effect compared to Emirati schools. The South Asian $\times$ Treatment interaction for letter identification shows a marginally significant positive coefficient (0.141, p¡0.10), and the interaction for reading fluency shows a positive coefficient (0.207, p¡0.10), suggesting potentially larger gains for South Asian school students on these dimensions. Overall, the results suggest that the program's impact is largely consistent across school types. This is notable from an educational equity perspective, given that students in South Asian schools in the sample face relatively greater socio-economic disadvantages and attend schools with fewer resources.

While these findings provide initial evidence that Iqra reaches students across diverse educational settings, they capture only one dimension of potential heterogeneity. Further analysis is needed to assess treatment effect variation by other school-level characteristics, such as public versus private status, as well as by additional student characteristics beyond native language. Such analyses would provide deeper insight into the mechanisms driving program effectiveness and inform more targeted implementation and scaling strategies.

Table 8: Heterogeneity Analysis: Emirati vs South-Asian vs International Schools

| | (1) EGRA Overall Score | (2) Letter Identification Score | (3) Word Reading Score | (4) Non-word Reading Score | (5) Reading Fluency Score | (6) Reading Comprehension Score |
|---|---|---|---|---|---|---|
| Treatment | 0.0708 | 0.0754 | 0.215*** | 0.0817 | 0.0203 | -0.0295 |
| | (0.0568) | (0.0717) | (0.0652) | (0.0776) | (0.0611) | (0.0307) |
| Asian × Treatment | 0.0880 | 0.141* | 0.00999 | 0.0370 | 0.207* | 0.0298 |
| | (0.0782) | (0.0825) | (0.109) | (0.117) | (0.106) | (0.0389) |
| International × Treatment | -0.00780 | -0.0339 | -0.0791 | -0.00805 | 0.0715 | 0.0251 |
| | (0.0664) | (0.0839) | (0.0873) | (0.0889) | (0.0728) | (0.0483) |
| Observations | 1308 | 1308 | 1308 | 1308 | 1308 | 1308 |
| Control Mean | 0.000 | 0.000 | 0.000 | 0.000 | -0.000 | 0.306 |

*Notes:* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. This table presents a heterogeneity analysis examining differential treatment effects by curriculum. Emirati schools follow the Ministry of Education (MOE) curriculum and are used as the reference category. South Asian schools follow the Indian or Pakistani curriculum. International schools follow the International Baccalaureate, American, or British curriculum. **Treatment** represents the main effect of the intervention for students in Emirati schools (the reference group). **South Asian × Treatment** and **International × Treatment** are interaction terms testing whether the treatment effects differ for South Asian schools and International schools, respectively, compared to Emirati schools.

## 6.2 Exploratory Outcomes

### 6.2.1 Item-level analysis: Effect of font size and spacing

One mechanism through which Iqra intends to improve literacy is by leveraging perceptual learning, the ability of the visual system to combine individual shapes into larger configurations and detect patterns with practice (Gori 2014). To stimulate this perceptual learning, Iqra materials use larger fonts and greater spacing between letters than the default Ministry of Education curriculum. We test this mechanism by randomly assigning larger fonts and greater spacing among items in the word and non-word reading EGRA subtasks.[3] This structure allows us to test whether font size and spacing matter for literacy at the item level. We also interact dummies for large font and spaced letters with treatment, to test whether treatment influences the role of these characteristics on item-level performance.[4]

[3]Word and non-words are presented in grids of 12 items each. We vary the font size and spacing of all items in a grid. Then, we randomly order the grids presented to students. Because students are more likely to answer items on grids appearing earlier in the sequence, we control for grid order in all analyses.

[4]The estimating equation is:

$$Correct_{ijgcs} = \alpha_s + \rho_1 Large_j + \rho_2 Spaced_j + \rho_3 Iqra_c + \rho_4 (Large_j \times Iqra_c) + \rho_5 (Spaced_j \times Iqra_c) + \gamma_g + X_i\theta + \varepsilon_{ijgcs} \quad (3)$$

where $i$, $j$, $g$, $c$, and $s$ index students, test items, test grid order (1-8), classrooms, and schools, respectively; $Correct$ is an indicator for answering the item correctly; $Large$ is an indicator for the item appearing in large font; $Spaced$ is an indicator for the item appearing with wide spacing; $\alpha$ and $\gamma$ are school and grid fixed effects, repsectively; and $X$ includes student-level controls from equation 1. The coefficients $\rho_1$ and $\rho_2$ measure the effects of font size and spacing, respectively, on item-level performance in the control group. The coefficient $\rho_3$ measures the effect of assignment to Iqra on performance on

Table 4 presents results. We find no statistically significant effects of large font and spaced letters on non-word reading, either in isolation or interacted with treatment. For word reading, we also find no statistically significant effects of spaced letters. However, for word reading we find three statistically significant results. First, Iqra students are 6.6 percentage points more likely to correctly read words presented in standard font size and spacing. This result is consistent with the positive treatment effect of Iqra on word reading in Table 5. Second, large font increases the probability of reading a word correctly by 2.4 percentage points, consistent with perceptual learning. Third, Iqra *reduces* the effect of large font on word reading by a marginally significant 3.1 percentage points, sufficient to undo the effect of large font on word reading.

Put differently, large font improved word reading, but only for students in control classrooms. Moreover, Iqra students performed better than control students when reading words with standard font size and spacing. One interpretation is that by the endline, Iqra students had improved sufficiently in reading familiar words that large fonts provided no additional advantage. While counterintuitive at first sight, this result is highly reassuring: While large fonts and spaced letters are core techniques used within the Iqra curriculum, students do not become dependent on them. On the contrary, our results are consistent with students obtaining improved familiarity and automaticity thanks to this temporary scaffolding, with the result that they no longer need it after a certain time period. In contrast, control students – who had been less exposed to these reading supports - continued to benefit from them at endline.

items with standard font size and spacing. The coefficients $\rho_4$ and $\rho_5$ measure differential effects of Iqra on item-level performance on words presented in large and spaced fonts, respectively, compared to the control group.

Table 10: Impact of Text Formatting

| | (1) Non-words (Cat) | (2) Words (Cat) |
|---|---|---|
| Large | -0.00494 | 0.0235** |
| | (0.00842) | (0.0104) |
| Treatment | 0.0180 | 0.0661*** |
| | (0.0174) | (0.0194) |
| Treatment × Large | -0.000992 | -0.0317* |
| | (0.0139) | (0.0167) |
| Spaced | 0.00929 | -0.00186 |
| | (0.0128) | (0.0147) |
| Treatment × Spaced | 0.0218 | 0.0276 |
| | (0.0190) | (0.0207) |
| Observations | 14861 | 14175 |

*Notes:* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. This table presents item-level analysis of the effects of text formatting on reading performance. The dependent variable is a binary indicator equal to 1 if the student answered the item correctly, and 0 otherwise. Column (1) presents results for non-word reading items, and column (2) presents results for word reading items. **Large** is a dummy variable equal to 1 if the item was presented in large font (24-point vs. 18-point). **Spaced** is a dummy variable equal to 1 if the item was presented with wide letter spacing. **Treatment** represents the effect of Iqra assignment on correctly reading items with standard font size and spacing. **Treatment × Large** and **Treatment × Spaced** are interaction terms testing whether treatment effects differ for items presented with large font or wide spacing, respectively. All regressions include school fixed effects, grid order fixed effects (to control for item presentation sequence), and the same baseline student-level controls as Table 5: baseline EGRA overall score, baseline letter identification score, baseline participation dummy, native speaker status, gender, age in months, date-of-birth month missing dummy, and father presence at baseline and endline. Standard errors are clustered at the classroom level.

### 6.2.2 Numeracy and Executive Function Outcomes

This section examines whether Iqra influenced students' numeracy skills, executive function, or home literacy environment. Although numeracy was not directly targeted by the program, improvements in literacy could theoretically spill over into mathematics performance if enhanced decoding supports the understanding of numerical symbols. Conversely, the intervention could crowd out mathematics learning by reallocating instructional time. We measure numeracy using a timed one-minute set of single- and double-digit addition and subtraction problems, normalized using the control-group mean and standard deviation.[5]

Executive function outcomes capture students' ability to maintain focus and cognitive endurance. We assess attentional capacity using performance on the first iteration of a shape-identification task, and sustained attention as the change in performance between the first and second iterations, conducted at the start and end of the assessment period. These measures aim to detect whether Iqra improved students' ability to concentrate over

[5]The task aligns closely with the Early Grade Mathematics Assessment (EGMA) Addition/Subtraction Level 1 subtask.

time.

Finally, we also report impacts on the home literacy environment, constructed from parent- and student-reported indicators on reading frequency, availability of books, and literacy-related interactions at home.

Across all four domains—numeracy, attentional capacity, sustained attention, and home literacy environment—we find no statistically significant effects of Iqra. Point estimates are small in magnitude and not consistently signed, providing no evidence that the program produced spillover effects on mathematics, cognitive endurance, or home learning conditions.

Table 11: Impact on Secondary Mechanisms

| | (1) Math score | (2) Attentional Capacity | (3) Sustained Attention | (4) EL: Home literacy environment |
|---|---|---|---|---|
| Treatment | 0.0421 | 0.0404 | -0.0193 | 0.0139 |
| | (0.0476) | (0.0439) | (0.0468) | (0.0102) |
| Observations | 1308 | 1308 | 1308 | 1277 |
| Control Mean | -0.000 | -0.000 | -0.000 | 0.779 |

*Notes:* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

**(1) Math Score** Refers to the endline mathematics score based on single- and double-digit addition and subtraction problems answered within one minute, normalized to the control group mean and standard deviation. The measure was computed as the number of correct responses minus incorrect ones, with "don't know" responses also deducted. **(2) Attentional Capacity** captures the normalized number of target shapes correctly identified within one minute. **(3) Sustained Attention** represents the normalized change in attentional capacity between the first and second exercise (end – start). **(4) Home Literacy Environment** is an index ranging from [0,1], measured as the mean of (i) an indicator for whether the student has Arabic books at home, and (ii) the reported frequency with which someone reads with the child (0 = never, 0.25 = rarely, 0.5 = sometimes, 0.75 = most days, 1 = every day). "Don't know" responses were also deducted. Measures were taken at endline. All outcome variables are normalized to the control group mean and standard deviation. All regressions include school fixed effects. Regressions (1–4) include have the same baseline student-level controls as Table 5: baseline EGRA overall score, baseline letter identification score, baseline participation dummy, native speaker status, gender, age in months, date-of-birth month missing dummy, and father presence at baseline and endline. Regression (4) additionally controls for the **Baseline Home Literacy Environment**, constructed using baseline data following the same definition described above. Standard errors are clustered at the classroom level.

# 7 Program Implementation and Perception Outcomes

How well was the program implemented? How did teachers perceive the program? We use two data sources to measure implementation and perceptions: classroom observations and a teacher survey. Observations were conducted by independent enumerators at endline. Teacher surveys were administered by AQF during the second half of the academic year. Each teacher provided information for every class they taught as part of the study, so for both classroom and teacher surveys, the unit of observation is the classroom.[6]

Each data source has strengths and weaknesses. Classroom observations provide direct measures of implementation, but are limited to one session, which may not be representative of classroom practice throughout the academic year. Moreover, teacher behavior may be influenced by the presence of an observer. Teacher surveys reflect practice during the academic year, but these self-reported measures may reflect social desirability bias (i.e., teachers say what they think the surveyor wants to hear). This bias may be exacerbated because surveys were conducted by AQF, the implementing partner. Using both data sources for analysis allows us to check for consistency in results.

## 7.1 Iqra pedagogy index

The primary implementation outcome is the Iqra pedagogy index, which captures teacher adherence to core instructional practices emphasized in the Iqra curriculum. The index aggregates a set of binary indicators reflecting practices such as reviewing prior content, introducing one letter or vowel per lesson, reading aloud with students, engaging students in individual reading, presenting letter combinations using the target letter, and providing timely feedback. Full definitions of all components are provided in Annex Section 10.2.1.

Most of these practices were also measured in the teacher survey, allowing us to construct both reported and observed measures of Iqra pedagogy. Table **??** presents treatment effects on the composite index and all major components.

The observed index shows a 50 percentage-point increase in Iqra-aligned practices, rising from a control mean of 31%. This represents a substantial improvement in fidelity to Iqra instructional routines. Individual practices also show large, statistically significant gains: reading aloud increases by 64 percentage points, individual reading by 79 percentage points, letter-combination presentation by 59 percentage points, index-finger tracking by 54 percentage points, and feedback provision by 26 percentage points. These results demonstrate that teachers implemented Iqra's pedagogy extensively and consistently during classroom instruction. The only component without a statistically significant improvement is lesson review.

[6]We initially planned to observe 83 classrooms, but due to challenges in endline data collection, 6 classroom observations are missing. The following schools have missing classroom surveys: Al Rams School (1 missing class), Alpha Private School (2), Ethen School (2), Harvest Private School (1).

To complement the observed data, we also construct a reported Iqra pedagogy index based on teacher survey responses. Teachers reported practices for each class they taught as part of the study. The index includes binary indicators for introducing letters in the sequence specified by the Iqra textbook and holding information sessions with parents, as well as frequency-based indicators for reviewing previous lessons, teaching one new letter per class, reading aloud, presenting letter combinations, instructing individual reading, focusing only on known materials, providing immediate feedback, conducting formal assessments, grouping students by ability, sending lesson copies to parents of weaker students, and implementing paired reading activities.

The reported index shows a 7 percentage-point increase from a high control mean of 89%. While smaller in magnitude than the observed effects—likely due to upper-bound ceiling effects—the pattern across individual components mirrors the observational results. Significant gains are seen in individual reading (+19 percentage points), letter-combination work (+12 percentage points), index-finger tracking (+8 percentage points), feedback (+7 percentage points), and use of known materials (+47 percentage points). We find declines in several components (such as formal assessment or certain parent-related practices), concentrated among practices not expected to occur at daily frequency within classrooms.[7] Accordingly, when separately measuring the effect of Iqra on a broad index of practices and a subindex restricted to daily activities, we find a significant increase of 9 percentage points in the daily activities index, but no effect on the broad index.

Together, results from both reported and observed measures show strong adoption of Iqra pedagogy, consistent with high program fidelity. Teachers systematically shifted their instructional routines to align with Iqra's structured approach to early literacy development. However, practices expected outside the daily routine, such as communication with parents and formal assessments, were not implemented as strongly. This suggests implementers should emphasize these non-routine practices in future iterations of the program.

[7]For this reason, these measures were not included in classroom observations, as we did not expect to observe them during a single visit.

Table 12: Treatment Effects on Iqra Pedagogy: Reported vs Observed Practices

| | Reported | | | | Observed | | | |
|---|---|---|---|---|---|---|---|---|
| Practice | Coef | SE | Control Mean | N | Coef | SE | Control Mean | N |
| Iqra Pedagogy Index | 0.07*** | (0.02) | 0.89 | 138 | 0.50*** | (0.04) | 0.31 | 77 |
| Review Previous | -0.07** | (0.03) | 0.97 | 138 | 0.10 | (0.12) | 0.32 | 77 |
| One Letter/Lesson | 0.11*** | (0.04) | 0.87 | 138 | 0.59*** | (0.08) | 0.24 | 77 |
| Read Aloud | 0.01 | (0.02) | 0.96 | 138 | 0.64*** | (0.07) | 0.29 | 77 |
| Individual Reading | 0.19*** | (0.04) | 0.76 | 138 | 0.79*** | (0.06) | 0.16 | 77 |
| Letter Combinations | 0.12*** | (0.04) | 0.86 | 131 | 0.59*** | (0.09) | 0.24 | 77 |
| Index Finger | 0.08*** | (0.03) | 0.92 | 138 | 0.54*** | (0.09) | 0.21 | 77 |
| Feedback | 0.07*** | (0.02) | 0.90 | 138 | 0.26*** | (0.07) | 0.71 | 77 |
| Known Material | 0.47*** | (0.06) | 0.37 | 136 | | | | |
| Formal Assessment | -0.28*** | (0.06) | 0.80 | 138 | | | | |
| Group Abilities | -0.11** | (0.04) | 0.79 | 138 | | | | |
| Send Parents | -0.41*** | (0.07) | 0.84 | 70 | | | | |
| Custom Activities | -0.40*** | (0.07) | 0.76 | 68 | | | | |
| Pair Practice | 0.07* | (0.04) | 0.81 | 138 | | | | |
| Follow-up Parents | -0.37*** | (0.07) | 0.79 | 110 | | | | |
| Broad Reported Index | -0.00 | (0.02) | 0.82 | 138 | | | | |
| Daily Activities Index | 0.09*** | (0.01) | 0.82 | 138 | | | | |

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Each coefficient represents the treatment effect of the Iqra program on Iqra-aligned pedagogical practices. Reported practices come from the teacher survey; observed practices come from classroom observations. The Iqra Pedagogy Index is constructed from an overlapping set of components across both data sources. The Broad Reported Index and Broad Observed Index include all Iqra-related pedagogical components measured within each respective data source. Standard errors are shown in parentheses.

## 7.2 Iqra materials index

As a secondary implementation outcome, the Iqra materials index measures the extent to which classrooms use instructional materials aligned with the Iqra curriculum. The index includes multiple components capturing the availability of Iqra student books, the presence of Iqra teacher materials, and the effective use of Iqra-aligned visual aids. Examples of these practices include having Iqra textbooks distributed to students, using the Iqra teacher guide during instruction, and displaying visual materials with large, clear text to support reading tasks. Full definitions of all components are provided in Annex Section 10.2.3.

Most of these practices were also captured in the teacher survey, allowing us to construct a reported Iqra materials index that parallels the observed index. Reported measures include whether the Iqra textbook is used as the primary Arabic textbook and the frequency with which teachers use visual aids or materials with large-font text.

Table 13 presents treatment effects on both the composite Iqra materials index and its individual components for reported (teacher survey) and observed (classroom observation) data. The reported index shows a 53 percentage-point increase in the use of Iqra materials, while the observed index shows an even larger impact of 82 percentage points.

These increases reflect substantial adoption of Iqra materials in treatment classrooms.

The individual components also demonstrate large and statistically significant improvements. In the observed data, the proportion of students with Iqra textbooks rises by 80 percentage points from a control mean of only 3%, and the likelihood that teachers have Iqra curriculum materials increases by 99 percentage points from a near-zero baseline. Effective use of visual aids increases by 64 percentage points, reflecting widespread take-up of Iqra instructional practices. Results from the teacher survey show similarly strong adoption patterns, with near-universal reporting of Iqra teacher materials in treatment classrooms. Reported use of visual aids does not increase in response to Iqra, but the high control mean (80%) suggest teachers did not distinguish between types of visual aids used in Iqra and MoE classrooms.

Together, these findings show strong implementation fidelity with respect to Iqra materials: classrooms not only substituted away from MoE textbooks and materials but also adopted the full suite of Iqra instructional tools.

Table 13: Treatment Effects on Iqra Materials: Reported vs Observed Practices

| | Reported | | | | Observed | | | |
|---|---|---|---|---|---|---|---|---|
| Practice | Coef | SE | Control Mean | N | Coef | SE | Control Mean | N |
| Iqra Materials Index | 0.53*** | (0.03) | 0.40 | 138 | 0.82*** | (0.04) | 0.16 | 77 |
| Teacher Has Materials | 1.00 | | 0.00 | 138 | 0.99*** | (0.02) | 0.00 | 77 |
| Effective Visual Aids | 0.06 | (0.07) | 0.80 | 138 | 0.64*** | (0.07) | 0.32 | 77 |
| Students Own Book | | | | | 0.80*** | (0.07) | 0.03 | 77 |
| Iqra Broad Index | | | | | 0.81*** | (0.04) | 0.11 | 77 |

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Each coefficient represents the treatment effect of the Iqra program on Iqra-aligned material use. Reported practices come from the teacher survey; observed practices come from classroom observations. The Iqra Materials Index is constructed using an overlapping set of components across both data sources. The Iqra Broad Index includes all observed Iqra material-related components, including those not captured in the teacher survey. Standard errors are shown in parentheses.

## 7.3 Ministry of Education (MoE) pedagogy index

As another secondary implementation outcome, the Ministry of Education (MoE) pedagogy index assesses teacher practices under the MoE curriculum using both classroom observations and the teacher survey. The index captures a range of practices that typically characterize MoE instruction, including use of MoE materials, lesson-structuring techniques, and enrichment activities. Examples include having MoE textbooks available for students, using MoE curriculum guides or lesson plans, writing lesson objectives, and incorporating enrichment activities such as songs, pictures, stories, multimedia, or Arabic writing tasks. Full definitions of all components are provided in Annex Section 10.2.2.

Most of these practices were also measured in the teacher survey, which enables direct comparison between what teachers report and what enumerators observed in the

classroom. Table 14 summarizes treatment effects on both reported and observed MoE practices. The left panel presents self-reported MoE practices from teachers, while the right panel presents the same practices from classroom observations.

The results show consistent and substantial reductions in MoE-style practices in Iqra classes. The MoE Pedagogy Index decreases by 38 percentage points in reported data and 34 percentage points in observed data. Large and statistically significant reductions are also visible across individual components: 39 percentage points in the likelihood of students having their own book; writing instruction by 52 percentage points; the use of pictures by 54 percentage points; and the use of stories by 34 percentage points. These declines indicate strong program fidelity—teachers substantially reduced MoE instructional routines and adopted the distinct pedagogical practices of Iqra.

Table 14: Treatment Effects on MoE Pedagogy: Reported vs Observed Practices

| | Reported | | | | Observed | | | |
|---|---|---|---|---|---|---|---|---|
| Practice | Coef | SE | Control Mean | N | Coef | SE | Control Mean | N |
| MoE Pedagogy Index | -0.38*** | (0.03) | 0.90 | 138 | -0.34*** | (0.03) | 0.69 | 77 |
| Write Objectives | -0.31*** | (0.05) | 0.96 | 138 | -0.03 | (0.06) | 0.87 | 77 |
| Songs | -0.19*** | (0.05) | 0.78 | 138 | -0.11 | (0.11) | 0.42 | 77 |
| Pictures | -0.34*** | (0.04) | 0.96 | 138 | -0.54*** | (0.09) | 0.76 | 77 |
| Stories | -0.48*** | (0.05) | 0.84 | 138 | -0.34*** | (0.09) | 0.61 | 77 |
| Multimedia | -0.28*** | (0.05) | 0.88 | 138 | 0.06 | (0.04) | 0.95 | 77 |
| Teach Writing | -0.81*** | (0.05) | 0.99 | 110 | -0.52*** | (0.09) | 0.68 | 77 |
| Students Own Book | | | | | -0.39*** | (0.08) | 0.42 | 77 |
| Teacher Materials | | | | | 1.16*** | (0.23) | 2.74 | 77 |
| MoE Broad Index | | | | | -0.83*** | (0.06) | 0.84 | 77 |

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Each coefficient represents the treatment effect of the Iqra program on MoE-aligned pedagogical practices. Reported practices come from the teacher survey; observed practices come from classroom observations. The MoE Pedagogy Index is constructed using an overlapping set of components across both data sources. The MoE Broad Index includes all observed MoE pedagogy-related components, including those not captured in the teacher survey. Standard errors are shown in parentheses.

## 7.4 Iqra Perceptions Index

We construct an Iqra perceptions index from teacher survey responses to capture how important teachers consider various instructional practices for learning Arabic. The index reflects the extent to which teachers value key Iqra-aligned practices, such as reviewing previous lessons, using visual aids and large fonts, encouraging repeated reading, providing feedback, grouping students by ability, and engaging parents. It also reflects how teachers view practices less aligned with Iqra, such as incorporating songs, pictures, or multimedia. This index and its components differ from both the classroom observation and teacher survey measures reported above, which measure teacher *practices*. The perceptions index measures teacher *beliefs* about the importance of these practices to teach Arabic, regardless of whether they use these practices in the classroom.

We asked teachers about their overall perceptions for teaching Arabic, rather than specific to any classroom. We therefore have one observation of perceptions for each teacher in the dataset. We then estimate how the perceived importance of each practice varies with whether the teacher was assigned to teach any Iqra classes, with results in Table 15. For Iqra-aligned practices, higher values of the outcome mean the teacher considers the practice more important. For MoE-aligned practices, higher values of the outcome mean the teacher considers the practice less important.[8] Section **??** lists all Iqra- and MoE-aligned practices. With this coding, positive coefficients on assignment to Iqra Table 15 are consistent with Iqra teachers updating their beliefs to match Iqra's pedaogogical philosophy.

We find a 4 percentage point increase in the Iqra perceptions index in response to teaching Iqra (marginally significant at 10%), relative to a control mean of 69%. Most individual components of the index show no response to Iqra. Among components with significant effect, all but one (formal assessment of students) are in the same direction as recommended practices of Iqra. The positive coefficient on use of pictures means that Iqra teachers believe this practice is less important than non-Iqra teachers, consistent with Iqra pedagogy.

We conclude that Iqra had a modest effect on teacher perceptions about the importance of Iqra-aligned practices for teaching Arabic. This modest effect may have dampened the effectiveness of Iqra, as its teachers may not have been fully invested in the program. Implementers may wish to emphasize the pedagogical philosophy underlying Iqra in future iterations of the program.

[8]Specifically, Iqra-aligned practices are coded as: not important = 0; somewhat important = 0.33; important = 0.67; and very important = 1. MoE-aligned practices are coded as: not important = 1; somewhat important = 0.67; important = 0.33; and very important = 0.

Table 15: Effect of Teaching at least one Iqra class on Iqra Perceptions Index

| | Coefficient | SE | N | Control Mean |
|---|---|---|---|---|
| Overall Perceptions Index | 0.04* | 0.02 | 57 | 0.69 |
| ***Perceived Importance of*** | | | | |
| Reviewing previous lesson | 0.01 | 0.06 | 57 | 0.93 |
| Teaching one letter per lesson | -0.02 | 0.11 | 57 | 0.78 |
| Reading aloud with students | 0.06 | 0.06 | 57 | 0.91 |
| Presenting letter combinations | -0.01 | 0.09 | 54 | 0.86 |
| Instructing individual reading | 0.17** | 0.08 | 57 | 0.72 |
| Using only known material | 0.13 | 0.13 | 56 | 0.47 |
| Using index finger to follow text | 0.06 | 0.07 | 57 | 0.87 |
| Giving immediate feedback | -0.04 | 0.08 | 57 | 0.87 |
| Formally assessing students | -0.19** | 0.09 | 57 | 0.95 |
| Grouping by abilities | 0.07 | 0.08 | 57 | 0.78 |
| Sending activities to parents | 0.18** | 0.07 | 57 | 0.72 |
| Pairing students for practice | -0.10 | 0.07 | 57 | 0.94 |
| Using visual aids | 0.09 | 0.07 | 57 | 0.87 |
| Displaying large font | 0.14 | 0.11 | 57 | 0.20 |
| Writing lesson objectives | 0.03 | 0.10 | 57 | 0.31 |
| Using songs | 0.07 | 0.08 | 57 | 0.13 |
| Using pictures | 0.25*** | 0.07 | 57 | 0.11 |
| Using stories | 0.10 | 0.09 | 57 | 0.17 |
| Using multimedia | 0.10 | 0.15 | 55 | 0.53 |
| Ordering letters logically | 0.00 | 0.02 | 55 | 1.00 |
| Pronouncing correctly (teacher) | -0.08 | 0.08 | 55 | 0.90 |
| Pronouncing correctly (students) | -0.10 | 0.08 | 55 | 0.90 |
| Walking around classroom | 0.05 | 0.05 | 55 | 0.92 |
| Helping weaker students | -0.01 | 0.07 | 55 | 0.90 |

*Notes:* This table presents coefficients from regressions of teaching at least one Iqra class on teachers' perceptions of the importance of various pedagogical practices for teaching Arabic literacy. Standard errors are reported in a separate column. All regressions include school fixed effects. *** p<0.01, ** p<0.05, * p<0.10. The overall index is constructed as the mean of all component perception variables.

# 8 Policy implications and recommendations

The Iqra program improved student literacy in Arabic across multiple domains, including letter identification, word reading, and reading fluency. Effects are largest on word reading. Room for further improvement remains in other reading skills, particularly in "downstream outcomes" such as reading comprehension, which typically develops at more advanced literacy stages.

The positive results of the Iqra pilot echo the World Bank's emphasis on bridging the gap between colloquial Arabic dialects and Modern Standard Arabic (Boni et al 2021). Iqra's approach to teaching MSA may offer a replicable model for addressing this fundamental challenge across Arabic-speaking countries, subject to replication and testing at scale within the UAE first. It is recommended that policymakers invest in interventions that explicitly address diglossia, rather than assuming students will naturally acquire MSA literacy skills.

The results show a significant improvement in foundational skills (letter identification, word reading), but not in reading comprehension, which was not explicitly part of the program. This is an indicator that governments must first ensure students have strong foundational literacy skills before focusing on higher-order comprehension skills.

The program had similar effects for native and non-native Arabic speakers, suggesting it is appropriate for a wide range of students. Nonetheless, the lack of stronger effects among native speakers is puzzling, given the program's focus on improving literacy in MSA in the presence of Arabic diglossia. This pattern merits further analysis and research.

The program is likely to be cost-effective, as it utilizes existing staff and resources, requiring only a few days of additional teacher training. Out of the 26 schools in our study, at least 21 implemented the programme fully within regular curriculum hours.

Implementation quality was high, as reflected in strong adherence to the study protocol and fidelity to Iqra's teaching methods. There is room to improve implementation quality further, particularly for practices (such as communication with parents and formal assessment of students) not expected to occur daily. Additionally, there is room to strengthen teacher beliefs in the importance of Iqra's pedagogical practices, to ensure strong teacher commitment to the program.

The study's demanding design, which involved randomizing classrooms within the same school to receive Iqra or not, makes this high-quality implementation particularly impressive. A school-wide adoption of Iqra is likely easier to implement when scaling up the program, and could simplify coordination across early-grade Arabic classrooms.

Sustaining high implementation quality at scale may pose challenges. While direct program costs are likely to be manageable (a few days of teacher training), scaling will require intensive monitoring support to ensure consistent implementation quality until

Iqra has been established as the standard approach.

In this evaluation, AQF played a crucial role in supporting teachers and ensuring the fidelity of implementation, which required the time and effort of skilled facilitators. This need for monitoring, which is common for programs introducing novel pedagogical approaches, will be essential to maintain consistent quality as Iqra expands across RAK and the wider UAE.

# 9 References

# References


Abadzi, H., "Improving Arab students' academic achievement: The crucial role of rapid reading and grammar mastery in the early grades," *Policy Paper*, 2017, (20).

Abadzi, Helen and Marialuisa Martelli, "Efficient Reading for Arab Students: Implications from Neurocognitive Research," *World Summit of Innovation in Education (WISE)*, 2014.

Brown, Christina, Supreet Kaur, Geeta Kingdon, and Heather Schofield, "Cognitive Endurance as Human Capital," *The Quarterly Journal of Economics*, 2025, *140* (2), 943–1002.

Eckert, Max, Elisabeth Wilson, Helen Abadzi, and Soohyun Jeon, "Improving Arabic Reading Fluency: Results from Iqra, an Early-Grade Reading Intervention in Ras Al Khaimah," Technical Report 39, Al Qasimi Foundation August 2020.

Evans, David K. and Fei Yuan, "How big are effect sizes in international education studies?," *Educational Evaluation and Policy Analysis*, 2022, *44* (3), 532–540. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Eviatar, Zohar and Raphiq Ibrahim, "Why Is It Hard to Read Arabic?," in "Handbook of Arabic Literacy," Springer, 2014, pp. 77–96.

John, Anett and Todd Pugatch, "Promoting Literacy Through Cognitive Science: Experimental Evidence from the United Arab Emirates," 2024.

# 10 Annex

## 10.1 Schools by Type and Curriculum

| Type | School | Curriculum |
|---|---|---|
| International | RAK Academy – Al Rams | British |
| | RAK Academy – Khuzam | British |
| | RAK Academy – Al Hamra | British |
| Private | Wellspring School | American |
| | Harvest Private School | British |
| | Emirates National School RAK Campus | IB |
| | New Indian School | Indian |
| | Scholars Indian Private School | Indian |
| | Ideal English School | Indian |
| | Alpha Private School | Indian |
| | Indian Public High School | Indian |
| | Indian Private School | Indian |
| | Pakistan Higher Secondary School | Pakistani |
| | Sheikha Hessa Bint Saqr Private School | Ministry of Education |
| | Al Salaf Al Saleh Private School | Ministry of Education |
| | Al Nasr Private School – Mamourah Branch | Ministry of Education |
| | Al Nasr Private School – Julfar Branch | Ministry of Education |
| | Sheikh Saud Private School – Branch 1 | Ministry of Education |
| | Sheikh Saud Private School – Branch 3 | Ministry of Education |
| Public | Ahmed bin Majed School | Ministry of Education |
| | Al Rams School | Ministry of Education |
| | Ethen School | Ministry of Education |
| | Khatt School | Ministry of Education |
| | Mahra Bint Ahmed School | Ministry of Education |
| | Shamal School | Ministry of Education |
| | Al Dhait School | Ministry of Education |

## 10.2 Practices Description

### 10.2.1 Iqra Pedagogy Index

| Practice | Description |
|---|---|
| Reviewing Previous Lesson | Review the previous lesson before starting a new lesson. |
| Teaching One Letter per Lesson | Teach only one new letter or syllable per class. |
| Reading Aloud with Students | Read aloud with students. |
| Presenting Letter Combinations | Present words/syllables showing how the target letter connects. |
| Instructing Individual Reading | Students read contents individually two or more times. |
| Using Only Known Material | Students read texts containing only previously taught letters/syllables. |
| Using Index Finger to Follow Text | Students follow the text with their index fingers. |
| Giving Immediate Feedback | Provide immediate feedback to individual students. |
| Formally Assessing Students | Formally assess Arabic literacy during the academic year. |
| Grouping by Abilities | Group students by ability and tailor teaching strategies accordingly. |
| Sending Activities to Parents | Follow up with parents of struggling students with at-home activities. |
| Customised Activities for Weaker Students | Provide customised at-home learning activities for weaker students. |
| Pairing Students for Practice | Pair students with each other to practice reading. |
| Following Up with Parents of All Students | Follow up with parents of all students for at-home learning. |

### 10.2.2 MoE Pedagogy Index

| Practice | Description |
|---|---|
| Writing Lesson Objectives | Write lesson objectives on the board. |
| Using Songs | Incorporate songs into the lesson. |
| Using Pictures | Incorporate pictures into the lesson. |
| Using Stories | Incorporate stories into the lesson. |
| Using Multimedia | Incorporate multimedia (video or audio recordings) into the lesson. |
| Teaching Arabic Writing | Teach students to write in Arabic. |

### 10.2.3 Iqra Materials Index

| Practice | Description |
|---|---|
| Iqra Textbook as Primary | Use Iqra textbook as the primary Arabic textbook (binary). |
| Using Visual Aids | Use a visual aid, such as cards, presentation slides, or projector, to display writing. |
| Using Materials with Large Fonts | Display writing in large font. |

### 10.2.4 Iqra Perceptions Index

| Curriculum | Perception (*Perceived Importance of*) | Description |
|---|---|---|
| IQRA Practices | Reviewing Previous Lesson | Review previous lesson before starting new lesson. |
| | Teaching One Letter per Lesson | Teach only one new letter or syllable per lesson. |
| | Reading Aloud with Students | Read aloud with students. |
| | Presenting Letter Combinations | Present a variety of words or syllables that show how the target letter connects with other letters. |
| | Instructing Individual Reading | Instruct students to read the contents of the lesson individually two or more times. |
| | Using Only Known Material | Students are only asked to read texts which contain exclusively letters/syllables previously taught. |
| | Using Index Finger to Follow Text | Instruct students to follow reading with index fingers. |
| | Giving Immediate Feedback | Provide immediate feedback to individual students on reading. |
| | Formally Assessing Students | Formally assess students on Arabic literacy outside main exams. |
| | Grouping by Abilities | Group students according to their learning abilities and tailor teaching strategies to support their needs. |
| | Sending Activities to Parents | Follow up with parents of struggling students to provide activities for at-home learning. |
| | Customised Activities for Parents of Weaker Students | Follow up with parents of weaker students to provide customised activities for at-home learning. |
| | Pairing Students for Practice | Pair students with each other to practise reading. |
| | Following up with Parents of All Students | Follow up with parents of all students to provide activities for at-home learning. |
| | Using Visual Aids | Use a visual aid, such as cards, presentation slides, or projector, to display writing. |
| | Displaying Large Font | Display writing in large font. |
| | Ordering Letters Logically | Logical ordering of letters when introduced for the first time. |
| | Pronouncing Correctly (Teacher) | Correct pronunciation by teacher. |
| | Pronouncing Correctly (Students) | Correct pronunciation by student. |
| | Walking Around Classroom | Walk around classroom to observe student reading. |
| | Helping Weaker Students | Assist weaker students with reading. |
| | Encouraging Home Reading | Meeting with parents to encourage reading practice at home. |
| MOE Practices | Writing Lesson Objectives | Write lesson objectives on the board. |
| | Using Songs | Incorporate songs into the lesson. |
| | Using Pictures | Incorporate pictures into the lesson. |
| | Using Stories | Incorporate stories into the lesson. |
| | Using Multimedia | Incorporate multimedia (video or audio recordings) into the lesson. |