WORKING PAPER

Developing Arabic Language Skills in Egypt

Assessing the Implementation and Impact of the National Programme for Developing Arabic Language Skills

Contents

Executive summary 4

Introduction 7

1.1 National Programme for Developing Arabic Language Skills (NPDA) 7

1.1.1 Overview 7

1.1.2 Eligibility and enrolment 8

1.1.3 NPDA components 8

1.1.4 Implementation research 9

Research Methods 9

2.1 Research objectives 9

2.2 Research methods 10

2.3 Limitations 10

Results 11

3.1 Learning assessments 11

3.1.1 Baseline data 12

3.1.2 Baseline characteristics of NPDA participants 15

3.1.3 Endline learning assessments 16

3.2 Progression of learning amongst NPDA participants 16

3.2.1 Progression of learning for NPDA participants by grade 18

3.2.2 Progression of learning for NPDA participants by governorate 18

3.3 Impact of the NPDA on learning outcomes 20

3.3.1 Quasi-experimental method 20

3.3.2 Regression discontinuity design (RDD) model 21

3.3.3 The impact on learning outcomes 21

3.3.4 Impact results by gender, governorate and grade 22

3.4 Relationship between baseline scores and endline outcomes 24

3.4.1 Outcomes for students scoring zero at baseline 24

3.4.2 Ordinary least square results on the relationship between baseline and endline scores 25

3.5 Understanding implementation 25

3.5.1 School monitoring visits 25

3.5.2 Classroom and school characteristics 26

3.5.3 Classroom observation of teaching practices 27

3.5.4 Classroom observation of teaching materials and tools across governorates 29

3.5.5 Classroom observation of teaching activities 29

3.5.6 Stakeholder feedback 30

Conclusion **33**

Key findings 33

Recommendations 35

Annexes **36**

Endnotes **49**

Executive summary

Foundational literacy is a crucial step in a child's development. The shift from "learning to read" to "reading to learn" marks a critical turning point, opening the door to all subsequent steps of their education. Yet in Egypt, nearly six in ten children are unable to read or write a simple story by the age of 10 (PIRLS, 2021). This literacy gap limits their ability to acquire the skills needed to participate in further learning, and subsequently their participation in society and the economy.

In February 2025, the National Programme for Developing Arabic Language Skills (NPDA) was launched by the Ministry of Education and Technical Education (MoETE) as an urgent response to the foundational literacy gap. Over a 12-week period, the programme reached 199,225 students in grades 3 to 6, delivering five hours of structured literacy instruction per week (equivalent to 60 hours in total). NPDA provided a comprehensive approach, combining teacher training (provided by the Professional Academy for Teachers), structured lesson plans, and dedicated teaching and learning materials. Implementation was led and monitored by MoETE in collaboration with the National Center for Examinations and Educational Evaluation (NCEEE).

This report presents the key findings and recommendations from research conducted on the first stage of the NPDA programme in 980 schools across 10 governorates. The selection of the governorates was based on the lowest scores achieved on a Grade 4 Arabic and mathematics assessment that was conducted by the NCEEE in October, 2023.

Key findings

1. **The NPDA had a substantial positive impact on literacy outcomes.**

- Using a quasi-experimental Regression Discontinuity Design (RDD) the programme is estimated to have increased literacy scores by 17 per cent, significant at the 1 per cent level, meaning there is less than a 1 per cent chance that the result is occurring by chance.

- This estimate is conservative, additional analysis found that students with the lowest baseline scores had the largest improvements.

- The RDD analysis also finds larger impacts of the programme for girls than boys.

2. **There were large learning gains for NPDA participants across literacy domains, on average, scores increased by 50 percentage points.**

- The proportion of students who scored zero dropped from 33 per cent (or one in three) at baseline to 1.7 per cent at endline.

- Improvements were seen across all five literacy domains with the highest increases seen in Reading Comprehension and Fluency, followed by Spelling, Reading Letters and Words and Sound Awareness.

3. **Arabic literacy levels varied by gender, grade and governorate.**

- Girls and students from higher grades scored better at baseline, the initial Arabic assessment, making them less likely to qualify for the programme. As a result, 40 per cent of participants were girls.

- Among those who did participate in the programme, girls showed greater gains than boys, especially those starting from a zero score.

- Although all governorates showed strong improvements, some differences in outcomes and implementation were observed, warranting further exploration.

4. **Implementation and pedagogical uptake levels were high.**

- Classroom observations showed fidelity to the programme design, this is corroborated by feedback from teachers of whom 95 per cent reported that they were able to implement according to the timeline.

- Observations show that teachers increasingly used recommended pedagogical techniques over time, suggesting growing confidence.

- Observations showed that use of techniques and materials varied by region, highlighting the need for localized support.

5. **Feedback from parents, teachers and administrators was all overwhelmingly positive.**

- There were however some key challenges that emerged from feedback data which should be further explored for the next stage of the programme. Key challenges identified include teachers lacking consistent access to materials (21 per cent), teachers reporting insufficient initial training (15 per cent), and inadequate levels of teacher incentives (43 per cent).

Recommendations

1. **Institutionalize Arabic literacy remediation programme in the education system**

- Phase 1 of the NPDA has showcased the potential impact of remediation interventions, especially those embedded within school hours. Acceleration and institutionalization of such interventions are critical to support Egypt reach its commitment of halving learning poverty (56 per cent) by 2029 to be at 27 per cent.

2. **Track long-term outcomes**

- Conduct follow-up assessments to understand the sustainability of learning gains and long-term impact.

- This research demonstrates the strong initial impact of the NPDA. It will be important to understand how this investment by the MoETE will be sustained and if learners continue their learning growth trajectory.

3. **Strengthen support for teachers**

- Address gaps in training, materials, and incentives, and tailor support to the needs of different regions.

- Feedback from teachers illustrates that there are areas for improvement, the uneven use of materials is also seen in classroom observation data. Understanding the reasons for these disparities and what can be done to support teachers from different localities can refine the programme in the subsequent stage of implementation.

4. **Study cost-effectiveness and implementation variation**

- Identify what types of support for teachers are most cost effective across different settings.

- Putting in place a research protocol to evaluate various adaptations in programme support would allow the programme to adapt and course correct.

5. **Investigate spillover into regular classrooms**

- The NPDA was implemented by teachers, and each school implemented the programme at an hour that best suited their learning environment. This meant that the learning hour could vary to anytime of the day (e.g. first hour, last hour, after school).

- To understand the broader impact of the NPDA it is critical to understand how the NPDA may influence teachers' daily teaching practices in their regular classrooms.

I. Introduction

Reading proficiency is a gateway for student learning and is a foundational skill that paves the way for children to be successful in school and beyond. In Egypt, however, more than half of children (56 per cent) are unable to read a simple text by the age of 10.[1] While improvements in reading literacy have been observed in international assessments such as the Progress in International Reading Literacy Study (PIRLS),[2] 45 per cent of fourth graders in Egypt scored at the low reading proficiency benchmark in 2021.[3] Gender disparities also persist, with girls outperforming boys by 16 points in the same assessment year.[4] Similar trends were observed in a national examination conducted in 2021 by Egypt's National Center for Examination and Education Evaluation (NCEEE); 30.2 per cent of fourth grade students did not achieve the minimum expected level of skills and knowledge, while 47.3 per cent only met the minimum level.[5]

In response to the need to improve foundational literacy in primary education, Egypt's Ministry of Education and Technical Education (MoETE) in partnership with the National Center for Examination and Education Evaluation (NCEEE) and UNICEF, launched the "National Programme for Developing Arabic Language Skills" (NPDA) in February 2025. NPDA is a structured remedial education programme designed to improve foundational Arabic literacy among students in Grades 3 to 6.

NPDA was designed for phased implementation. Phase I ran from February to May 2025 and was implemented in 980 schools representing 10 governorates. It assessed 437,857 students, of whom 199,225 were identified as eligible for NPDA support. Approximately 8,000 Arabic language teachers also received programme support. Phase II, scheduled from October 2025 to December 2026, will scale the programme to reach an additional 1,000 primary schools across additional governorates, targeting a further half million students.[6]

This report presents key findings from research embedded into the first stage of NPDA implementation, led by NCEEE and further analysed by UNICEF Innocenti - Global Office of Research and Foresight. The primary objective of the research was to generate evidence on implementation and programme effectiveness to inform the planned scale-up.

1.1 National Programme for Developing Arabic Language Skills (NPDA)

1.1.1 Overview

The NPDA is a structured Arabic literacy programme and is comprised of weekly teacher guides and structured session guidance, capacity building for teachers, on-the-job mentoring and coaching, and enrichment readings. After teachers received online training on the programme, eligible students received a total of 60 sessions, each lasting 45 minutes to one-hour. The sessions were mostly conducted five days a week, over a 12-week period, with some variation in school-level scheduling.

The teacher training equipped teachers with a wide range of instructional strategies, including cooperative learning, brainstorming, case study analysis, peer learning, decision-making exercises, role play, and inquiry-based approaches. Teachers were provided with a dedicated online facilitator manual that included session-by-session objectives, instructional methods, and expected learning outcomes for each session.

1.1.2 Eligibility and enrolment

Student eligibility for NPDA was determined through a standardized literacy assessment. Students scoring below 36 out of 60 points (the assessment benchmark) were selected to participate in the programme. The NCEEE conducted a baseline assessment with 437,874 students across 10 governorates. Of these, 199,225 students (45.5 per cent) scored below the threshold and were deemed eligible. Endline assessment data was collected in June 2025. Comparable baseline and endline data are available for 189,938 students, comprising 185,697 programme participants and a sample of 4,241 non-participants enabling analysis of programme impact. The initial selection of schools that were invited to participate in the programme was based on the results of the Grade 4 Arabic and maths basic assessment report in October, 2023.

1.1.3 NPDA components

The programme was implemented over three months (March–May 2025), involving a wide range of stakeholders including teachers, school principals, education departments, the Central Administration for Public Education, NCEEE, and UNICEF. Teacher-led instruction in classrooms was structured into foundational, constructivist, and expressive phases, with 60 lessons adapted to student levels. These were distributed as follows: five approximately one-hour teaching sessions per week, with a total of 20 lessons monthly.

A variety of educational aids and teaching strategies, such as direct instruction, play-based and cooperative learning, and guided reading were used to engage students. Teachers were provided with a weekly teaching plan that can be modified based on the needs of students and their levels of progress. Plans included a focus on practical exercises and interactive activities to enhance student learning, and the need to use educational games to strengthen understanding and comprehension. Teachers were also required to closely monitor student performance throughout implementation, to offer feedback, foster positive participation and to follow up on student progress. Teachers were provided with tools and strategies to continuously assess student learning. In addition, 47 Performance Improvement centres were introduced across governorates, which were open on weekends to provide more support to students who needed it. The centres were established at the *idaras* (administrative) level, within a cluster of schools as part of the "Comprehensive Inclusion Programme", supported by UNICEF, to provide ongoing technical support to neighbouring schools. Monitoring and evaluation were integrated throughout the programme, with daily, weekly, and monthly assessments using reading tests, observations, and activities to guide teaching and student grouping. Monthly tests were validated and used to track progress and adjust instruction. Quality control and technical support were central to ensuring the NPDA was implemented according to the required standards. These processes were designed to monitor progress, identify areas for improvement, and provide real-time guidance to support effective delivery across schools.

A comprehensive system of supervision, testing, and continuous monitoring and evaluation was established with several key objectives: to track actual implementation on the ground, verify adherence to quality standards, reduce errors, enhance participant performance, and provide timely technical support and recommendations. To operationalize this, 11 weekly quality control and technical support forms were developed capturing comprehensive school-level data, teacher practices, student attendance, use of learning materials, and indicators of student progress.

A dedicated technical support team of approximately 250 members was formed, comprising representatives from the MoETE, NCEEE, and educational directorates and departments. Support team members conducted weekly school visits to monitor implementation, provide feedback, and offer hands-on support to teachers. Each visit covered administrative and instructional practices, with a focus on support rather than inspection. Weekly reports documented attendance, instructional practices, and student learning, and were shared with the MoETE to inform decision-making.

1.1.4 Implementation research

This research is part of a collaborative effort between the MoETE, NCEEE, UNICEF Egypt, and UNICEF Innocenti - Global Office of Research and Foresight and will be used to inform the second stage of the NPDA and its expansion beyond the 980 schools. Design and implementation of the programme, data collection, and processing was led by the MoETE/NCEEE. UNICEF Egypt provided technical support to the design of the programme and advice to the implementation and coordinated between the MoETE, NCEEE and the research team at Innocenti. UNICEF Innocenti designed a research strategy utilizing data being collected by the programme, conducted the impact assessment and implementation research, and produced analysis and a research report based on the findings.

II. Research Methods

2.1 Research objectives

This research was embedded within programme implementation and data were collected to answer three key research questions:

1) **Impact -** What is the impact of the NPDA on children's learning outcomes?

2) **Implementation -** How did implementation vary across schools, and what factors supported or hindered improvements in learning?

3) **Sustainability and scale** - What programme adaptations could strengthen its implementation as it scales?

2.2 Research methods

A mixed-methods implementation research approach was applied to address three research objectives. To test the effectiveness of the programme on learning outcomes, a quasi-experimental regression discontinuity design (RDD) approach was used. Quantitative and qualitative data collected during implementation; including baseline and endline assessments, classroom observations, and stakeholder surveys (teacher, headteachers, and parents), were analysed to understand how the programme was implemented.

1) Impact. What is the impact of NPDA on children's learning outcomes?

The impact of the NPDA on children's learning outcomes was measured using empirical analysis with econometric techniques. Specifically, a RDD was applied to measure the causal effect of the programme on children's literacy skills. As detailed in Section 3.1.2., the eligibility cutoff score of 36 out of 60 points served as the running variable for the RDD analysis.

2) Implementation. How did implementation vary across schools, and what were the factors that led to or hindered improvements in learning outcomes?

Fidelity of implementation was examined, focusing on how teachers utilized the provided resources in practice and identification of factors at the school, teacher, and contextual levels that influence execution. To conduct this analysis the research draws from:

1. Observation data from regular quality assurance monitoring visits from the MoETE and NCEEE.

2. Feedback data from school principals, teachers, parents and quality assurance monitors at the end of the programme.

These data were matched with learning data to track how implementation factors and stakeholder perceptions were associated with different outcomes.

3) Sustainability and scale. What are adaptations that the programme could make to strengthen implementation as it scales?

This study utilizes the combination of insights from the implementation and impact studies to provide recommendations to inform improvements in Phase II. This research domain will also utilize additional qualitative data collection to dig deeper into implementation, which has not yet been conducted at the time of the drafting of this report.

2.3 Limitations

1. At the time of reporting, qualitative data on participant perception was limited. This constrained the ability to explore in-depth views on implementation challenges and programme strengths. This qualitative data collection is scheduled and will be utilized in informing the programme's scale up and an updated version of this report.

2. A regression discontinuity design (RDD) allows for causal estimates of programme impact but only for students near the eligibility cutoff (36 out of 60 points). It does not measure the impact for students scoring far below the cutoff. Notably, about 15 per cent, of the more than 400,000 baseline test-takers scored zero. To address this, supplementary analysis, including NCEEE data were used to examine outcomes for students at varying baseline levels. Findings strongly suggest that the programme's positive impact was particularly pronounced for lower-performing students.

3. Due to logistical reasons endline data collection for students who were not in the NPDA (the control group) was collected from just four out of 10 governorates. This means it was not possible to conduct impact analysis across all 10 governorates in which the programme took place.

4. Additionally, while the research was able to draw on rich monitoring data from NCEEE monitors, it is important to note that as NCEEE monitors were part of the government's programme implementation team. Thus, responses from teachers and administrators to questions may face some level of social desirability bias, meaning the tendency to evaluate the programme based on socially desirable qualities rather than objective performance criteria.

III. Results

This section presents results from the NPDA analysis. It begins by describing how baseline and endline assessment data were collected and analysed. The analysis then explores learning progression among NPDA participants and assesses the programme's impact on learning outcomes. Finally, the section examines key aspects of programme implementation.

3.1 Learning assessments

All participants across Grades 3-6 completed the same foundational literacy assessment. Each student was assessed and scored across five different sub-topics in basic literacy skills:

1. Reading letters and words

2. Sound awareness (phonological awareness)

3. Fluency

4. Reading comprehension

5. Orthography

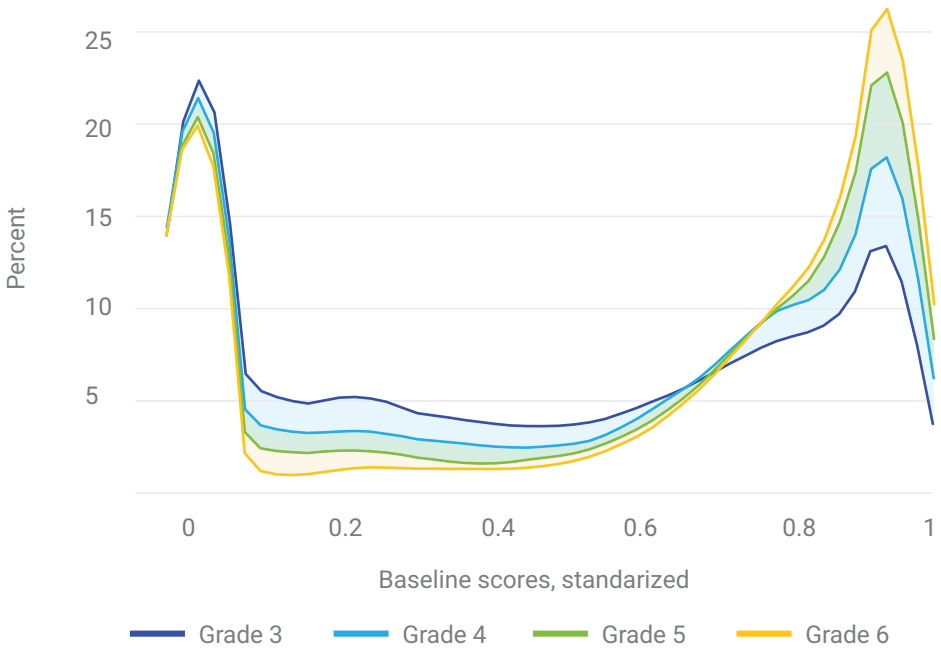
3.1.1 Baseline data

Baseline learning assessments were collected with 437,857 students (46 per cent female) across 980 schools representing 59 administrative regions (*idaras*), within 10 governorates: Alexandria, Aswan, Asyut, Cairo, Damietta, Faiyum, Giza, Ismailia, North Sinai, and Sohag. The schools and students were concentrated in three governorates, Sohag (18.3 per cent), Cairo (15.2 per cent), and Giza (12.4 per cent), representing nearly half (47 per cent) of all assessed students. Results revealed significant variation in Arabic literacy levels. A large proportion (16 per cent) of students scored zero at baseline (*Table 1*), and one out of five students scored zero on at least one of the five subdomains. While there is a considerable proportion of students that scored zero, 7 per cent achieved a perfect score and around one quarter of students scored more than 90 per cent. This produced a U-shaped distribution of scores, with large proportions scoring either very low or very high, as illustrated by the U-shaped distribution of baseline scores in Annex 1. The pattern was consistent across Grades 3-6 (*Figure 1*). There is also a modest trend of students in higher grades achieving higher scores on average.

Table 1: Scores from the baseline assessment

SCORE	NUMBER OF STUDENTS	% TOTAL
Zero	70,951	16%
Below 36 points	128,266	29%
36 and above	238,640	55%
Total	**437,857**	100%

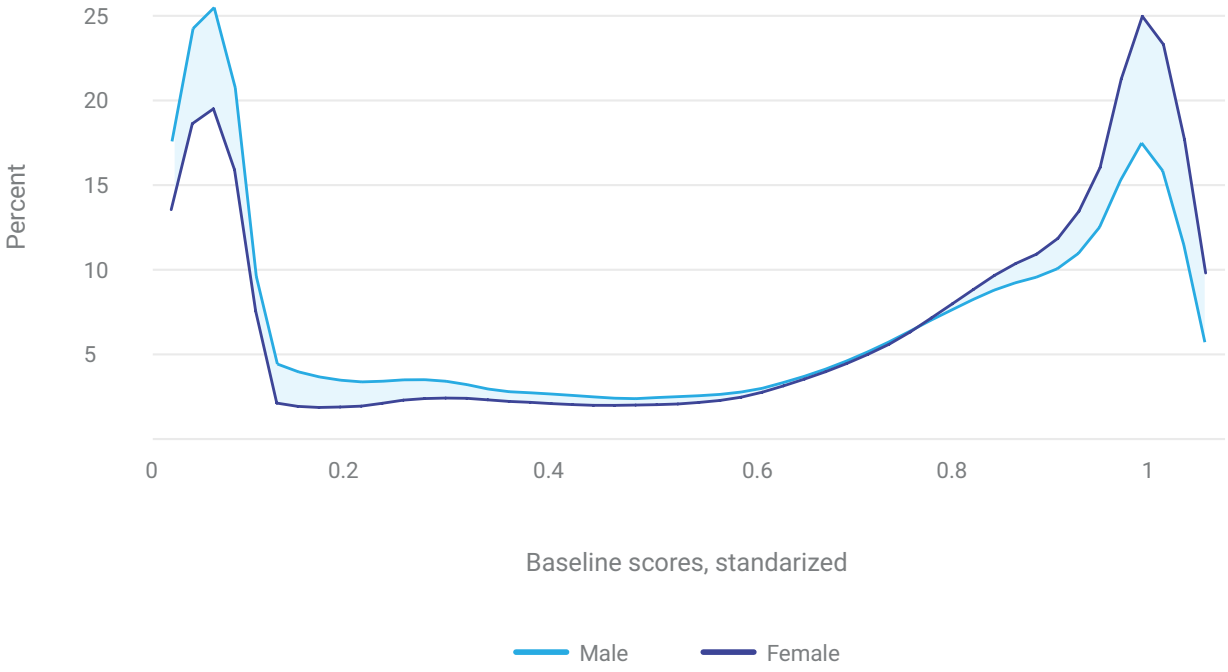
Figure 1: Distribution of baseline scores by grade



Source: Authors' analysis of baseline assessments, the sample used for the above figure includes all students who completed the baseline literacy assessment (N=438,875 students).

Analysis of baseline scores revealed clear differences in student performance by gender, grade and location. Figure 2 shows standardized baseline scores for test takers, by gender. Both boys and girls showed a U-shaped score distribution with clustering at 0 per cent and 100 per cent, but girls were more often at the top of the range while boys were more clustered at the bottom. On average, girls outperformed boys, also across all governorates (*Annex 2*). Because girls scored higher than boys at baseline, fewer met the eligibility threshold of scoring below 36 points. As a result, girls were slightly underrepresented in the NPDA, accounting for 41 per cent (77,506) while the share of boys was 59 per cent (112,440).

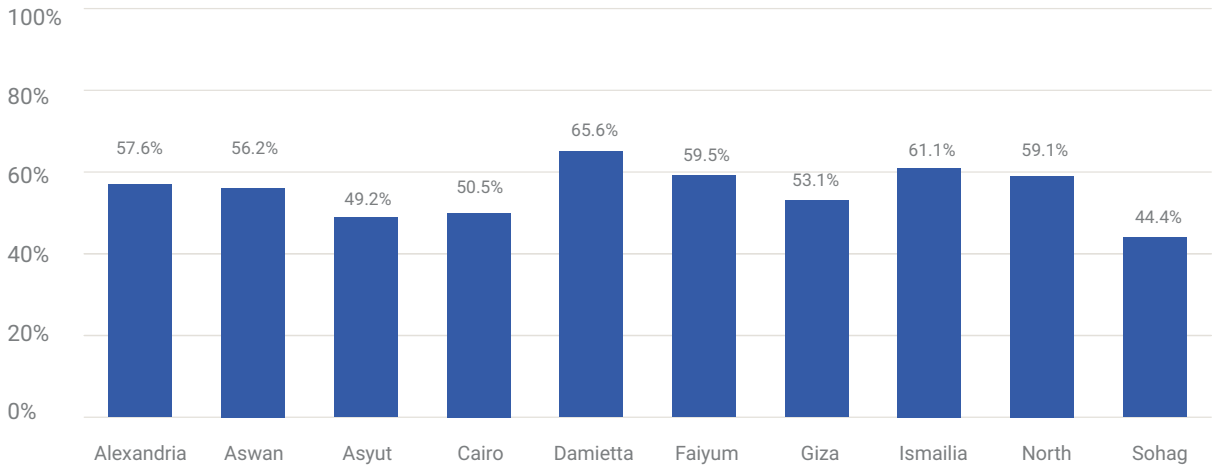
Figure 2: Baseline score, standardized, distribution by gender



Source: Authors' analysis of baseline assessments, the sample used for the above figure includes all students who completed the baseline literacy assessment (N=438,875 students).

Further disaggregation by grade level and governorate reveals regional disparities. There is a large variation in mean scores across governorates, with the highest aggregate mean score in Damietta, and the lowest in Sohag (*Figure 3*). When exploring disaggregation by both governorate and grade level, students in higher grades tended to score higher on the baseline assessment across each governorate (*see Annex 3*). A boxplot showing the distribution of scores by governorate is included in Annex 4.

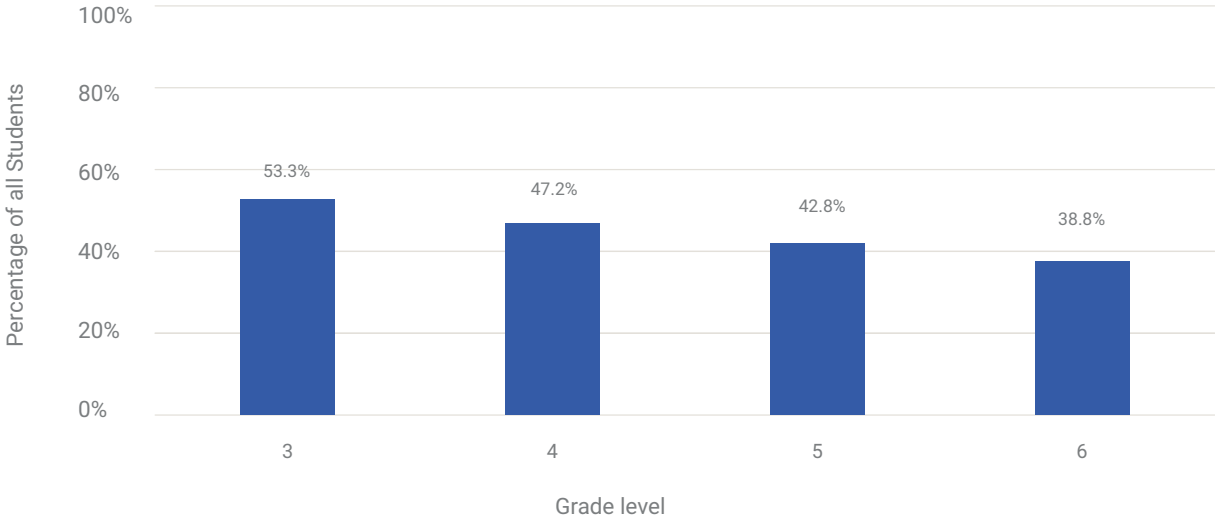
Figure 3: Mean baseline score, by governorate



Source: Authors' analysis of baseline assessments, the sample used for the above figure includes all students who completed the baseline literacy assessment (N=438,875 students).

Students in higher grades were more likely to meet the assessment benchmark. Thus, more students in lower grades were eligible for and received the programme, 53.3 per cent of Grade 3 students needed support and were entered into the programme, compared to 38.8 per cent of Grade 6 students (*Figure 4*).

Figure 4: Per cent of all assessed students who were eligible for the programme, by grade



Source: Authors' analysis of baseline assessments, the sample used for the above figure includes all students who completed the baseline literacy assessment (N=438,875 students).

3.1.2 Baseline characteristics of NPDA participants

This section presents the baseline characteristics of the students who were deemed eligible based on their performance in the baseline assessment. Overall, Sohag had the most NPDA participants with 46,423 students from that governorate, followed by Cairo with 28,344 and Giza with 23,526. This can be seen by grade level in Table 2.

Table 2: Number of NPDA students per governorate and grade

	GRADE LEVEL				
GOVERNORATE	3	4	5	6	TOTAL
Alexandria	4,191	3,980	4,014	3,510	15,695
Aswan	3,491	3,155	2,933	2,720	12,299
Asyut	5,598	5,493	4,873	4,492	20,456
Cairo	8,108	7,718	6,401	6,117	28,344
Damietta	2,824	2,697	2,637	2,405	10,563
Faiyum	4,884	1,668	1,554	1,474	9,580
Giza	6,596	6,027	5,641	5,262	23,526
Ismailia	4,076	3,710	2,963	2,626	13,375
North Sinai	1,628	1,482	1,230	1,096	5,436
Sohag	13,301	12,093	11,206	9,823	46,423
Total	**54,697**	**48,023**	**43,452**	**39,525**	**185,697**

Source: Authors' analysis of baseline assessments, the sample used for the above figure includes students who met the eligibility for the NPDA programme (N=185,697 students). 40.8 per cent of these students are female.

Overall, NPDA students from Alexandria had the lowest average scores at baseline. On a percentage basis, students in Alexandria on average scored just over 10 per cent out of 100 in their baseline test, while students in Ismaila scored the highest on average at just 21 per cent (*see Annex 5*). These averages are driven by large amounts of students who scored zero at baseline. Annex 6 also shows the per cent of students who scored zero on the baseline assessment by governorate, grade, and gender. Interestingly a higher number of students in Grade 6 scored zero (41 per cent) than in Grade 3 (27 per cent), so while overall there are less Grade 6 students that scored low enough to be eligible to be in the programme, those that were eligible had high degrees of need. In Alexandria, 68 per cent of NPDA eligible students scored zero at baseline, this is the highest amount of any governorate, followed by 48 per cent in Giza.

3.1.3 Endline learning assessments

Endline learning assessments were conducted for 185,697 students who were eligible and took part in the NPDA. In addition, assessments were carried out for a sample of 4,241 students who scored above 36 points and were therefore ineligible for the NPDA. The endline learning assessment was the same as the pre-assessment.[7] Table 3 shows summary statistics of the sample size of the treatment group and control group, as well as the percentage of female students, by governorate. It also shows the percentage of baseline students who had a zero score, as well as the baseline and endline scores and their standard deviation.

Table 3: Overview of baseline and endline assessments of NPDA participants by region

	SAMPLE SIZE				TOTAL	BASELINE ASSESSMENT (MIN=0, MAX=35)		ENDLINE ASSESSMENT (MIN=0, MAX=60)		
	TREAT-MENT	% FEMALE	CONTROL	% FEMALE		% ZERO SCORES	MEAN	SD	MEAN	SD
Alexandria	15,695	33.1	887	20.7	15,695	65%	7.1	11.8	50.3	12.4
Aswan	12,299	40.6			12,299	18%	14	11.9	41.3	14.5
Asyut	20,456	42.7	1,003	53.8	20,456	21%	13.4	11.7	36.5	15.4
Cairo	28,344	42.5	1,245	40.7	28,344	31%	10.4	12.4	45.7	14.2
Damietta	10,563	39.4			10,563	42%	11.8	12.7	45.8	14.9
Faiyum	9,580	38.6			9,580	22%	15.1	12.1	44.5	12.4
Giza	23,526	35.7			23,526	48%	10.6	12.5	45.9	13.7
Ismailia	13,375	41.4			13,375	20%	15.1	12.4	42.8	16.2
North Sinai	5,436	42.8			5,436	26%	15.1	12.3	43.8	12.8
Sohag	46,423	44.4	1,106	49.3	46,423	28%	11.8	11.5	36	16.2
Total	**185,697**	**40.8**	**4,241**	**41.9**	**189,938**					

Source: Authors' analysis of baseline and endline assessments, the sample used for the above figure includes students who met the eligibility for the NPDA (N=185,679 students).

3.2 Progression of learning amongst NPDA participants

Students who participated in the NPDA achieved an average of 50 percentage points higher on the post-test compared to pre-test (70.12 per cent compared to 19.78 per cent, respectively). Positive gains were seen across all sub-domains of the assessment, with the largest gains seen in reading comprehension (59.4 per cent increase) (*Table 4*).

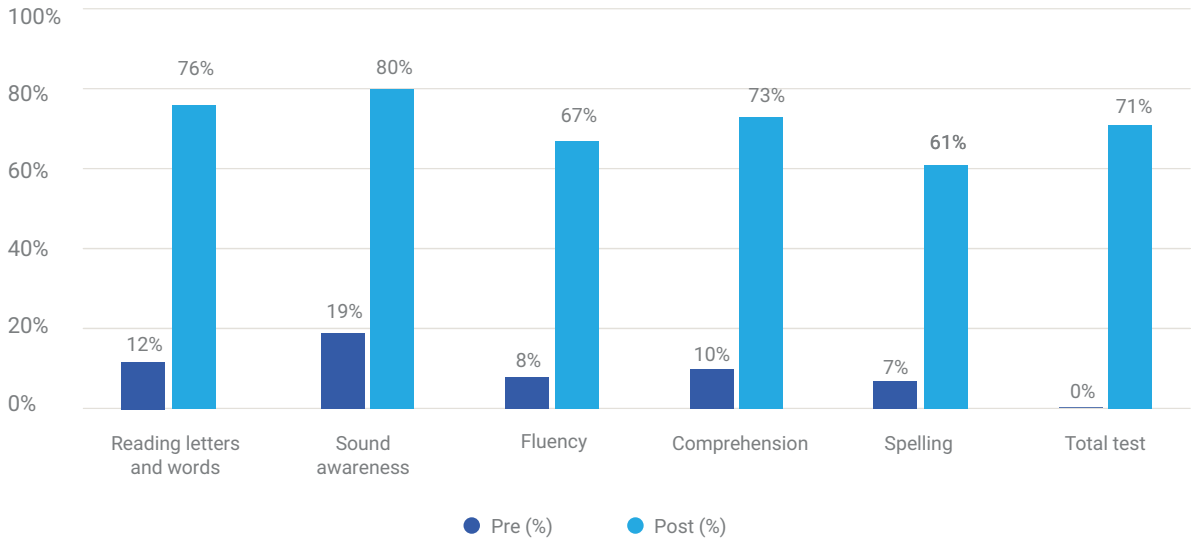
Table 4: Average scores at pre-test and post-test on the NPDA assessment overall and across the sub-domains

TEST DOMAIN (MAX)	BASELINE		ENDLINE		PERCENTAGE POINT CHANGE BETWEEN PRE- AND POST
	AVG.	% SCORE FROM SUB-DOMAIN	AVG.	% SCORE FROM SUB-DOMAIN	
Reading letters and words (18)	4.37	24.3%	13.19	73.3%	49.0
Sound awareness (17)	4.34	25.5%	12.67	74.5%	49.0
Fluency (5)	0.49	9.8%	3.18	63.6%	53.8
Reading comprehension (5)	0.57	11.4%	3.54	70.8%	59.4
Spelling (15)	2.11	14.1%	9.49	63.3%	49.2
Total (60)	**11.87**	**19.8%**	**42.07**	**70.1%**	**50.3**

Source: Authors' adaptation from the NCEEE report (N = 185,697).

These big increases also mean that there was a substantial increase in the proportion of students who surpassed the threshold score for each of the sub domains and overall. At endline 71 per cent of students who received the NPDA passed the minimum threshold compared to 0 per cent at baseline. This trend is also seen across the five sub-domains of the assessment (*Figure 5*), with increases in the proportion of students reaching the respective thresholds across each domain. Overall, Spelling and Fluency sub domains, which had the lowest pass rates in the baseline at 7 and 8 per cent respectively, remain the lowest in the endline though much improved with 61 per cent of students passing Spelling and 67 per cent passing Fluency.

Figure 5: Percentage of NPDA participants who reached the minimum competency threshold (36 out of 60 points) at pre- and post-test by test domain



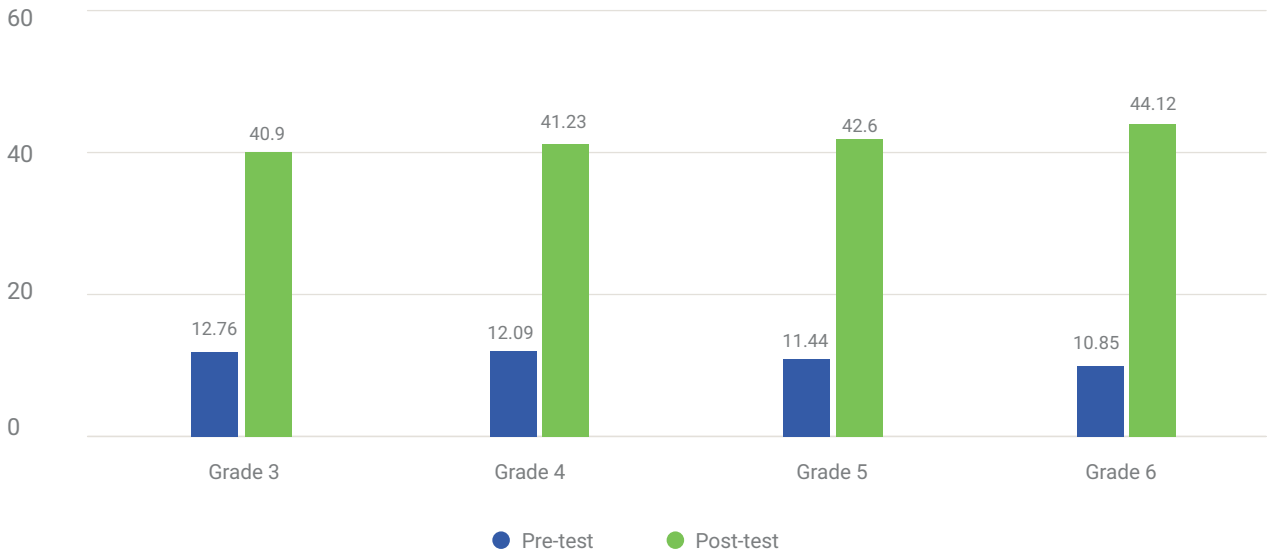
Source: Authors' adaptation from the NCEEE report and analysis of baseline data (N = 185,697).

Girls and boys demonstrated substantial improvements from pre- to post-test across all subdomains, with comparable gains in overall performance. Girls showed a slightly larger gain in overall scores, achieving a total of 43.6 at post-test, equivalent to a 52.5 per cent increase in overall score. Meanwhile, boys achieved an average of 41.0 at post-test, equivalent to a 48.8 per cent increase.

3.2.1 Progression of learning for NPDA participants by grade

Across all grade levels, students participating in the NPDA demonstrated improvement from pre- to post-test, and on average, achieved scores above the threshold value of 36 (*Figure 6*). Grade 6 students, who began with the lowest average pre-test score of 10.85, showed the highest gains, obtaining an average score of 44.12 at post-test, the highest across all grades. This reflects a 55.5 per cent increase in overall score for sixth grade students, compared with 51.9 per cent for students in fifth grade, 48.6 per cent for students in fourth grade and 46.9 per cent for students in third grade. A distribution of scores at baseline and endline by grade level is included in Annex 7.

Figure 6: Average scores on baseline and endline assessment for NPDA participants by grade level



Source: Authors' adaptation from the NCEEE report (N = 185,697).

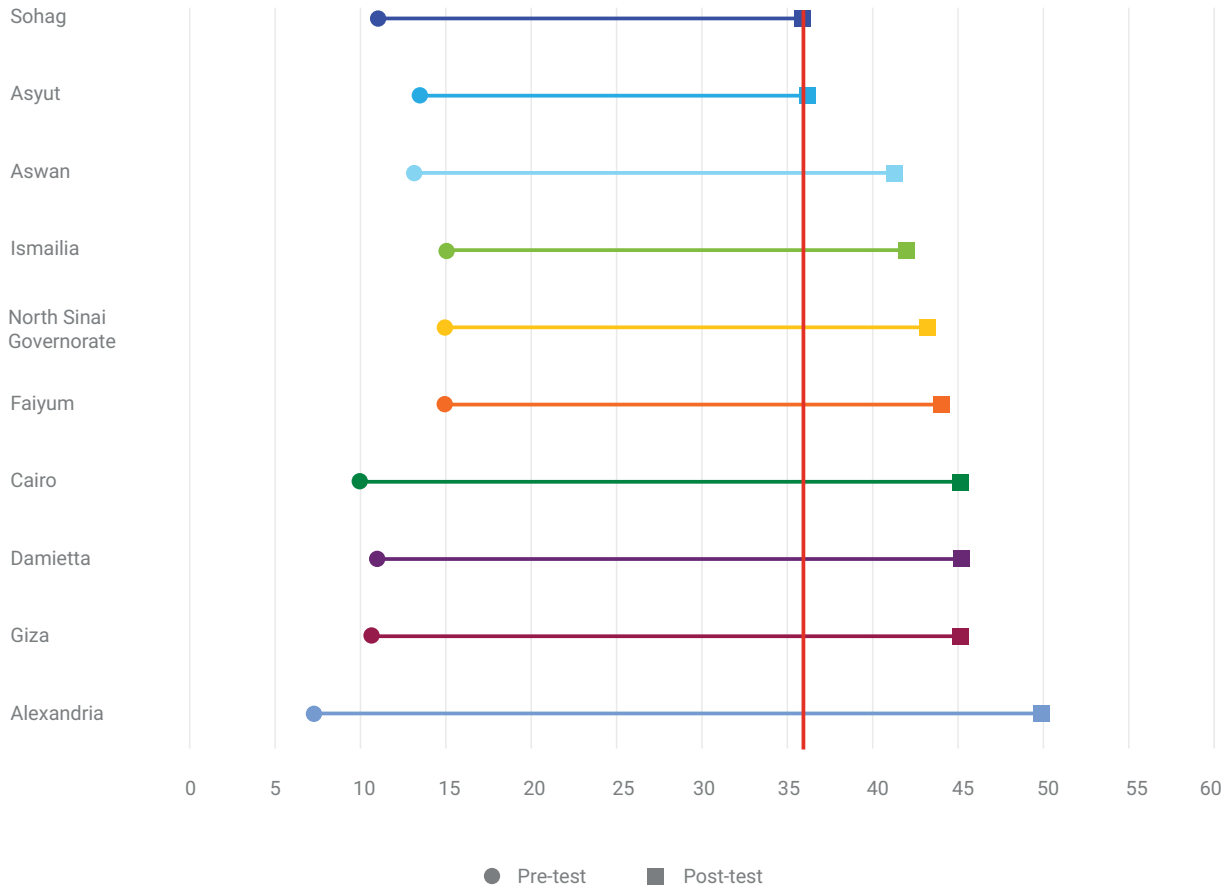
3.2.2 Progression of learning for NPDA participants by governorate

Across all participating governorates, students demonstrated notable gains in performance from pre-test to post-test. Students in the Alexandria governorate had the lowest average pre-test score of 7.14 and achieved the highest average score at post-test of 50.31, reflecting the most substantial improvement across governorates equivalent to a 72.0 per cent average increase in score (*Figure 7*).

Similarly, students in governorates such as Giza and Cairo started with low pre-test averages (10.59 and 10.45, respectively) but showed extensive gains, achieving post-test averages of 45.86 and 45.74, respectively, equivalent to an average increase of 59 per cent. Students in the governorates of Faiyum, North Sinai, and Ismailia, began with some of the higher pre-test scores (around 15), and showed improvement with post-test scores reaching between 42.82 and 44.53, or an increase of between 46 per cent to 50 per cent. Meanwhile, in Sohag and Asyut, it is observed that students improved from pre-test, yet in both governorates, students are on average just achieving the threshold value (36.01 and 36.52, respectively reflecting 40.4 per cent and 38.5 per cent increases).

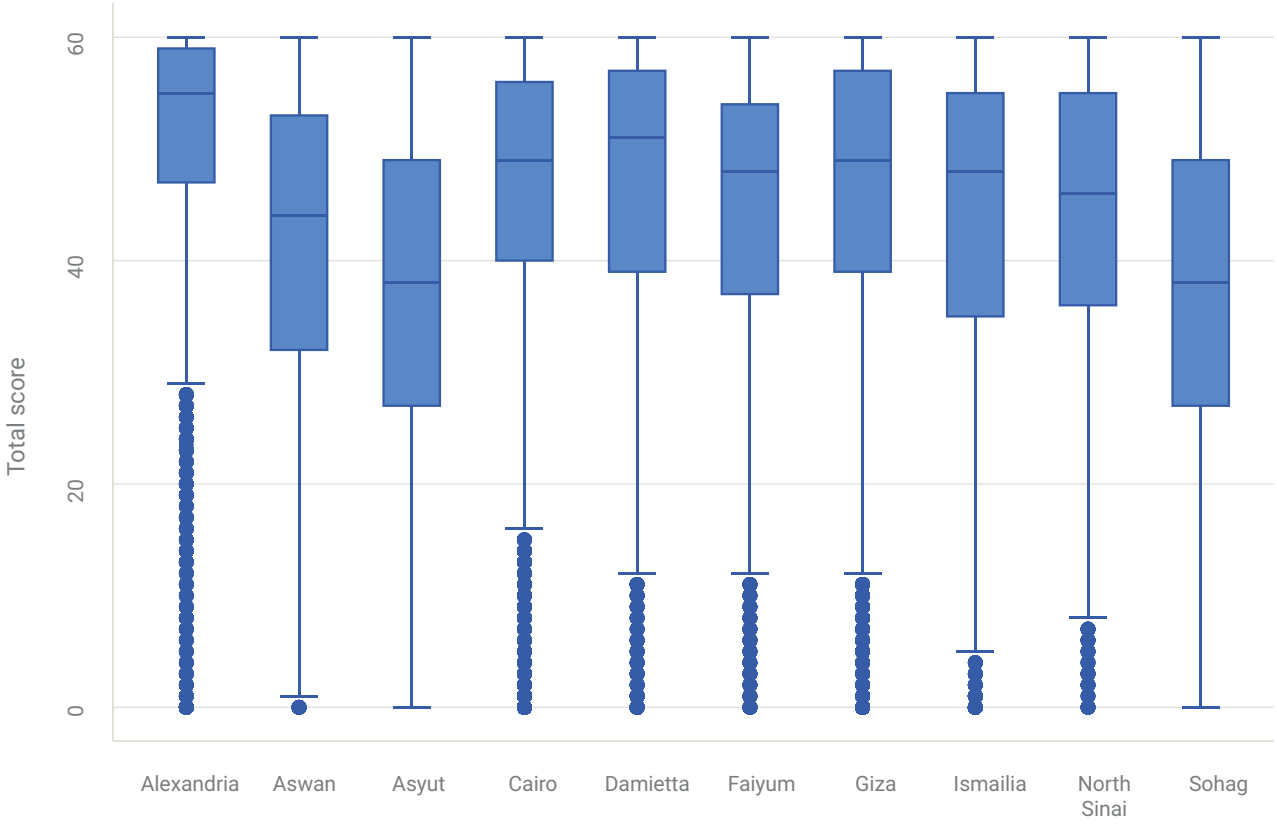
Overall, the data highlights consistent learning improvements across all governorates, with some of the largest gains observed in governorates that began with the lowest initial scores. Figure 8 shows a boxplot and median scores across governorates. It shows that not only was the mean high, but the distribution of scores for Alexandria was highly concentrated at the upper end of the test.

Figure 7: Average pre-test and post-test scores for students who participated in the NPDA, by governorate, with the vertical line showing the assessment cut-off value



Source: Authors' adaptation from the NCEEE report, N = 185,697 (full treatment that matched baseline and endline assessments).

Figure 8: Distribution of endline scores by governorate



Source: Authors' analysis of endline data, N = 185,697 (full treatment that is matched baseline and endline assessments).

3.3 Impact of the NPDA on learning outcomes

3.3.1 Quasi-experimental method

Section 3.1.4 presented an analysis of the results of the programme for NPDA participant students. This shows that participating students had large gains to their scores after the NPDA, scoring higher on each domain. What the analysis in section 3.1.4 does not do however is estimate the causal impact of the programme. The pre-post analysis of the programme cannot alone test impact as it does not account for what would have happened if the programme had not existed, what is known as the 'counter-factual'. This section utilizes data collected by the NCEEE at baseline and endline for both treatment and non-treatment students to understand the causal relationship between the NPDA and learning outcomes. To estimate the impact of the NPDA, a regression discontinuity design (RDD) identification strategy was used. The RDD technique enables the estimation of the causal impact of the programme by comparing students who scored just below and just above the eligibility cutoff. The RDD method assumes that the students just below and above a threshold are very similar to each other, making the comparison valid. The analysis focused on students within a narrow range (+/- 4 points) of the 36-point cutoff. This quasi-experimental method was utilized for three main reasons:

1) random allocation of students into treatment was not feasible or desirable due to programme design,

2) there was a clearly defined rule to assign students into treatment based on baseline test cutoff score of 36, and

3) the assignment of a student to the treatment or not, around the cutoff score is plausibly exogenous, meaning that no other characteristics of students, or other factors determined if students were given the treatment other than what they scored on the baseline test.

3.3.2 Regression discontinuity design (RDD) model

A parametric local[8] polynomial RDD was applied using a triangular kernel, which assigns greater weight to students closer to the cutoff score and less weight as one moves further away from the cutoff. Giving greater weight to those closer to the cutoff score is done as the assumption is that students who are closer to the cutoff scored very similar on baseline are more comparable. The treatment effect was estimated using a model that accounted for baseline scores, treatment status (whether they received the programme or not), and their interaction. This sharp RDD model captures the immediate effect of the intervention at the threshold and any differences in trends on either side of the cutoff. The effect was also modeled as a global linear RDD using ordinary least squares (OLS), yielding similar results (*Annex 9, Table 10*).. Specifically, the change in learning outcomes measured by the change in score of the endline assessment was modeled as:

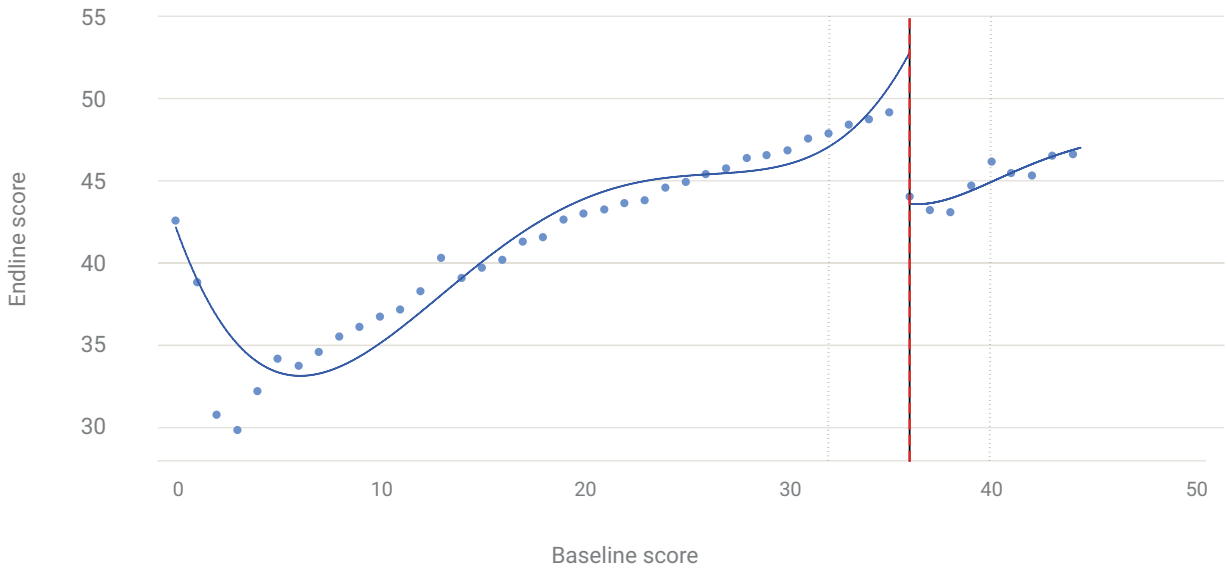
$$Post_i = \alpha + \beta 1 * Treatment + \beta 2 * Running + \beta 3 * (Running * Treatment) + \varepsilon_i.$$

In the above estimation, *Post* is the endline assessment score, which is the same assessment at baseline. *Treatment* is a binary variable taking the value of one if the baseline score was less than 36 points; 0 otherwise. Running, is the centered running variable which measures the distance from the cutoff, such that the cutoff equals zero, so in our case, baseline score - 36. This centering allows us to interpret α, our constant term, as the outcome for the control group at the cutoff. $\beta 1$ is the local average treatment effect, representing the discontinuity at the cutoff score of 36 points. The effect of the baseline score on the endline score is captured by $\beta 2$ for the control group and by $\beta 2 + \beta 3$ for the treatment group. Because we are using a local randomization assumption with a small bandwidth (+/- 4), meaning that we assume that assignment to treatment is as good as random around the cut off, and thus the model controls for both observable and unobservable characteristics. εi is the error term, where standard errors are clustered at the grade level.

3.3.3 The impact on learning outcomes

Results from the RDD analysis show a clear positive effect of the NPDA on student learning. Students who received the intervention (those just below the 36-point cutoff) scored, on average, 6.14 points higher on the endline assessment compared to similar students just above the cutoff. This amounts to a 17.1 per cent increase on the endline assessment. This demarcation is shown in Figure 9, where the red vertical line represents the cutoff for the NPDA. The grey vertical lines on either side represent the bandwidth for comparison groups selected for the analysis. This difference is statistically significant at the one-per cent level, providing strong evidence that the intervention improved learning outcomes for students near the eligibility threshold.

Figure 9: Effect of the NPDA on learning outcomes



Source: Authors' analysis of baseline assessments, the sample used for the above figure includes students who met the eligibility for the NPDA (N=185,697 students).

3.3.4 Impact results by gender, governorate and grade

Both boys and girls benefited from the NPDA, though girls experienced slightly greater gains in literacy skills. Boys who participated in the NPDA scored 4.95 points more than boys who did not participate in the programme, and girls scored 7.74 points more than non-participating counterparts, 2.18 points higher than boys. The difference among boys and girls receiving the NPDA was also statistically significant.[9]

Across the ten governorates, two per cent of students who scored 36 points or higher on the baseline were administered an endline assessment. These students make up the control group and are concentrated in four governorates (see Table 5). For each governorate, the effect of participating in the NPDA was estimated separately, using RDD. The effects of being in the NPDA were positive and significant across all four governorates, with the effects being largest in Cairo, followed by Assiut, Sohag, and Alexandria (*see Table 6*).

Table 5: Number of students in control and treatment groups, by governorate

GOVERNORATE	CONTROL	TREATMENT
Cairo	1,245	28,344
Sohag	1,106	46,423
Asyut	1,003	20,456
Alexandria	887	15,695
Giza	0	23,526
Ismailia	0	13,375
Aswan	0	12,299
Damietta	0	10,563
Fayoum	0	9,580
North Sinai	0	5,436
Total	**4,241**	**185,697**

Source: Authors' analysis of baseline and endline assessments, the sample used for the above table includes students who met the eligibility for the NPDA (N=185,697 students). And the control group sample N=4,241.

Table 6: RDD estimates of the effect of NPDA on endline assessment, by governorate

GOVER-NORATE	COEFFICIENT	AVG. GAIN, PER CENT	TOTAL OBSERVATIONS		SAMPLE USED IN RDD (IN BANDWIDTH)	
			TREATMENT	CONTROL	TREATMENT	CONTROL
Cairo	10.26***	24.1%	28,344	1,245	2,433	395
Alexandria	1.28*	2.4%	15,695	887	1,529	348
Sohag	5.77***	13.2%	46,423	1,106	3,238	386
Asyut	6.28***	14.5%	20,456	1,003	2,725	537

Note: *, **, *** denotes statistical significance levels of respectively p<0.10, p<0.05, p<.01. Average gain was calculated as the point increase in change for the treatment group, over the average baseline score for students in the control group in the same governorate, i.e. those scoring >35 and <41 points.

In addition to geographic location, an analysis by grade level shows that the effect of participating in NPDA was positive across all grades. Table 7 shows the change in scores was largest in Grade 6, followed by Grades 4, 3, and 5.

Table 7: RDD estimates of the effect of NPDA on endline assessment, by grade level

GRADE	COEFFICIENT	AVG. GAIN, PER CENT	TOTAL OBSERVATIONS		SAMPLE USED IN RDD (IN BANDWIDTH)	
			TREATMENT	CONTROL	TREATMENT	CONTROL
3	5.5***	12.9 %	54,679	1,037	4,811	347
4	6.91***	15.4 %	48,023	1,033	5,202	443
5	4.94***	10.9 %	43,452	1,136	3,759	353
6	7.50***	16.7 %	39,525	1,035	4,982	473

Note: *, **, *** denotes statistical significance levels of respectively p<0.10, p<0.05, p<.01. Average gain was calculated as the point increase in change for the treatment group, over the average baseline score for students in the control group in the same class, i.e. those scoring >35 and <41 points.

These results show that the NPDA had a strong positive impact on student learning outcomes. Girls seemed to benefit slightly more than boys, and while the impacts were positive across geographies, in the four governorates where change to control students' data was available there were differences in the size of the impact. While these results show the positive impact of the programme it is important to stress that this impact is estimated for students who scored close to the cut-off score of 36, therefore the results do not take into account the results of the large proportion of students who entered the programme who scored 0 or close to 0 at baseline. The following section provides additional analysis for the learning progression of students, with a particular focus on students with low baseline scores.

3.4 Relationship between baseline scores and endline outcomes

3.4.1 Outcomes for students scoring zero at baseline

As discussed in section 3.1.2 Baseline data, there was a large proportion of NPDA students who scored zero or close to zero at baseline. Out of the students who enrolled in the programme around one third (32.6 per cent (N=61,995) scored zero at baseline. This section explores how students who started at different levels progressed through the programme. There were substantial gains in scores by students who started with zero scores. The number of students who scored zero dropped dramatically by endline, from nearly 33 per cent of students at baseline to 1.7 per cent (N=3,264) (*see Annex 8*) at endline. Students who began with zero scores showed substantial gains, with an average endline score of 42.91 (gain of 42.91), compared to 41.65 (gain of 23.83) for students in the treatment group who had scored below 36 but not zero. This 19.08-point difference in improvements is statistically significant, suggesting that the remedial programme was particularly effective for students with the most severe learning gaps. While the average endline score of the zero-score students is higher, the difference is small, only 1.26 points. As such, the NPDA helped zero-score students catch-up with other students.

Students with a zero score at baseline demonstrated substantial gains across all literacy sub-domains between pre- and post-tests (Table 8). The largest gains were seen in students' scores in sound awareness (74.47 percentage points) and reading letters and words (74.11 percentage points), followed closely by reading comprehension (71.80). Gains were also notable in spelling (66.27) and fluency (67.60), though these represented the smallest relative increases among the sub-domains.

Table 8: Mean post-test scores for students who scored zero at baseline, by test sub-domain

SUB-DOMAIN	MEAN	STANDARD DEVIATION	PERCENTAGE POINT INCREASE
Reading letters and words	13.34	5.59	74.1
Sound awareness	12.66	5.3	74.5
Fluency	3.38	1.95	67.6
Reading comprehension	3.59	1.98	71.8
Spelling	9.94	5.51	66.3
Total	**42.91**	**18.94**	**71.5**

Source: Authors' analysis of baseline and endline data (N=61,995).

3.4.2 Ordinary least square results on the relationship between baseline and endline scores

An ordinary least squares (OLS) regression analysis was conducted to examine the relationship between student's baseline and endline scores (see Annex 9 for full results). The findings indicate a positive and significant association between baseline and endline scores, with girls scoring 3.6 points higher than boys on average. Higher grade levels were also linked to higher endline scores. A subsample analysis revealed that students who scored zero at baseline had the largest learning gains, consistent with a negative relationship between baseline scores and gains (*see Specification (5) of Annex 9, Table 10*). Additionally, girls and higher-grade students were more likely to score zero at baseline, suggesting they needed more support upon entering the programme. These results imply that the RDD estimates may represent a conservative lower bound of the impact of NPDA, as RDD does not fully account for the gains among the lowest-performing students. Further analysis will explore this in Phase II.

3.5 Understanding implementation

3.5.1 School monitoring visits

To provide context for the results on learning, and to understand the implementation of the NPDA, structured monitoring visits were conducted at each school by monitors from the NCEEE. Throughout the implementation of the NPDA, 783 schools (80 per cent of participating schools) received at least one monitoring visit. Observations mostly took place between mid-March through May during programme implementation. A total of 487 monitors made 3,892 visits (*Figure 10*). While the number of school visits per monitor ranged from 1 to 20, on average, there were 3.9 visits per monitor (*Figure 11*).

Figure 10: Distribution of monitor visits per week (n=3,892)

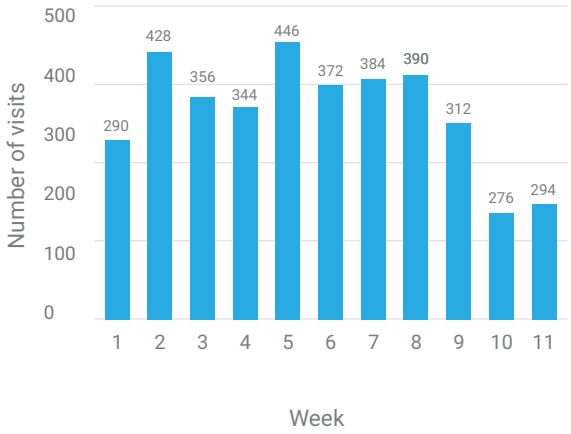
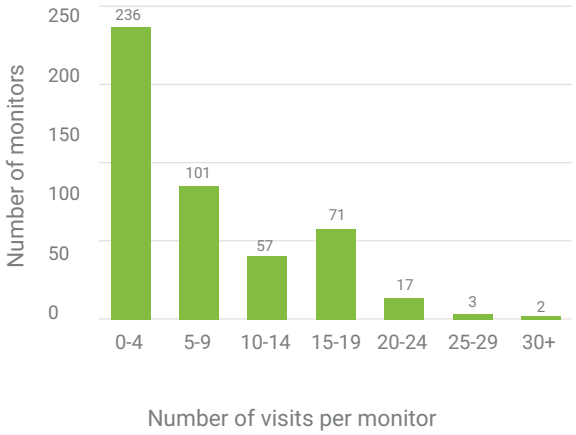


Figure 11: Distribution of number of visits by monitors (n=3,892)



Source: Authors' analysis of monitoring data.

3.5.2 Classroom and school characteristics

On average, schools participating in the NPDA had a total of 446 students across Grades 3 to 6 with the largest school having 2,380 students across those grades. For students in the target grades, on average 203 students per school or 45 per cent of students were eligible to take part in the programme. On average, there were approximately 216 programme participants at each school per visit, across 4.9 classes, with 4.7 teachers participating (Table 9). There is large variation in the number of students participating per observation, reflected in the high standard deviation. There is also large variation in the number of students which were enrolled in the programme across governorates.

Table 9: Average number of students participating in schools in the NPDA, by governorate

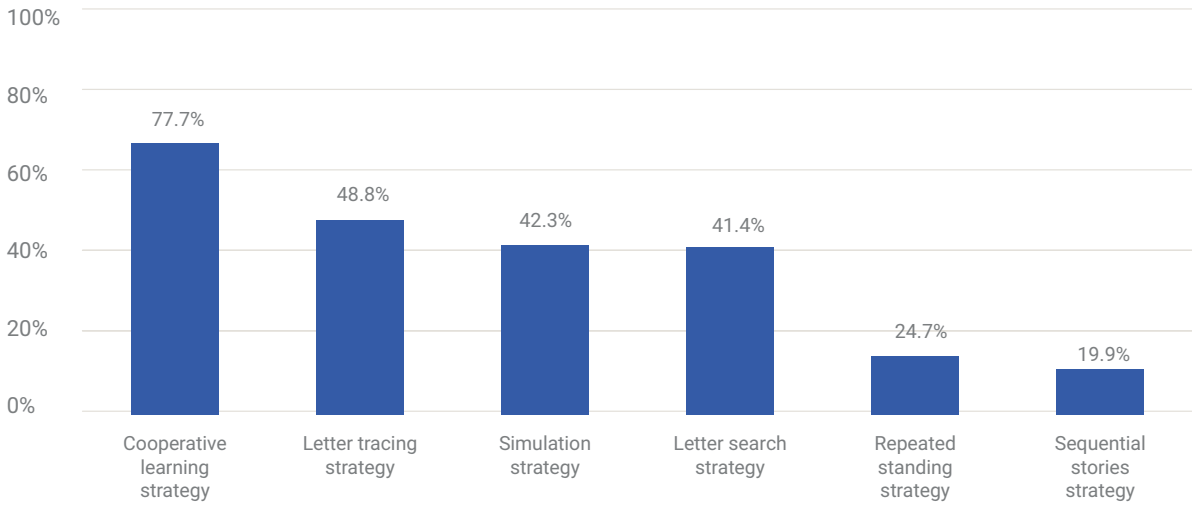
GOVERNORATE	STUDENTS/ SCHOOL, MEAN	SD	MIN	MAX	N
Alexandria	210	127.20	10	479	79
Aswan	132	111.70	2	796	93
Asyut	215	135.40	11	662	100
Cairo	213	152.00	5	677	139
Damietta	134	101.10	6	412	79
Faiyum	192	151.00	19	678	50
Giza	180	127.50	10	571	131
Ismailia	134	108.50	6	690	100
North Sinai	129	63.60	36	341	42
Sohag	295	158.10	16	952	161
Total	**195**	**143.40**	**2**	**952**	**974**

Source: Authors' analysis of baseline assessments, the sample used for the above table includes students who met the eligibility for the NPDA (N=185,697 students).

3.5.3 Classroom observation of teaching practices

During classroom observations, monitors checked a variety of aspects of programme implementation. This included if various learning tools were deployed, such as using different types of games (e.g., crosswords) and what activities teachers were employing (e.g., writing, coloring, organizing) as well as pedagogical techniques used by teachers. Teachers used a variety of pedagogical techniques in the classroom, as shown by Figure 12, the use of cooperative learning strategies was observed in 78 per cent of the school visits, the use of letter tracing was seen observed in just under half of classes, while the use of sequential stories strategy was only observed in 20 per cent of the visits.

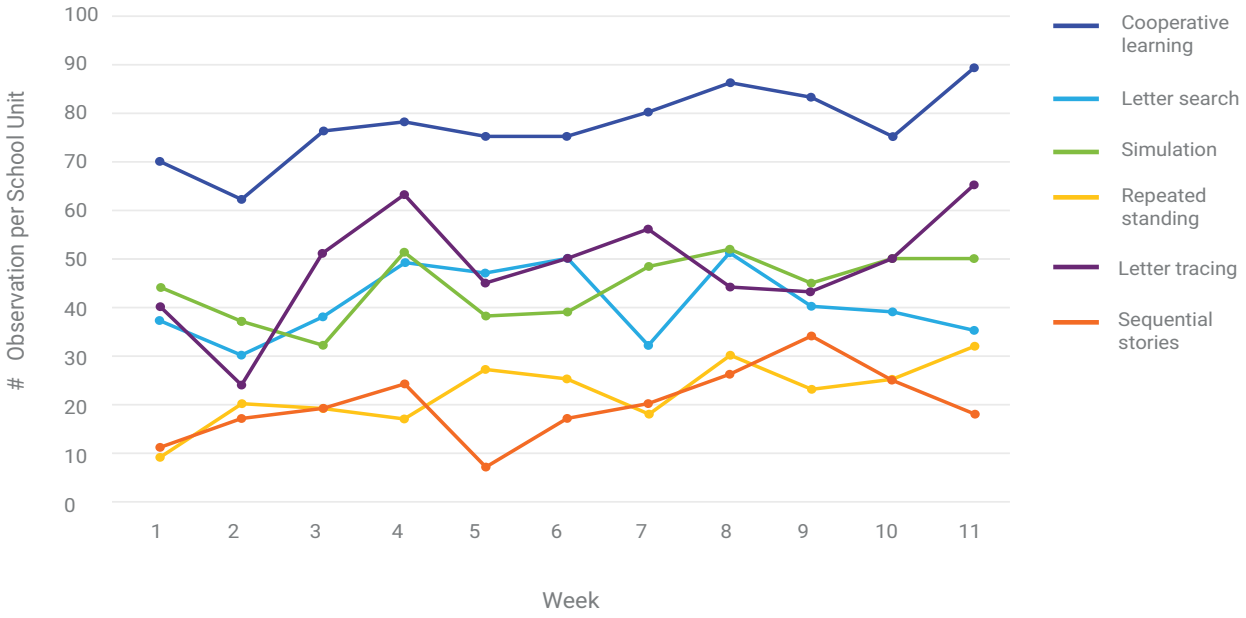
Figure 12: Percentage of visits where different strategies were observed



Source: Authors' analysis of monitoring data, N=3,892.

Overall, most strategies show a gradual increase or stabilization in usage over the period of the programme. The graphs in Figure 13 reveal trends in the use of teaching strategies over the 11-week period when observations took place. Notably, strategies like Cooperative Learning and Repeated Standing show steady or rising use throughout, with peaks towards the end, indicating sustained engagement. More complex strategies such as Simulation and Sequential Stories start with lower use but increase significantly in later weeks, indicating teachers growing confidence utilizing those techniques.

Figure 13: Strategy used in class by week

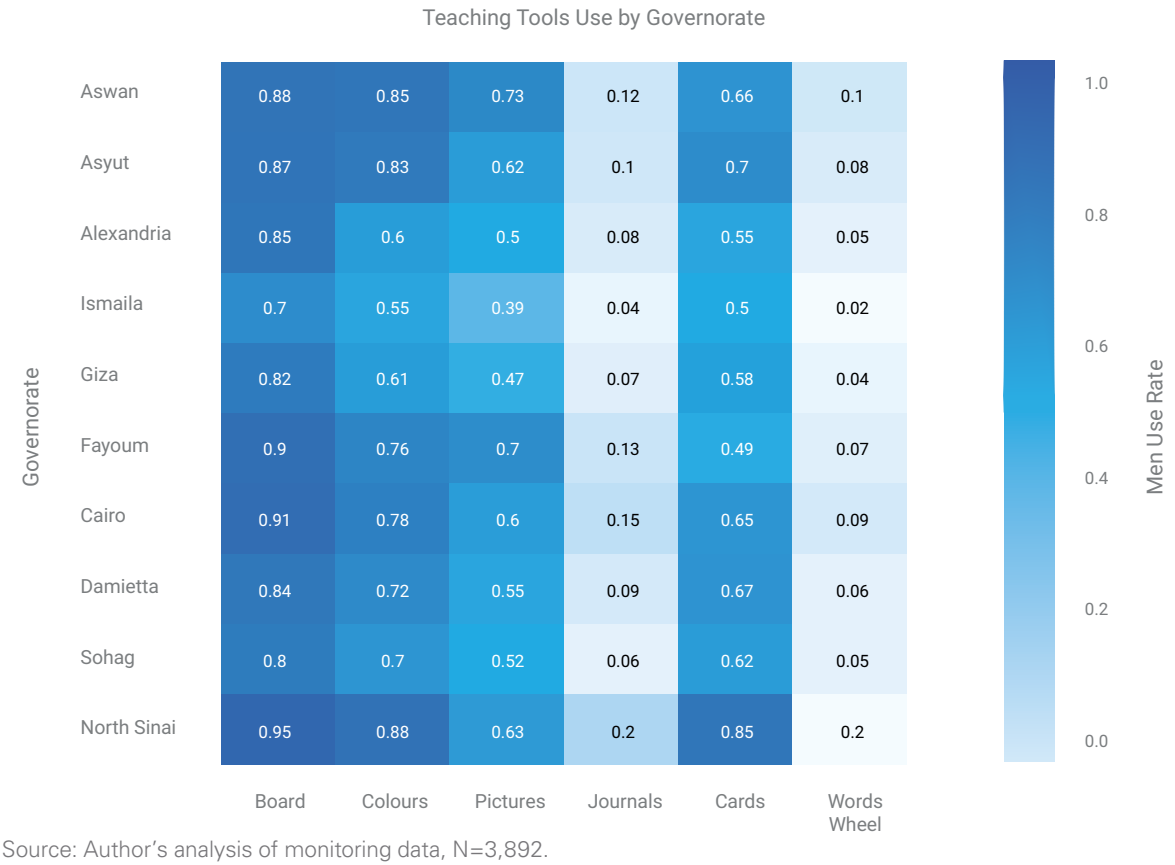


Source: Author's analysis of monitoring data, N=3,892. The y-axis represents the number of times that the strategy was observed over the total number of observations of that strategy.

3.5.4 Classroom observation of teaching materials and tools across governorates

Monitors observed the teaching tools and materials that were available in the classroom, and were used for teaching, during the visit. These tools included boards, colors, pictures, journals, cards, and the words wheel (*Figure 14*). The data shows that boards and colors were the most widely and consistently used materials across all governorates. On the other hand, tools like journals and the words wheel were observed far less frequently in most geographic locations. While governorates such as North Sinai demonstrate relatively high usage of nearly all tools, others – such as Ismailia, Alexandria, and Giza – showed much lower level of tool usage. These findings suggest a need for targeted support to ensure all classrooms are equally equipped with effective teaching tools.

Figure 14: Teaching tools used by governorate



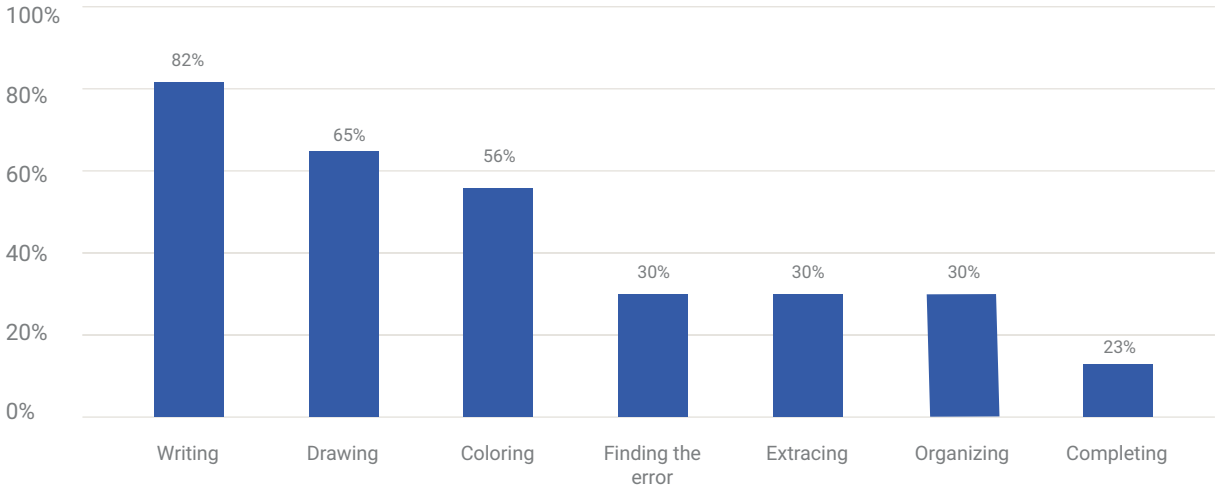
Source: Author's analysis of monitoring data, N=3,892.

3.5.5 Classroom observation of teaching activities

Class observations included questions on the types of learning activities that were used in classrooms during visits. These activities included writing, colouring, drawing, finding errors, organizing, extracting information, and completing sentences. The data reveals that writing and drawing were the most frequently observed activities, with consistently high usage across all weeks, suggesting that these tasks are well-integrated into daily teaching instructions and likely perceived as essential by teachers (*Figure 15*).

Colouring followed a similar trend, showing moderate use with an increase toward the end of the 11-week period. In contrast, more cognitively demanding activities such as "find the error", organize, extract, and especially "complete the sentence" were used less frequently, although some showed gradual increases over time. It is important to note that depending on the week and day of the week when the observation took place some of these practices could be more or less relevant for the specific lesson objective. Qualitative data collection will explore further the experience of teachers and differences in their use of activities and pedagogical techniques.

Figure 15: Observed activities in NPDA classrooms



Source: Authors' analysis of observation data, N=3,892.

3.5.6 Stakeholder feedback

To understand the perceptions of the NPDA across several stakeholder groups, surveys were conducted with i) parents of learners; ii) the quality assurance team at the NCEEE; iii) school principals; and iv) teachers.

The analysis grouped related variables into key domains to simplify interpretation across various respondents. This includes:

1. **Implementation** - reflecting how consistently and effectively the programme was delivered.
2. **Training** - capturing the quality and usefulness of teacher preparation and materials.
3. **Support** - including the support of leadership, coordination, and technical guidance.
4. **Engagement** - relating to the motivation and participation of either students or teachers.
5. **Outcomes** - measures the perceived improvements in student literacy and overall satisfaction with the programme.

For parents, a distinct communication domain assesses the clarity of information shared by schools. In the teachers' survey, a feedback domain addresses supervision, incentives, and performance input. Collectively, these domains provide a structured view through which to explore perceptions around implementation quality and programme impact.

Survey results reveal high overall satisfaction from stakeholders across all these domains. Across all groups, satisfaction with the NPDA was consistently high (*Figure 16*). Overall, the support and implementation domains received consistently high ratings from all groups of respondents, with scores clustering around 3.5 out of 4 (*See Annex 10* for details on perceptions towards sub-domains and individual items). This suggests that delivery quality and structural implementation were seen as effective. Parents gave the highest score overall to the "outcomes" domain (3.6 out of 4), indicating

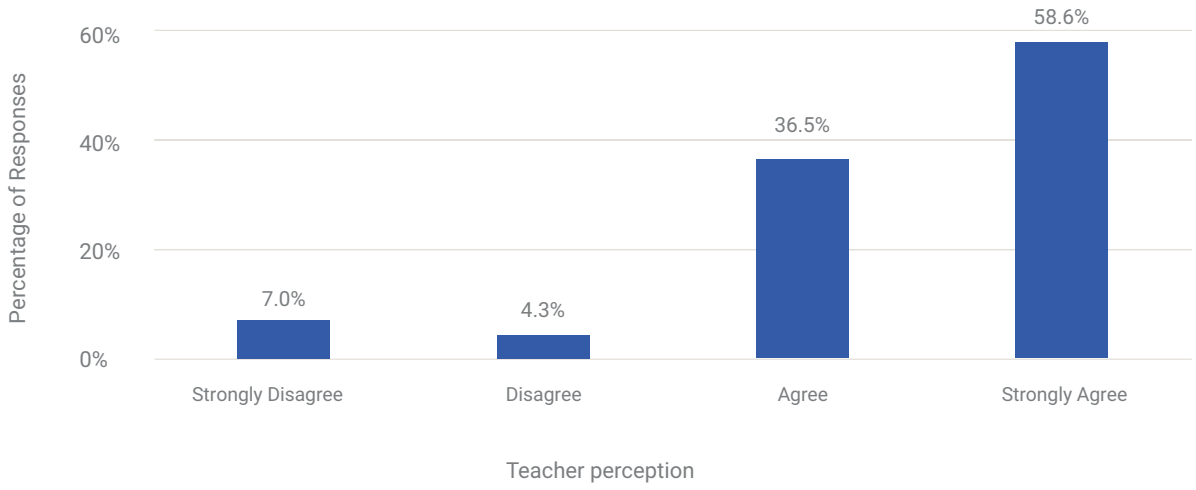
strong satisfaction with the perceived impact of the programme on their children's literacy. In contrast, while still giving a high score, teachers rated the outcomes domain the lowest (3.32 out of 4). Teachers also gave the lowest score overall to the "feedback" domain (3.11 out of 4), highlighting a clear area for improvement in supervision and recognition within the programme. Among principals, the lowest rating was given for training (3.29 out of 4), reflecting potential gaps in teacher preparation. Meanwhile, the quality assurance team gave the lowest scores to engagement (3.37 out of 4). Delving deeper into these domains, it is observed that most respondents "agreed" or "strongly agreed" with positive statements about the programme, especially around implementation clarity, teacher commitment, and student progress. For example:

- Over 95 per cent of parents agreed or strongly agreed that the programme was clearly communicated and consistently delivered to their children.

- More than 97 per cent of quality assurance monitors and principals reported that the goals of the literacy programme were clearly defined and communicated.

- Over 90 per cent of teachers reported positive views on their ability to deliver the programme and the effectiveness of programme-related supporting materials (such as teaching guides).

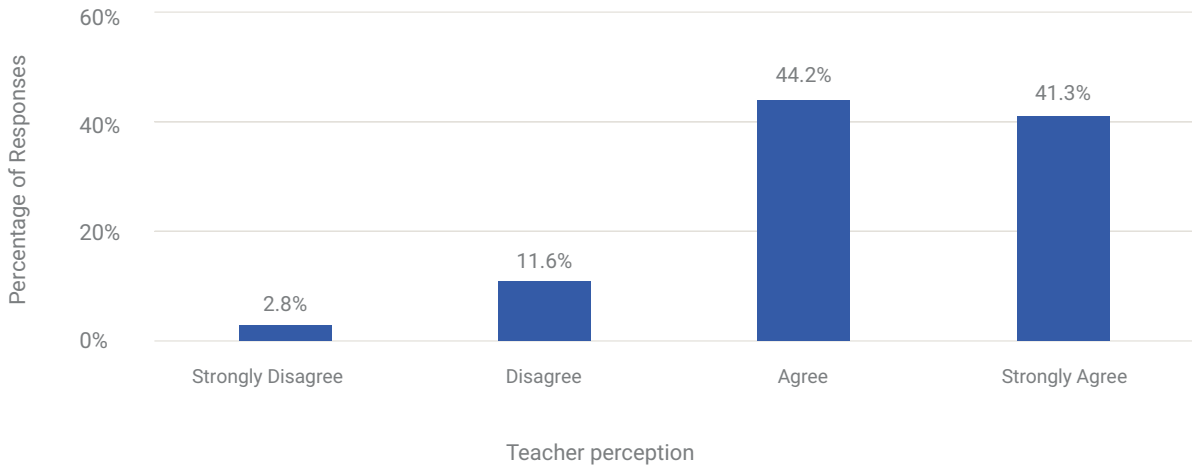
Figure 16: Teacher perceptions on delivering the programme according to the schedule



Source: Authors' analysis of surveys, N=913 teachers.

Participants largely agreed that the programme fostered student engagement and learning improvements, with above 90 per cent agreeing or strongly agreeing that children were motivated and showed progress in literacy skills. Nearly all respondents across teachers (97 per cent), principals (97 per cent), and quality assurance monitors (99.5 per cent) agreed or strongly agreed that the programme goals were clearly communicated. 93.7 per cent of principals reported receiving adequate technical and supervisory support. Despite positive overall perceptions regarding different programme domains, some areas of variation were identified in the data, including around: 14.4 per cent of teachers disagreed or strongly disagreed that the teacher training was adequate for them to deliver the programme (*Figure 17)*.

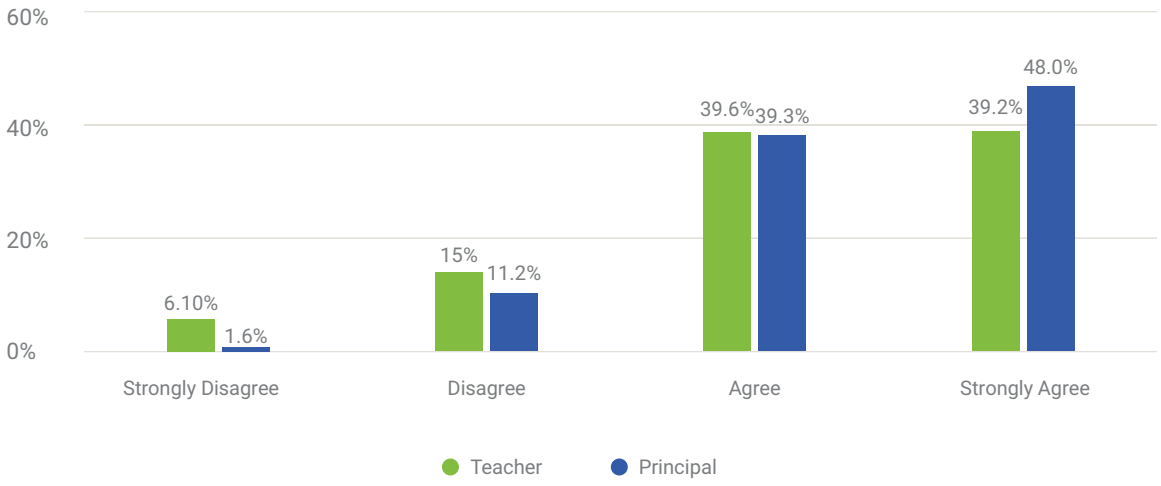
Figure 17: Teacher perceptions on teacher training adequacy



Source: Authors' analysis of surveys, N=913 teachers.

On teaching material availability, 21 per cent of teachers, 14 per cent of the quality assurance monitors, and nearly 13 per cent of principals rated it as weak or insufficient. Showing that concerns were shared across both groups (*Figure 18*).

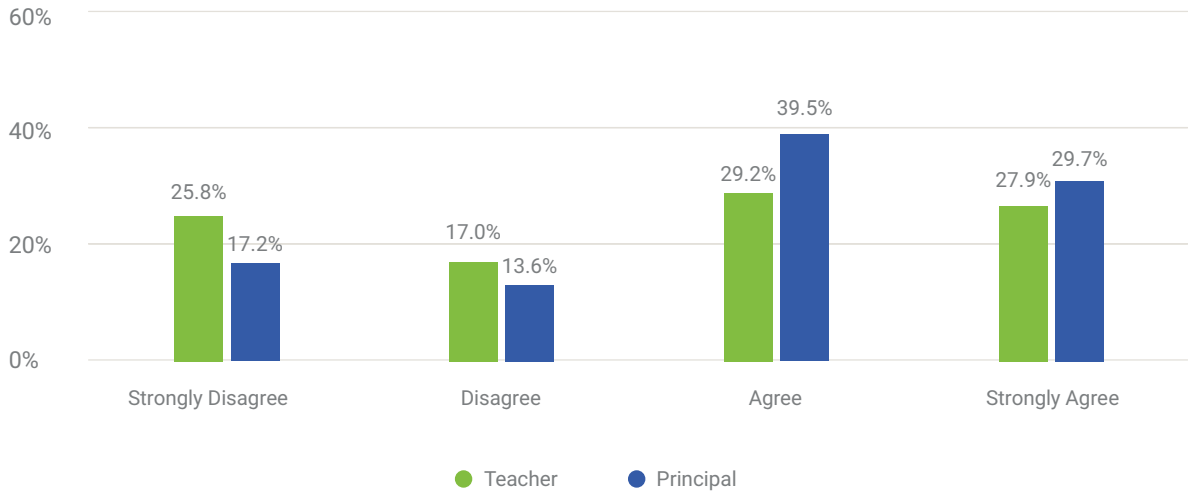
Figure 18: Teacher and principal perceptions around the availability of useful teaching materials



Source: Authors' analysis of surveys, N=913 teachers and N=448 for principals.

It should be noted that the NPDA represents a substantial increase in workload for teachers and administrators. As such teachers and administrators were provided financial incentives for their participation, 58 per cent of teachers agreed or strongly agreed that they were satisfied with the incentives that they are receiving, while just over a quarter of teachers (25.8 per cent) reported that financial incentives were weak (*Figure 19*). Among principals, over 17 per cent reported that financial incentives for the programme were weak.

Figure 19: Teacher and principal perceptions around financial incentives



Source: Authors' analysis of surveys, N=913 teachers and N=448 for principals.

IV. Conclusion

NPDA was designed as a targeted strategy to support students who were falling behind in foundational literacy skills. More than half of children (56 per cent) are estimated to not be able read or write a simple story by the age of 10. When a student cannot read or write a simple sentence, it means they cannot make the crucial shift from learning to read to reading to learn. Practically, this means that these students are limited in how much they can engage in learning, and in achieving the skills they need for their futures to participate in academics, society and the economy.

To tackle this problem, NPDA provided intensive literacy instruction to 199,225 struggling students. The programme took place for one hour, five days a week over the 12 weeks for a total of 60 hours of learning. The programme was comprehensive, it comprised of teacher training for teachers, structured lesson plans for teachers, teaching and learning materials. All these activities were carried out in a highly coordinated manner led by the NCEEE whose monitors provided support and carried out classroom observations throughout the 12 weeks of implementation.

This report provides the main findings from this experience as well as some recommendations for further development and exploration.

Key findings

1. **Impact on literacy outcomes:** NPDA had a significant positive impact on students' literacy. Using a Regression Discontinuity Design (RDD) technique, the causal impact of the NPDA is estimated at 6 points, which is equivalent to a 17 percentage point increase in learning. This is likely a lower-bound estimate of the true effect of the programme. Additional analysis shows that students with lower baseline scores experienced the greatest gains.

2. **Progression in learning for NPDA participants:** Students participating in NPDA showed large improvements between baseline and endline, with average scores increasing by 50 percentage points. Notably:

    a. The proportion of students scoring zero dropped from 33 per cent at baseline to 1.7 per cent at endline.

    b. Improvements were seen across all five literacy domains with the highest increases seen in Reading Comprehension and Fluency, followed by Spelling, Reading Letters and Words and Sound Awareness.

3. **Variations by gender, geography and grade**

    a. Girls and older students outperformed their peers at baseline, making them less likely to be eligible for the programme. As a result, only 40 per cent of NPDA participants were girls.

    b. However, among girls and older students who did qualify, a higher proportion had zero baseline scores—indicating greater vulnerability among those left behind. Results from the RDD also find that the programme had a larger impact on girls than on boys.

    c. Regional variation was also significant in both baseline characteristics and programme outcomes. Implementation fidelity varied slightly across regions, based on monitor observations, more investigation is needed into what accounts for some of these differences.

4. **Classroom observations and the use of tools and pedagogical techniques**

    a. Classroom observations overall revealed strong adherence to programme guidelines. They also show a general trend that teachers utilized more of the pedagogical techniques and recommended tools as the weeks progressed, indicating that they became more comfortable with the tools. There were variations between regions in the uptake of tools which warrant further investigation as the programme continues its development.

5. **Stakeholder feedback:** Overall feedback for the programme from teachers, principals, NCEEE staff, and parents was overwhelmingly positive. Several areas, however, emerged from feedback as potential areas for improvement in the next stage.

    a. 21 per cent of teachers reported that materials were not always available.

    b. 15 per cent of teachers felt the training was not sufficient.

    c. 43 per cent of teachers indicated that the incentives for participation in the programme were not sufficient.

Recommendations

1. **Institutionalize the Arabic literacy remediation programme.** NPDA Phase I has confirmed the need and demonstrated the potential impact of remediation interventions, especially those embedded within school hours. Acceleration and institutionalization of such interventions are critical to support Egypt reach its commitment of halving learning poverty (from 56 per cent) to 27 per cent by 2029.

2. **Track the sustainability of learning gains.** This research demonstrates the strong immediate impact of the programme on student literacy. It will be essential to understand whether these gains persist over time. Follow-up assessments should be considered to understand the longer-term learning trajectories of students who participated in NPDA.

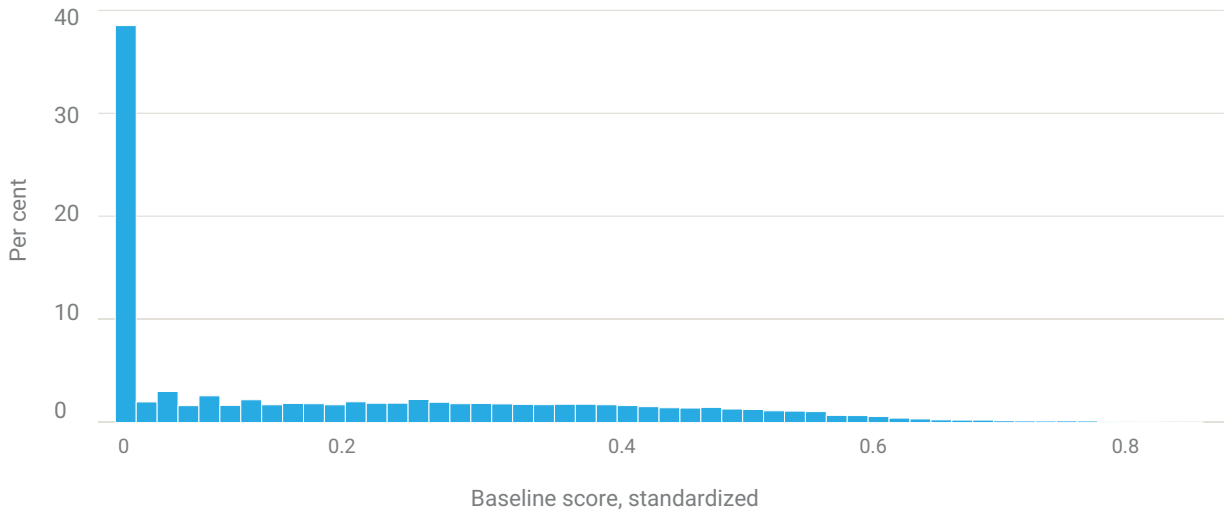
3. **Identify ways to improve teacher support.** Despite positive feedback, issues related to availability of materials, incentives for the programme and teacher training were clear in the feedback from teachers and administrators. Observations also found that materials were not evenly used across regions. The programme should work with teachers to identify where there could be gaps to improve and further refine the programme package.

4. **Study Variation in Implementation and Cost-Effectiveness.** As the programme progresses different models of support could be evaluated to understand what types of support are most cost effective. Understanding first the differences in context between regions which led to differences in implementation would be critical, potentially to be evaluated in Phase II.

5. **Explore the effects of the programme on regular teaching practices.** The NPDA was effective, it improved learning for students and was implemented by teachers in schools with variation during and outside of regular teaching hours. To understand the sustainability of the programme and its implementation, it will be important to investigate whether experience of the programme has influenced the practice in teachers' regular practice.

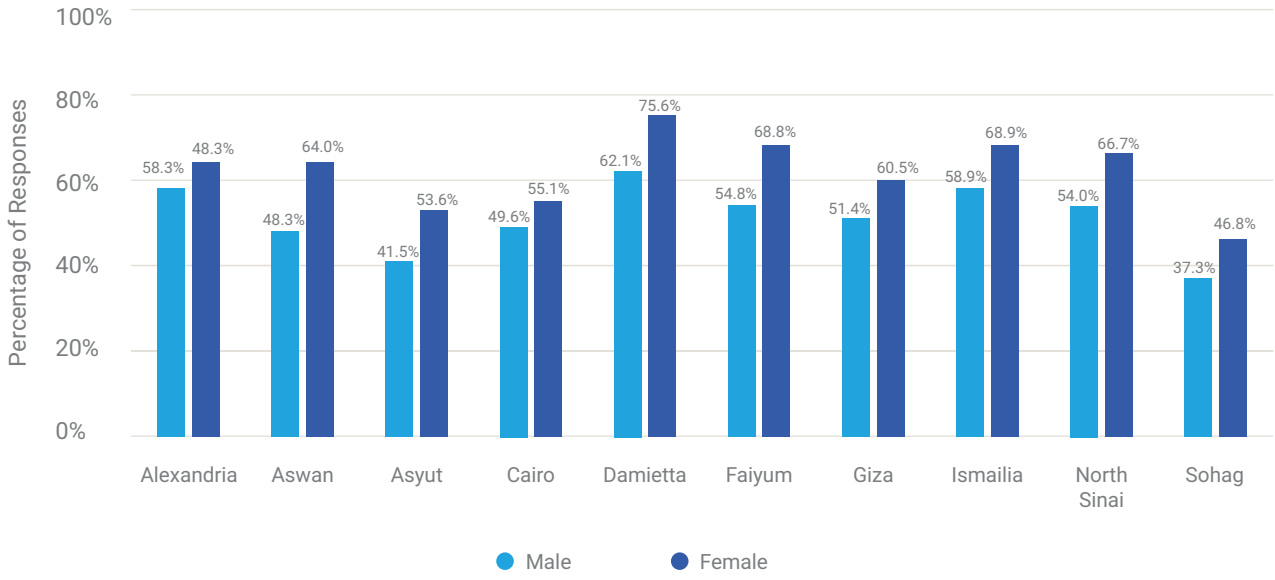
Annexes

Annex 1: Distribution of standardized baseline scores



Source: Authors' analysis of baseline assessments, the sample used for the above figure includes all students who completed the baseline literacy assessment (N=438,875 students).

Annex 2: Percentage of students scoring 36 points or more on the baseline assessment, by governorate and gender

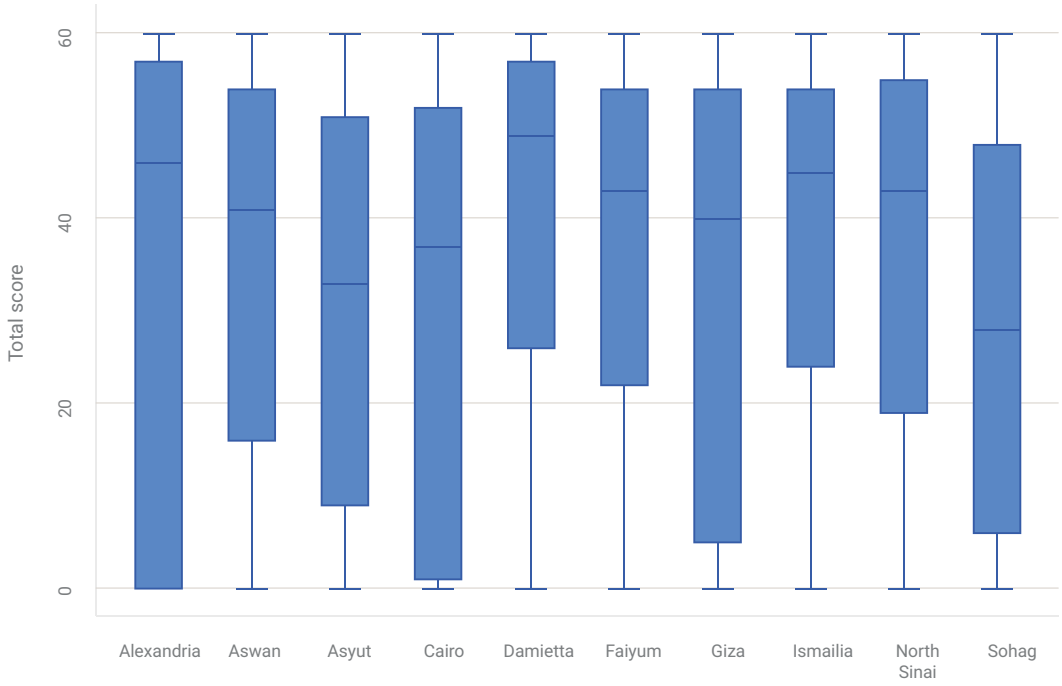


Source: Authors' analysis of baseline assessments, the sample used for the above figure includes all students who completed the baseline literacy assessment (N=438,875 students).

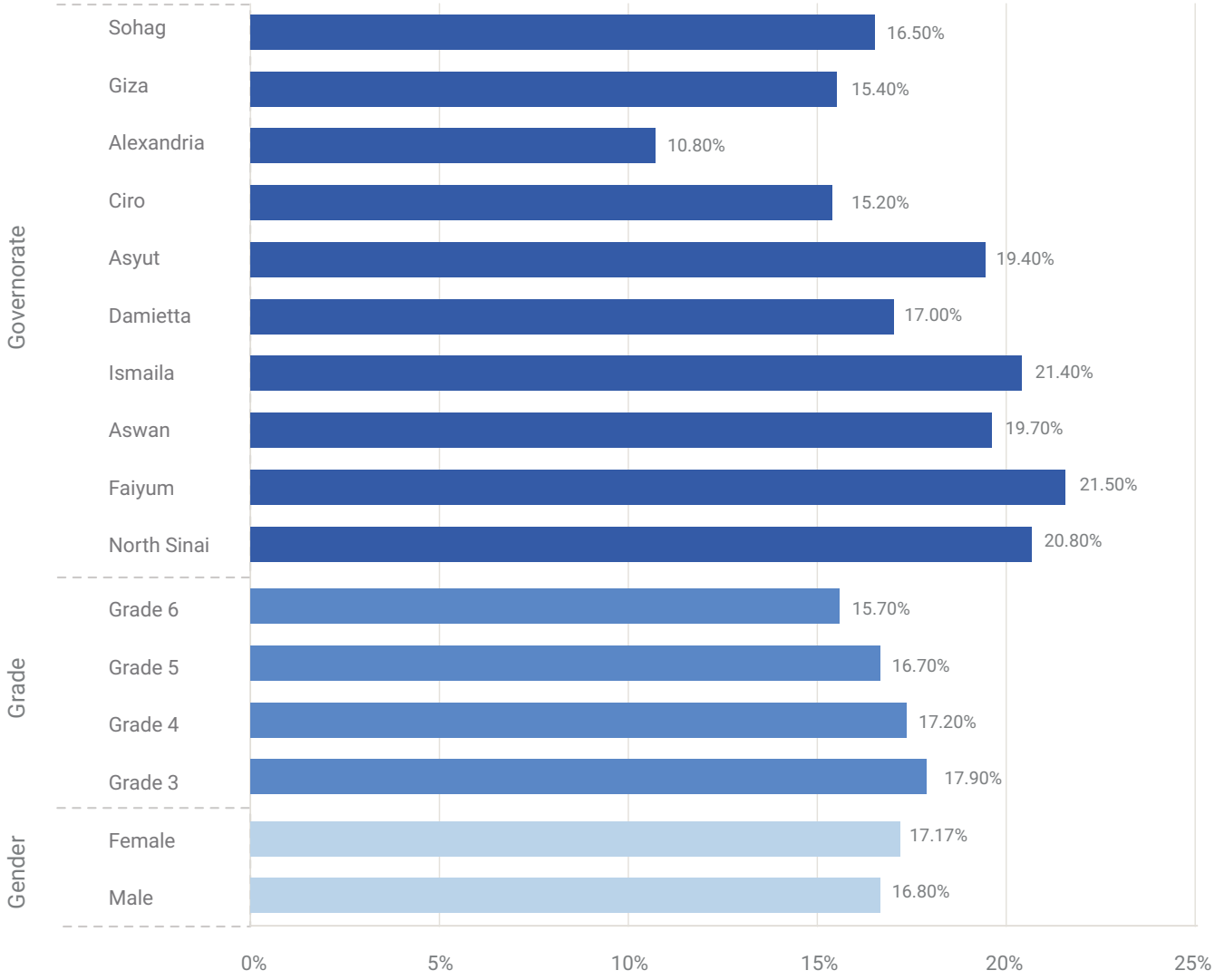
Annex 3: Average baseline scores by grade level and governorate

Source: Authors' analysis of baseline assessments, the sample used for the above figure includes all students who completed the baseline literacy assessment (N=438,875 students).

Annex 4: Distribution of raw baseline scores by governorate

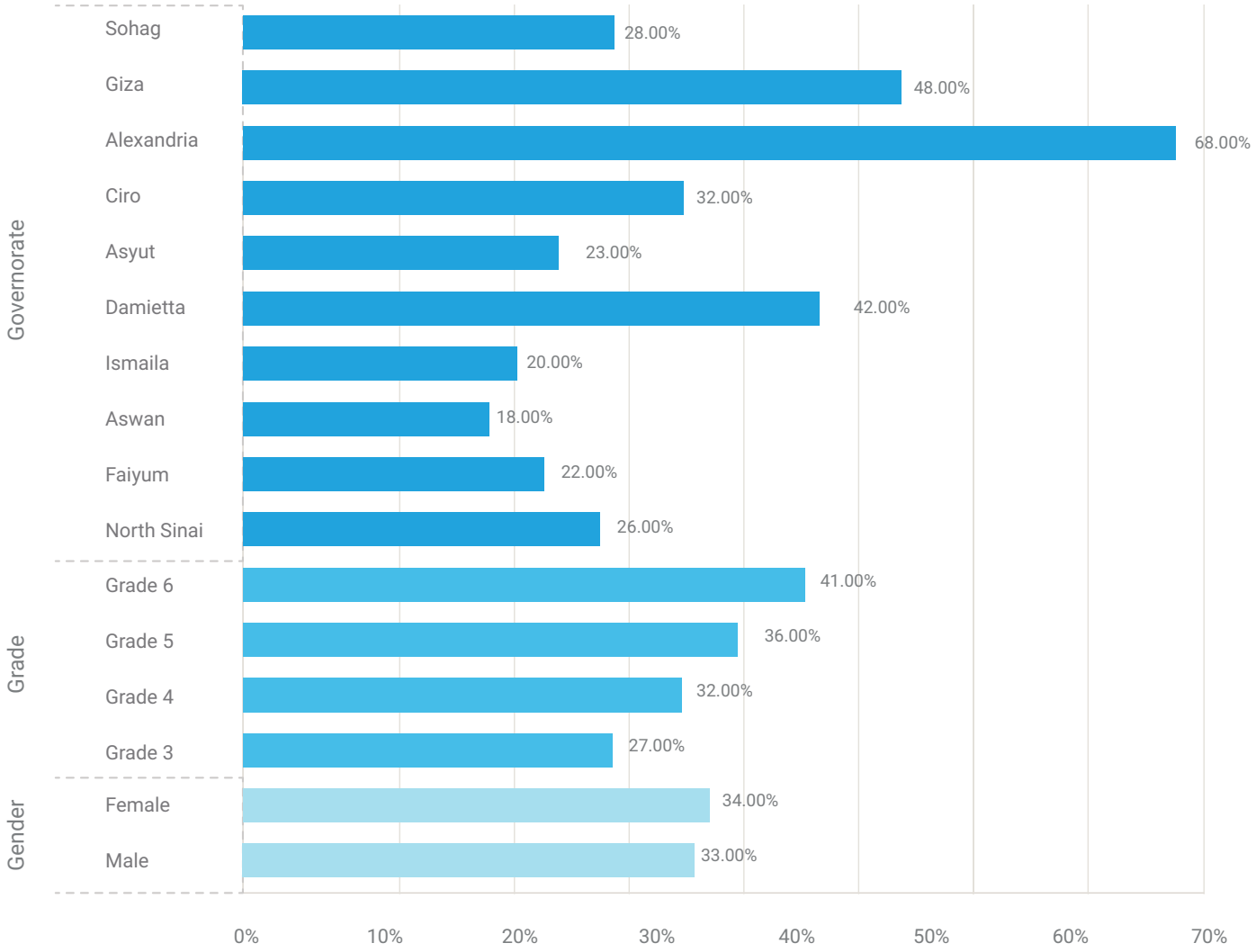


Annex 5: Average baseline score (percentage of correct responses), per governorate, grade, and gender



Source: Authors' analysis of baseline assessments, the sample used for the above figure includes students who met the eligibility for the NPDA (N=185,679 students). In total, there were 75,730 female students (40.8 per cent) and 109,967 male students (59.2 per cent). A total of 61,996 students scored zero on the baseline assessment (33.4 per cent).

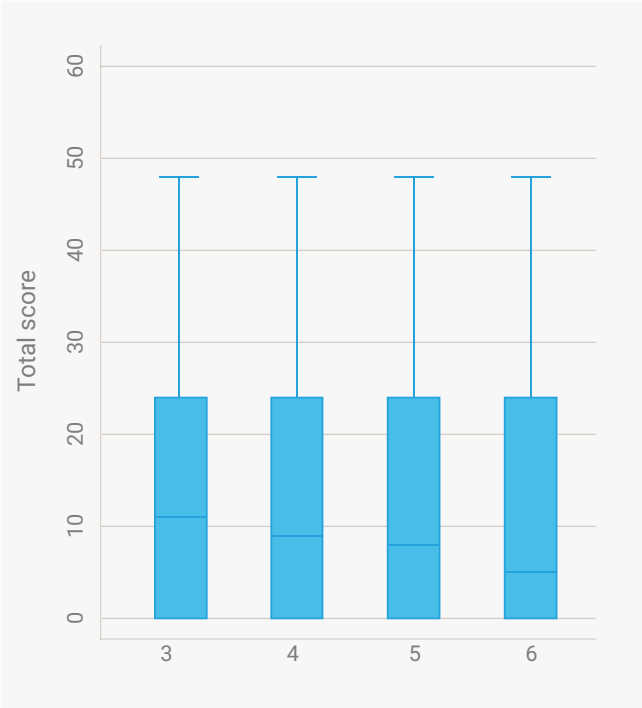
Annex 6: Per cent of students who scored zero at baseline by governorate, grade and gender



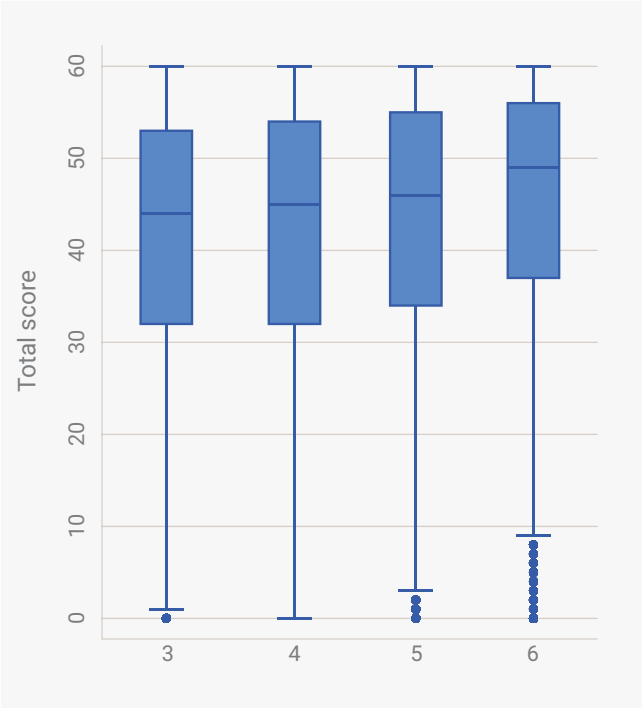
Source: Authors' analysis of baseline assessments, the sample used for the above figure includes students who met the eligibility for the NPDA (N=185,679 students). In total, there were 75,730 female students (40.8 per cent) and 109,967 male students (59.2 per cent). A total of 61,996 students scored zero on the baseline assessment (33.4 per cent).

Annex 7: Distribution of baseline and endline scores by grade level

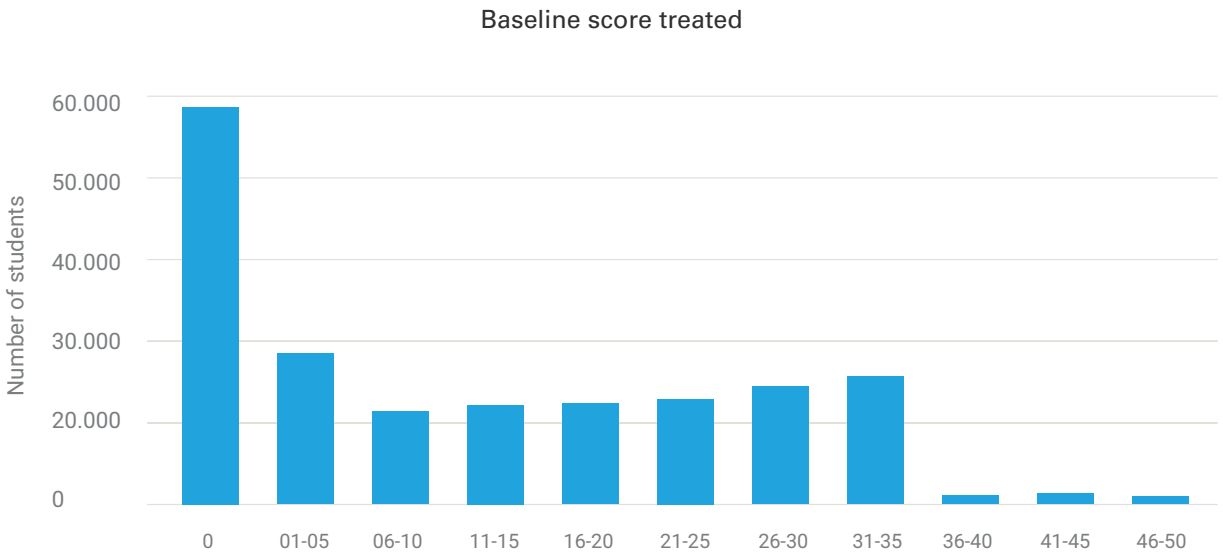
Baseline scores

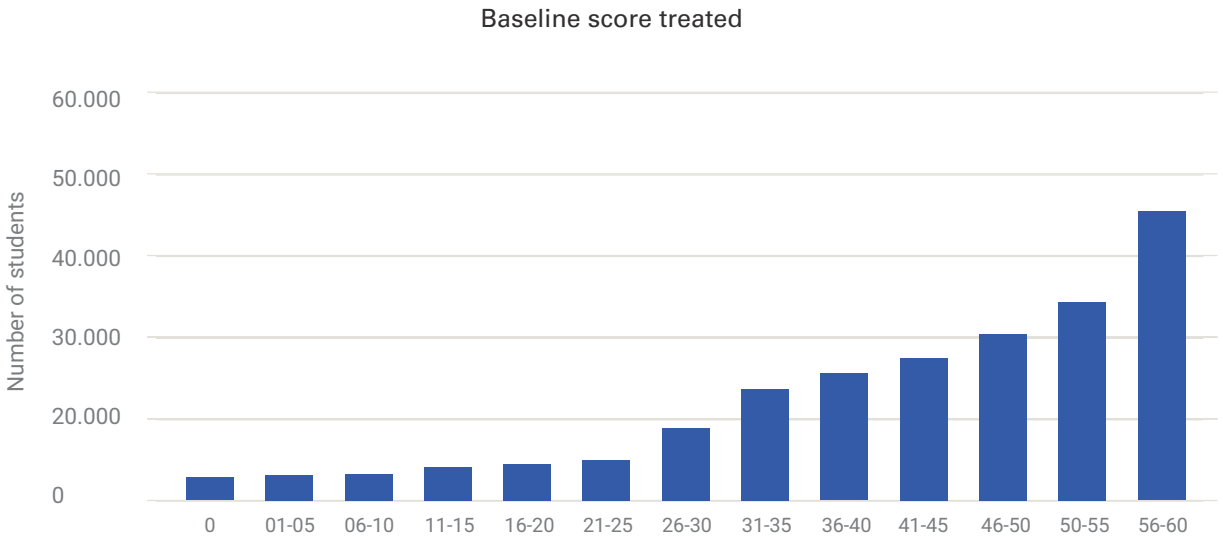


Endline scores



Annex 8: Percentage of students who received the intervention achieving different scores at baseline and endline





Source: Authors' analysis of baseline and endline data.

Annex 9: OLS estimation of baseline and endline scores

To understand the relationship between a student's baseline scores and endline scores, we estimated the association using ordinary least squares (OLS) regression. Table 10 below presents the results of the analysis. Specification (1) models the endline scores as a function of baseline scores, gender, grade level, and governorate. It shows that there is a positive and significant relationship between baseline score and endline score. Girls also score 3.16 points more than boys at endline. In terms of grade levels, there is an association between higher grade levels and higher endline scores. Specification (2) is the same as (1) but includes zero scores as a control variable, with results that are similar in direction, magnitude, and significance as specification (1). Specificatsion (3) and (4) are modeled the same way as our main specification but limits the data for a subsample analysis, with the samples restricted to students scoring a zero at baseline, and students scoring greater than zero at baseline, respectively.

Specification (5) models learning gains as a function of baseline scores and the same control variables. Here, there is a negative relationship, i.e., lower baseline scores are associated with higher gains, and higher baseline scores are associated with lower gains. This is consistent with an analysis conducted above that children who score zero have the largest gains. Specification (6) estimates the relationship between the likelihood of getting a zero score and gender, grade level, and governorate. Findings suggest also that there is a significant association between the likelihood of scoring zero and being female, as well as being in a higher grade level. Meaning that while less girls and students in higher grades were eligible for the programme those that did enter the programme were more likely to have the highest need.

Table 10: Regression analysis estimating the relationship between baseline and endline scores

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS - DV: POST	OLS - DV: POST	OLS - DV: POST, BASELINE=0	OLS - DV: POST, BASELINE>0	OLS - DV: GAINS	LOGIT - DV: ZERO SCORES
main						
Baseline score	0.263***	0.433***	0	0.455***	-0.737***	
	(0.00276)	(0.00373)	(.)	(0.00313)	(0.00276)	
Female=1	3.160***	2.924***	4.229***	2.380***	3.160***	0.150***
	(0.0676)	(0.0669)	(0.144)	(0.0683)	(0.0676)	(0.0106)
Grade level						
Ref= 3rd Grade						
4th Grade	0.663***	0.417***	0.481*	0.345***	0.663***	0.259***
	(0.0896)	(0.0886)	(0.203)	(0.0882)	(0.0896)	(0.0144)
5th Grade	2.174***	1.786***	2.463***	1.367***	2.174***	0.435***
	(0.0921)	(0.0912)	(0.202)	(0.0924)	(0.0921)	(0.0146)
6th Grade	3.795***	3.188***	3.656***	2.769***	3.795***	0.657***
	(0.0946)	(0.0940)	(0.200)	(0.0975)	(0.0946)	(0.0148)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS - DV: POST	OLS - DV: POST	OLS - DV: POST, BASELINE=0	OLS - DV: POST, BASELINE>0	OLS - DV: GAINS	LOGIT - DV: ZERO SCORES
Governorate						
Ref= Cairo						
Alexandria	5.653***	3.804***	7.786***	-2.206***	5.653***	1.530***
	(0.142)	(0.143)	(0.250)	(0.189)	(0.142)	(0.0215)
Ismailia	-4.043***	-4.053***	-9.392***	-3.147***	-4.043***	-0.618***
	(0.150)	(0.148)	(0.383)	(0.142)	(0.150)	(0.0250)
Damietta	-0.322*	-1.172***	1.206***	-2.652***	-0.322*	0.396***
	(0.163)	(0.161)	(0.322)	(0.173)	(0.163)	(0.0236)
North Sinai	-3.086***	-3.482***	-0.547	-4.851***	-3.086***	-0.284***
	(0.212)	(0.209)	(0.498)	(0.205)	(0.212)	(0.0335)
Giza	0.248*	-0.835***	3.611***	-4.307***	0.248*	0.672***
	(0.126)	(0.125)	(0.247)	(0.137)	(0.126)	(0.0184)
Sohag	-10.15***	-10.10***	-11.29***	-9.803***	-10.15***	-0.196***
	(0.108)	(0.106)	(0.239)	(0.107)	(0.108)	(0.0164)
Assiut	-10.06***	-9.896***	-7.910***	-10.81***	-10.06***	-0.516***
	(0.131)	(0.130)	(0.317)	(0.126)	(0.131)	(0.0211)
Fayoum	-1.881***	-2.050***	0.575	-3.302***	-1.881***	-0.464***
	(0.170)	(0.168)	(0.427)	(0.162)	(0.170)	(0.0281)
Aswan	-5.360***	-5.015***	-3.251***	-5.855***	-5.360***	-0.771***
	(0.154)	(0.153)	(0.412)	(0.145)	(0.154)	(0.0267)
Baseline=zero		6.676***				
		(0.1000)				
_cons	40.16***	36.60***	40.87***	37.43***	40.16***	-1.118***
	(0.108)	(0.120)	(0.232)	(0.112)	(0.108)	(0.0163)
N	185697	185697	61995	123702	185697	185705
adj. R^2	0.156	0.175	0.141	0.240	0.374	

Annex 10: Frequency distributions of responses on surveys measuring participant perceptions on the NPDA, by respondent and domain

IMPLEMENTATION & SUPPORT DOMAIN				
SURVEY: PARENTS (N=304)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
The school clearly communicated and announced the programme	62.5% (n=190)	35.5% (n=108)	1.32% (n=4)	0.66% (n=1)
The programme was consistently delivered during the school day	61.18% (n=186)	30.59% (n=93)	6.9% (n=21)	1.32% (n=4)
Teachers were motivated and committed to delivering the programme	65.46% (n=199)	30.92% (n=94)	1.97% (n=6)	1.64% (n=5)
SURVEY: TEACHERS (N=913)				
I was able to deliver the programme according to the schedule	58.60% (n=535)	36.47% (n=333)	4.27% (n=39)	0.66% (n=6)
I had flexibility to adapt instruction to student needs	54.76% (n=500)	30.20% (n=367)	3.83% (n=35)	1.20% (n=11)
The goals of the literacy programme were clear and understandable	55.31% (n=505)	41.84% (n=382)	2.63% (n=24)	0.22% (n=2)
I felt supported by school leadership	68.57% (n=626)	28.48% (n=260)	2.19% (n=20)	0.77% (n=7)
SURVEY: QUALITY ASSURANCE TEAM (N=203)				
Programme followed a clear and realistic timeline	46.31% (n=94)	48.77% (n=99)	3.94% (n=8)	0.99% (n=2)
Schools adhered to intended delivery model	50.25% (n=97)	47.78% (n=97)	1.48% (n=3)	0.49 (n=1)
Classroom observations were conducted regularly	57.64% (n=117)	39.90% (n=81)	1.97% (n=4)	0.49% (n=1)
SURVEY: SCHOOL PRINCIPALS (N=448)				
Flexibility to schedule the programme to fit school needs	56.92% (n=255)	35.49% (n=159)	6.25% (n=28)	1.34% (n=6)
The programme was implemented consistently across all relevant grade levels	52.90% (n=237)	40.62% (n=182)	5.36% (n=24)	1.12% (n=5)
I regularly monitored classroom implementation of the programme	74.33% (n=333)	24.33% (n=109)	0.67% (n=3)	0.67% (n=3)

COMMUNICATION DOMAIN				
SURVEY: PARENTS (N=304)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
There was effective communication between the school and parents	64.47% (n=196)	31.25% (n=95)	2.30% (n=7)	1.97% (n=6)
SURVEY: QUALITY ASSURANCE TEAM (N=203)				
Goals of the literacy programme were clearly defined	55.17% (n=112)	44.33% (n=90)	0.49% (n=1)	---
SURVEY: SCHOOL PRINCIPALS (N=448)				
Goals of the literacy programme were clearly communicated	56.47% (n=253)	40.85% (n=183)	2.01% (n=9)	0.67% (n=3)

ENGAGEMENT AND DELIVERY DOMAIN				
SURVEY: PARENTS (N=304)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
My child was excited or motivated to participate in the programme	64.47% (n=196)	31.25% (n=95)	2.30% (n=7)	1.97% (n=6)
The literacy activities were engaging and appropriate for my child's age	60.53% (n=184)	29.61% (n=90)	7.57% (n=23)	2.30% (n=7)
The duration and frequency of sessions were appropriate	57.24% (n=174)	32.89% (n=100)	7.24% (n=22),	2.63% (n=8)
SURVEY: TEACHERS (N=913)				
Student engagement and participation in literacy activities	45.67% (n=417)	40.64% (n=371)	10.62% (n=97)	3.07% (n=28)
Teachers' preparedness and engagement observed in classrooms	46.80% (n=95)	42.36% (n=86)	9.36% (n=19)	1.48% (n=3)
SURVEY: SCHOOL PRINCIPALS (N=448)				
Teacher engagement and commitment to the programme	72.77% (n=326)	25% (n=112)	1.79% (n=8)	0.45% (n=2)
Student attendance and participation during programme sessions	35.71% (n=160)	48.21% (n=216)	14.06% (n=63)	2.01% (n=9)

COORDINATION DOMAIN				
SURVEY: PARENTS (N=304)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
The programme addressed the individual needs of my child	61.84% (n=188)	30.26% (n=92)	5.92% (n=18)	1.97% (n=6)
The materials used were effective and child-friendly	59.87% (n=182)	31.25% (n=95)	6.91% (n=21),	1.97% (n=6)
SURVEY: TEACHERS (N=913)				
I was able to deliver the programme according to the schedule	58.60% (n=535)	31.25% (n=333)	6.91% (n=21)	1.97% (n=6)
I had flexibility to adapt instruction to student needs	54.76% (n=500)	40.20% (n=367)	3.83% (n=35)	1.20% (n=11)
The goals of the literacy programme were clear and understandable	55.31% (n=505)	41.84% (n=382)	2.63% (n=24)	0.22% (n=2)
SURVEY: QUALITY ASSURANCE TEAM (N=203)				
Responsiveness of schools to feedback or guidance	55.17% (n=112)	40.39% (n=82)	4.43% (n=9)	---
Coordination and communication with other programme teams	63.05% (n=128)	33% (n=67)	2.96% (n=6)	0.99% (n=2)
SURVEY: SCHOOL PRINCIPALS (N=448)				
I received sufficient support from technical or quality assurance teams	57.37% (n=257)	36.38% (n=163)	5.13% (n=23)	1.12% (n=5)
The programme was aligned with my school's broader literacy objectives	66.29% (n=297)	30.80% (n=138)	1.79% (n=8)	1.12% (n=5)

OUTCOMES DOMAIN				
SURVEY: PARENTS (N=304)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
My child's reading and writing skills improved as a result of the programme	64.14% (n=195)	31.91% (n=97)	2.30% (n=7)	1.64% (n=5)
My child's self-confidence increased through participation in the programme	61.84% (n=188)	34.54% (n=105)	1.64% (n=5)	1.97% (n=6)
Overall satisfaction with the literacy programme	69.41% (n=211)	28.29% (n=86)	1.32% (n=4)	0.99% (n=3)
SURVEY: TEACHERS (N=913)				
Improvements in students' reading and writing abilities	44.58% (n=407)	46.44% (n=424)	7.78% (n=71)	1.20% (n=11)
Overall satisfaction with the literacy programme	58.93 % (n=538)	36.47% (n=333)	3.29% (n=30)	1.31% (n=12)
SURVEY: QUALITY ASSURANCE TEAM (N=203)				
Observable improvement in students' literacy performance	43.84% (n=89)	43.35% (n=88)	11.33% (n=23)	1.48% (n=3)
Overall satisfaction with the literacy programme	67.98% (n=128)	31.03% (n=63)	0.99 (n=2)	

OUTCOMES DOMAIN				
SURVEY: PARENTS (N=304)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
SURVEY: SCHOOL PRINCIPALS (N=448)				
Observable improvements in students' reading and writing skill	47.10% (n=211)	48.44% (n=217)	4.24% (n=19)	0.22% (n=1)
Overall satisfaction with the literacy programme	72.54% (n=325)	25.45% (n=114)	1.34% (n=6)	0.67% (n=3)

TRAINING AND MATERIALS DOMAIN				
SURVEY: TEACHERS (N=913)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
I received adequate training before implementing the programme	41.29% (n=377)	44.25% (n=404)	11.61% (n=106)	2.85% (n=26)
Effectiveness of training in preparing me to teach the programme	51.15% (n=467)	38.23% (n=349)	7.56% (n=69)	3.07% (n=28)
Programme related materials supported my teaching effectively	62.10% (n=567)	34.72 % (n=317)	3.07% (n=28)	0.11% (n=1)
Availability of materials and resources during implementation	39.21% (n=358)	39.65 % (n=362)	15.01% (n=137)	6.13% (n=56)
Clarity and usefulness of the teacher guide	61.12% (n=558)	32.42% (n=296)	5.15% (n=47)	1.31% (n=12)
SURVEY: QUALITY ASSURANCE TEAM (N=203)				
Clarity of programme expectations shared with schools	56.65% (n=115)	36.95% (n=75)	5.91% (n=12)	0.49% (n=1)
Availability and quality of teaching materials	40.89% (n=83)	44.33% (n=90)	12.81% (n=26)	1.97% (n=4)
SURVEY: SCHOOL PRINCIPALS (N=448)				
Teachers received adequate training before implementing the programme	39.29% (n=176)	41.74% (n=187)	16.52% (n=74)	2.46% (n=11)
The quality and clarity of teacher training	48.88% (n=219)	40.18% (n=180)	8.93% (n=40)	2.01% (n=9)
The availability and usefulness of programme materials	47.99% (n=215)	39.29% (n=176)	11.16% (n=50)	1.56% (n=7)

FEEDBACK & SUPERVISION DOMAIN				
SURVEY: TEACHERS (N=913)	**STRONGLY AGREE**	**AGREE**	**DISAGREE**	**STRONGLY DISAGREE**
I received constructive feedback on my teaching	51.90% (n=483)	43.37% (n=396)	2.63% (n=24)	1.10% (n=10)
Frequency and helpfulness of observation or feedback	41.62% (n=380)	45.01% (n=411)	9.97% (n=91)	3.40% (n=31)
Incentives for teachers participating in the programme	27.93% (n=255)	29.24% (n=267)	16.98% (n=155)	25.85% (n=236)
SURVEY: SCHOOL PRINCIPALS (N=448)				
Opportunities for school leadership to provide feedback and input on the programme	52.90% (n=237)	41.07% (n=184)	4.69% (n=21)	1.34% (n=6)
Financial incentives provided to school leadership for involvement in the programme	29.69% (n=133)	39.51% (n=177)	13.62% (n=61)	17.19% (n=77)

Endnotes

1 World Bank, UNESCO. *Egypt, Arab Republic - Learning Poverty Brief - 2024 (English)*. Learning Poverty Brief Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/099060124163512192

2 Patrinos, H. A., Jakubowski, M., & Gajderowicz, T. (2025). Evaluation of achievement change in Egypt using PIRLS data. *International Journal of Educational Development*, 117, 103371. https://doi.org/10.1016/j.ijedudev.2025.103371

3 PIRLS IEA. (2023). *PIRLS 2021 International Results in Reading: Performance at International Benchmarks.* https://pirls2021.org/wp-content/uploads/2022/files/P21_4-international-benchmarks.pdf

4 Mullis, I.V.S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K.A., & Wry, E. (2023). PIRLS 2021 International Results in Reading. Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb5342

5 MoETE (2021). National test for fourth grade. Executive summary for the baseline assessment.

6 Al-Masry Al-Youm. (2025, March 1). *The Ministry of Education launches the national program to develop reading and writing skills for one million primary school students.* Al-Masry Al-Youm. https://www.almasryalyoum.com/news/details/3389308

7 Because the baseline and endline assessments were identical, endline scores may be slightly inflated due to a practice effect. This could marginally overstate the gains in the descriptive statistics. However, statistical analysis at the 36-point threshold using a RDD estimation method still shows a significant positive impact of participating in the NPDA.

8 In Stata 14, a parametric localized approach was applied using the command rdrobust. A parametric globalized approach was also estimated using OLS, with the +/-4 band taken from the rdrobust output as a robustness check yielding similar results.

9 A subsample analysis was conducted by gender, using RDD to estimate the change in scores by gender. To estimate whether the difference is significant, we used OLS and included interaction terms for the treatment and gender, showing that the difference was significant. The OLS estimates yielded a slightly larger magnitude in score difference of 7.4 (5.9 for the reference of males and the 1.39 for females).

Acknowledgements

This report was written and prepared by Thomas Dreesen, Ghalia Ghawi, Sophia Kan, Lara Naber Ramadan Muhammad Ramadan, Hasib Muhammad Hasib, Omaima Riyad, and Abdullah Zekry. The report greatly benefited by the consistent leadership, guidance, support, input and feedback of Ebtehal Elghamrawy, Ahmed Mohamed Mahrous, Shiraz Chakera, Amira Fouad, Ayman Al-Husseini, Aferdita Spahiu, Chao Wang, Brenda Haiplik and Linda Jones. This research could not have been completed without the commitment and hard work of the programme management team of the Public Education Central Department under the leadership of Dr. Hala Abd El Salam, as well as the implementation from the National Center for Examinations and the Ministry of Education's central and directorate teams, and the dedicated and hardworking teachers, that made this programme possible.

Disclaimer: This is a working document. It has been prepared to facilitate the exchange of knowledge and to stimulate discussion. The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors. The statements in this publication are the views of the author(s) and do not necessarily reflect the policies or the views of UNICEF. The designations in this publication do not imply an opinion on legal status of any country or territory, or of its authorities, or the delimitation of frontiers.

Design: Afternorth
Photography: © UNICEF/Egypt

About us

UNICEF works in the world's toughest places to reach the most disadvantaged children and adolescents — and to protect the rights of every child, everywhere. Across 190 countries and territories, we do whatever it takes to help children survive, thrive and fulfill their potential, from early childhood through adolescence. And we never give up.

UNICEF Innocenti – Global Office of Research and Foresight tackles the questions of greatest importance for children, both current and emerging. It drives change through research and foresight on a wide range of child rights issues, sparking global discourse and actively engaging young people in its work.

UNICEF Innocenti equips thought leaders and decision-makers with the evidence they need to build a better, safer world for children. The office undertakes research on unresolved and emerging issues, using primary and secondary data that represents the voices of children and families themselves. It uses foresight to set the agenda for children, including horizon scanning, trends analysis and scenario development. The office produces a diverse and dynamic library of high-level reports, analyses and policy papers, and provides a platform for debate and advocacy on a wide range of child rights issues.

UNICEF Innocenti provides, for every child, answers to their most pressing concerns.

Published by

UNICEF Innocenti – Global office of Research and Foresight

Via degli Alfani, 58

50121, Florence, Italy

Tel: (+39) 055 20 330

Web: unicef.org/innocenti

Email: innocenti@unicef.org

Social media: @UNICEFInnocenti on Instagram, LinkedIn and YouTube

In partnership with

UNICEF Egypt

Building No 9 in 14 st, Maadi, Cairo

Tel: (+20) 225265083

Suggested citation

Dreesen, T., Ghawi, G., Kan, S., Naber, L., Ramadan, R. M., Hasib, H. M., Riyad, O., and Zekry, A., *Developing the Arabic Language Skills in Egypt: Assessing the Implementation and Impact of the National Programme for Developing Arabic Language Skills (NPDA)*, Working Paper, UNICEF Innocenti, Florence, September 2025.

© United Nations Children's Fund (UNICEF), September 2025

for every child, answers

unicef

for every child