PROJECT TITLE	**Let's Read Fluently!**
DEVELOPER (INSTITUTION)	Queen Rania Teacher Academy (QRTA)
EVALUATOR (INSTITUTION)	National Centre for Social Research (NatCen), with Integrated, Oxford Measured (OM) and STS International (STS)
PRINCIPAL INVESTIGATOR(S)	Sashka Dimova
STUDY PLAN AUTHOR(S)	Mansor Rezaian, Hannah Woodbridge, Julia Griggs, Enes Duysak, Andi Fugard, Lydia Marshall, Rachel Outhred, Nedjma Koval and Daniel Phillips
PUPIL AGE RANGE AND KEY STAGE	Ages 6 to 9 (Jordanian Grades 1 to 3)
NUMBER OF SCHOOLS/ SETTINGS	24 (**W/C:** 8; **C/U:** 8; **Control:** 8)
NUMBER OF PUPILS	587 (**W/C:** 180; **C/U:** 114; **Control:** 294)

Study plan version history

VERSION	DATE	REASON FOR REVISION
1.1		
1.0 [*original*]	28.09.21	

Intervention

Context: The importance of early literacy

A strong foundation in literacy is crucial for children to develop. Evidence shows early literacy difficulties can persist, limiting children's ability to achieve their potential (Brombacher et al., 2012). Results using the Early Grade Reading Assessment (EGRA) conducted in Jordan since 2012 have shown that primary school-aged children are failing to reach reading comprehension benchmarks (RTI International, 2018). This is consistent with other international assessments, which have found Arabic students tend to score lower in reading than others (Eckert et al, 2020). Evidence from other school systems suggests that it is very unlikely these students will make up for this learning loss[1] during the next stages of their education, leaving these children at a significant disadvantage throughout their schooling and life.

Learners of Arabic face unique challenges, including its diglossic nature. The colloquial variety pupils use at home differs from the formal Modern Standard Arabic (MSA) introduced at school (Abadzi, 2017). Pupils entering school are consequently tasked with absorbing MSA alongside developing their literacy. Furthermore, Arabic script is complex compared to others (Eckert et al, 2020)[2]. 'Let's Read Fluently!' (LRF) is an intervention that aims to support children in overcoming these obstacles to successfully develop their foundational literacy skills. LRF centres around a method and student practice book designed to reflect evidence about how Arabic reading fluency is best acquired. The approach utilizes an "I-do, we-do, you-do" teaching model that is explicitly designed to encourage repeated practice to help the students develop automaticity in letter reading and hence greater fluency and confidence in reading. Teachers receive one-day training from the Queen Rania Teacher Academy (QRTA) and up to three follow-up coaching sessions to support them to use the practice book.

The pilot will explore the intervention's feasibility and evidence of promise and assess readiness for trial. If we progress to an efficacy trial, it will provide evidence on the impact of LRF, mechanisms of change and lessons to inform future scale-up.

The *Let's Read Fluently!* intervention

The Let's Read Fluently approach involves a practice-focused pedagogy and student practice book was developed by cognitive psychologist Dr Helen Abadzi , and the Al Qasimi Foundation[3] in the UAE. The approach draws upon insights from studies in linguistics and cognitive science accounting for the Arabic script's visual complexities and the relationship between memory function and reading. It has been developed to help students build 'low level' neurological functions - rapidly distinguishing letter shapes, chunking and decoding sounds and words.[4] Similar approaches to early literacy teaching in Cambodia, the Gambia and Egypt have shown evidence of promise.[5] There are also early results from a small-scale pilot conducted in the UAE, which suggest an LRF approach may have a positive impact on Arabic reading fluency in early grade students.[6]

---

[1] World Bank (2019) Ending learning poverty: what will it take? Washington, D.C.: World Bank. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/32553/142659.pdf?sequence=7.

[2] Eckert, M., Wilson, E., Abadzi, H., & Jeon, S. (2020). Improving Arabic reading fluency: Results from Iqra, an early-grade reading intervention in Ras Al Khaimah (Policy Paper No. 39). Sheikh Saud bin Saqr Al Qasimi Foundation for Policy Research. http://dx.doi.org/10.18502/aqf.0142

[3] Ibid.

[4] Abadzi, H. (2013) 'Literacy for All in 100 Days: A research-based strategy for fast progress in low-income countries', *GPE Working Paper Series on Learning No. 7*.

[5] Ibid.

[6] Eckert, M., Wilson, E., Abadzi, H. and Jeon, S. (2020) 'Improving Arabic Reading Fluency: Results from Iqra, an Early-Grade Reading Intervention in Ras Al Khaimah', *Policy Paper No. 39*, Sheikh Saud bin Saqr Al Qasimi Foundation for Policy Research.

There are two models of LRF implementation, a Whole Class Teaching and Learning approach (W/C) and a Literacy Catch-Up (C/U) approach, both lasting one semester. W/C targets students in Grade 1 and is delivered to classes via three 30-minute classes per week. In class, the teacher adopts a 'I do', 'we do', 'you do' pedagogical approach using the practice book. C/U targets the 20% lowest-achieving students in Grades 1 to 3 and is delivered through small group tuition for 3 sessions per week. This protocol describes the pilot evaluation of both implementation models. The two models are described in more detail below (see 'What').

In the summer of 2021, a pre-pilot of both the W/C and C/U models was conducted. This was over an 8-week period in the summer[7]. The pre-pilot was carried out to inform the design of the pilot by considering how the program was implemented, and whether any components of the intervention should be adjusted. This included the practice book, the structure of the teacher and student interaction, the teacher training and the approaches to delivering C/U and W/C. To do this, the pre-pilot evaluation gathered contextual information, teacher feedback, classroom observation and fidelity of implementation (FOI) data.

Pre-pilot results suggested a set of adaptations for the pilot. Table 1 below summarises the changes.

Table 1: Let's Read Fluently – adaptations from pre-pilot

Adaptations made to:		No adaptations made to:
Practice book	**Classroom instruction**	
Recorded the lessons for the purpose of supporting teachers with an accurate model of reading in preparation of their lessons and supporting parents and/or students when they practice at home.	Encouraged teachers to draw on their professional judgement and knowledge of their students with regard to the flow of the routine of (I do, we do, you do) i.e. do 1-2 pages in one cycle and the rest in other cycle with careful attention that the independent reading time is not than 15 minutes in total. The purpose of this change is increasing students' attention span and engagement.	Coaching (delivery of LRF is perceived to be inexpensive, so despite recommendations to increase the number of coaching sessions no increase was made to keep delivery cost low).
Reviewed invented words to ensure they follow word phonetic rules in Arabic.	Emphasized starting where students are standing, explicitly introducing the new letter, diacritics if they feel the need.	Diagnostic tool
Reviewed language accuracy.	Emphasized the importance of achieving mastery before proceeding to the next lesson.	Teacher rewarding achievement
Added a self-tracking tool in the footnote of each page. This is a question asking the student: How many times did you read this page?	Emphasized techniques that motivate students to increase their independent reading stamina. i.e. increasingly challenging them to add to the minutes they are on task, asking them how much did exercise your brain today.	

[7] The original pre-pilot intervention launched in March 2021 but paused after less than two weeks due to COVID related schools' closures. Subsequently, the pre-pilot relaunched in two community centres in summer 2021. Students who participated in the pre-pilot had either finished grades 1 or 2.

Added another tool to track practice at home. For each lesson, students can colour a figure of a brain carrying weights to reflect their answer to the question (How much did you exercise your brain today?)	

Why

It is estimated that early readers in Arabic need a level of automaticity[8] in oral reading fluency[9] of 45-60 words per minute[10]. This fluency allows working memory to be freed-up for comprehension. Data from the use of the Early Grade Reading Assessment (EGRA) tool in 2018 suggests that only around 19% of Grade 2 and Grade 3 students meet, or exceed, the lowest levels of this benchmark. Alongside that, a significant number of students in Jordan (16.6% in 2018) scored zero in oral reading fluency.[11]

Early grade interventions can be beneficial in terms of helping students as they progress in grade level. The 2018 Program for International Student Assessment (PISA) found that, for 15-year-old Jordanian students', attainment levels were behind the OECD average by an equivalent of more than one grade in reading. Only one in five students performed at or around the average OECD reading score and two in every five performed below the minimum proficiency level in reading.[12]

There are also concerns with global levels of literacy, and in 2019, the World Bank announced its 'Literacy Makes Sense' approach to reduce what they describe as 'learning poverty'.[13] Within the context of Jordan, the report estimated that 52% of Jordanian 10- year-olds are unable to read and understand a short age-appropriate piece of text.

When diacritics[14] are used, Arabic is a transparent language – that is, there is a reliable relationship between letters and sounds. Given this, these low EGRA scores likely reflect a gap in phonics skills.[15] The importance of phonics is reflected in the EEF's Teaching and Learning Toolkit and other literature.[16]

---

[8] Automaticity is defined as being able to complete a task with no conscious effort, in much the same way as you are able to read this footnote or calculate 2×2.
[9] Oral reading fluency is the ability to read connected text quickly, accurately and with expression. In doing so, there is no noticeable cognitive effort associated with decoding the words on the page.
[10] RTI, Aug. 2012, Student Performance in Reading and Mathematics, Pedagogic Practice, and School Management inJordan,https://earlygradereadingbarometer.org/files/EGRA%20in%20Jordan.pdf
[11] RTI, Nov. 2018, Early Grade Reading and Mathematics Initiative Lot Quality Assurance Sampling Assessment. https://ierc-publicfiles.s3.amazonaws.com/public/resources/Jordan%20RAMP%20LQAS%20Summary%20Report%20FINAL.pdf
[12] QRF; 'Exploring Jordan's performance' (September 2020), https://www.qrf.org/en/what-we-do/research-and-publications/pisa-2018-exploring-jordan%E2%80%99s-performance
[13] World Bank (2019) Ending learning poverty: what will it take? Washington, D.C.: World Bank. Available at: https://openknowledge.worldbank.org/bitstream/handle/10986/32553/142659.pdf?sequence=7
[14] Diacritics are marks placed above or below (or sometimes next to) a letter in a word to indicate the short vowels.
[15] See: https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/phonics/
[16] See (1) Seidenberg, M. (2017) Language at the Speed of Sight: How we read, why so many can't, and what can be done about it. New York, NY: Basic Books or (2) Castles, A., Rastle, K. and Nation,

It is important to note that LRF and its evaluation are taking place amid the ongoing Covid-19 crisis, which has led to sustained school closures. This context reinforces the need for interventions to support literacy acquisition and strong evidence to understand what works[17].

Who

Teachers are at the core of the LRF intervention as they both receive and then deliver the program Teachers will receive a one-day training course from the Queen Rania Teacher Academy (QRTA) and up to three follow-up coaching sessions to support them in using the LRF practice book. They will deliver the LRF intervention to pupils in Grades 1 to 3. Thus, both pupils and teachers can be considered intervention recipients.

All pupils in the second semester of Grade 1 and at schools and classes selected for the **W/C implementation model** are eligible for the intervention. The rationale being that introducing the practice book at this stage allows an alignment with the sequence of letter-sound introductions as set out in the Grade 1 textbooks used in standard literacy teaching (i.e. usual practice).

The 20% of lowest achieving pupils in a class in the second semester of Grade 1, or in the first semester of Grades 2 and 3 and in a class selected for the **C/U implementation model**, are eligible for the intervention. Semester 2 was the earliest it was deemed feasible to identify struggling readers who were in Grade 1. QRF felt that the intervention could be more easily administered to pupils in Grades 2 and 3 in semester 1.

The Queen Rania Foundation (QRF) will recruit primary schools and QRTA will provide training and support teachers to deliver LRF. School principals and supervisors will attend three-hour orientation sessions which will inform them about what teachers need to do as part of the intervention and aim to equip them with the skills to support implementation.

What

Overview

Let's Read Fluently uses a reading practice book, which is delivered to primary school pupils in Grades 1 – 3 as part of their literacy learning. The practice book was developed by Dr Helen Abadzi and Al Qasimi Foundation[18] in the UAE.

W/C targets all students in designated classes in Grade 1 and is delivered by classroom teachers via three 30-minute classes per week. In class, the teacher adopts a 'I do', 'we do', 'you do' pedagogical approach using the practice book.

C/U targets the 20% of lowest-achieving students in Grades 1 to 3 and is delivered via small group tuition in each school's resource room. LRF is delivered by resource room teachers via three 30-minute sessions per week. Classroom teachers in schools selected for the C/U model will attend a training day but will not be directly involved in delivering LRF to pupils.

Table 2 shows the phases of the intervention and pilot evaluation for both the W/C and C/U implementation models. Specifically, the C/U model will run with Grades 2 and 3 in Semester 1 of the 2021-22 year, and with Grade 1 in Semester 2. The W/C model will run with Grade 1 in Semester 2.

K. (2018) 'Ending the reading wars: Reading acquisition from novice to expert', Psychological Science in the Public Interest, 19(1), pp. 5–51.

[17] UNICEF, 'The Impact of COVID-19 on children in the Middle East and North Africa' (November 2020), https://www.unicef.org/mena/media/10231/file/Impact%20of%20COVID%20on%20Children-Snapshot%20report.pdf%20.pdf

[18] http://www.alqasimifoundation.com/en/home

Table 2: Let's Read Fluently – phases for the pilot year

	2021-22 Academic Year	
Implementation model (Grade)	Semester 1 (Sept 21 – Jan 22)	Semester 2 (Feb 22 – July 22)
C/U (Grade 1)		**Intervention runs**
C/U (Grade 2, 3)	**Intervention runs**	
W/C (Grade 1)		**Intervention runs**

Teacher training and coaching

The Queen Rania Teacher Academy (QRTA) will train classroom and resource room teachers in how to use the LRF method and practice book. This will be one day of face-to-face training, which includes:

- the rationale for the project
- teachers' role in the intervention
- the learning experience teachers are being asked to facilitate
- how to appropriately communicate the project to parents/carers (this should include conducting a face-to-face awareness raising meeting for parents)
- how to support the involvement of parents/ carers – for example, encouraging students' use of the LRF practice book at home with parents/ carers, and supporting parents in this, (i.e. with WhatsApp messages).

The training will include opportunities to practice the new teaching and learning technique, and to explore potential barriers and how they can be overcome. School Supervisors have no formal role in implementing the programme but will be invited to attend the training session alongside teachers in their District. Principals of schools in the intervention group will attend a separate orientation session.

Following on from the training sessions, teachers will receive up to three coaching visits (accompanied by classroom observations from the coaches). The coaching visits will be delivered by QRTA staff. To facilitate these, QRTA will use a coaching model designed to enhance implementation on the part of teachers.

Teacher training and coaching will focus on just one of the two implementation models. Teachers will be asked to deliver either the W/C or C/U intervention (not both) based on the randomised assignment.

What: Whole Class (W/C)

The Whole Class teaching and learning approach (W/C) is targeted at students in the second semester of Grade 1. The rationale for delivery in the second semester is that introducing the practice book at this stage allows an alignment with the sequence of letter-sound introductions specified in the 'business-as-usual' Grade 1 textbooks.

The intervention will be delivered in three 30-minute sessions each week, for 12 weeks. Following approval from the MoE, the classroom time for this will be taken from one of the three 'free activity periods' in the Jordanian school curriculum, and two of the seven Arabic language classes. First, using large versions of the textbook, the classroom teacher will introduce the letter-sound, or letter combinations, and model how to 'read' it. ("I do"). This is followed by an opportunity for the whole class to practice 'reading' using either the choral or echo method ("we do"). These two steps should be completed in the first 10 to 15 minutes of the session. Following this, learners are asked to independently work through the student practice book, taking each item in turn and with their finger on the text sounding out the letter, or word ("you do"). At this stage of independent student practice, the teacher's role is to encourage engagement with the task and to provide feedback (namely, reinforcement and corrections). This stage of independent practice with teacher feedback should be 15-20 minutes, which is half to two-thirds of the session. This is a key feature of the LRF model as

research in cognitive science indicates individuals need to independently and repeatedly practice decoding to develop the automaticity needed for fluent reading.

All students will receive a copy of the practice book. They will be encouraged to take it home for extra practice with their parents/carers, with teachers supporting this form of parental engagement in two ways: (1) by raising awareness through an introductory meeting with parents, and (2) through communicating via WhatsApp messages to parents about the support needed with practice at home following the lesson.

In the two sessions that will be conducted in the Arabic classes, LRF sessions will replace 'business-as-usual' teaching. Joint working with the Jordanian Ministry of Education has ensured that this appropriately aligns with existing curriculum content so that students have a coherent learning experience.

Literacy Catch-Up (C/U)

The Literacy Catch-Up (C/U) model targets the 20% of lowest achieving students in a class either in the second semester of Grade 1, or in the first semester of Grades 2 and 3. The rationale for the former is that it is felt that the first semester of Grade 1 is too early to identify struggling readers. For the latter, the rationale is simply project planning.

Schools in the Literacy Catch-Up intervention arm will be expected to deliver this extra support to selected students in all three Grades (1 to 3), for three sessions a week over one semester.

Selection of students for the intervention will be carried out by teachers, using the coarse grained diagnostic tool developed by the Early Grade Reading and Mathematics Programme (RAMP)[19] - already in use in Jordanian classrooms. In the teacher training, teachers will be advised to exclude students who have severe learning difficulties [20], as the LRF intervention was not designed with their needs in mind[21]. That is, QRTA will advise that eligible students are those for whom it is plausible that a change in teaching and learning approach will be helpful.

In this model, the intervention will be delivered by resource room teachers (these are teachers whose role is to provide extra support to struggling students). Our current understanding is that the scheduling of extra support to struggling learners (referred to as 'Resource Room teaching' in Jordan) is agreed between the classroom teacher and the resource room teacher, that schools have autonomy over this scheduling and that this extra support is provided during the normal school day. The Jordanian Ministry of Education (MoE) has given permission to use one 'free activity period' for literacy catch-up. Teachers will be advised to schedule the remaining two sessions in a way that minimises any disruption to normal learning.

In effect, this is a form of extra small group tuition, using the practice book as the learning material. The C/U model will be delivered to groups of five-to-six students with similar literacy learning needs.

How

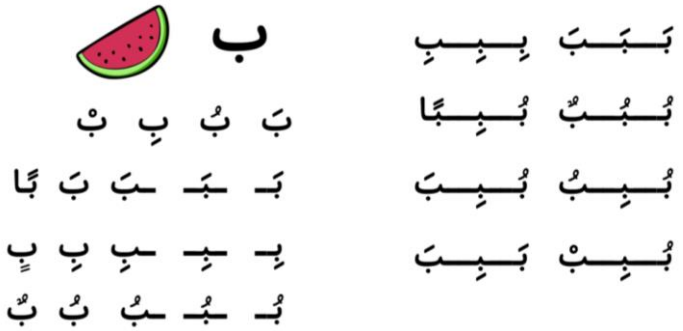
Students are taught to process written text more quickly by firstly repeating individual letters and words to the point of automation. This is intended to enable them to decode read faster, in order to read more fluently and free up working memory to recall important information and think critically. Time engaged in practice and receiving timely feedback (namely, reinforcement and corrections) are seen as important predictors of reading ability.

[19] https://www.usaid.gov/jordan/fact-sheets/early-grade-reading-and-mathematics-project-ramp
[20] The description used in QRTA communications with teachers is that LRF is suitable for 'students who are academically behind but don't suffer from mental or physical illnesses'.
[21] The description used in QRTA communications with teachers is that LRF is suitable for 'students who are academically behind but don't suffer from mental or physical illnesses'.

The LRF practice book is designed to encourage perceptual learning for decoding, as well as reading practice to attain fluency.[22] It includes a number of design features intended to tackle barriers to literacy and/ current understanding about what works for early readers:

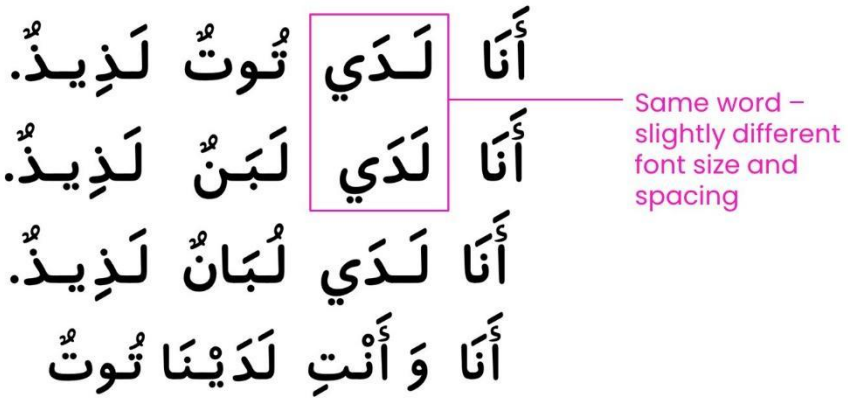
- Small font sizes negatively affect letter identification, so the book uses large font sizes and spacing.
- The Arabic script is dense and complex, and so creates a higher cognitive load for new readers than other languages. The book and LRF model more generally, place importance on repetition and teacher feedback.
- New letter shapes are introduced slowly, one by one.
- It follows a phonics-based approach in which children gradually decode words using their phonics knowledge rather than using other clues or seeking help
- Pattern analogies can assist learning, so common sounds are stressed (e.g., da di du, which links the ‘d’ sound with each of the short vowels).
- Students need to see meaning in text, so real words and sentences are introduced as soon as possible.
- The use of pictures in the text are minimised to ensure students learn letter sounds, rather than guessing.

The student practice book stresses repetition of patterns, alongside lots of practice in recognising them. See the examples below:



The practice book includes text with subtle differences to encourage pupils to recognise common words, even when presented slightly differently.

[22] Content is drawn from the LRF materials.



Where

The intervention will be delivered in the classrooms (W/C) or resource rooms (C/U) of participating schools.

When and how much

Intervention delivery will last 12 weeks, starting from 3 October 2021 and running to 24 December 2021 in semester 1 and 13 February to 19 May 2022 in semester 2 (there is a short break for Eid in semester 2). Pupils will receive three LRF sessions per week for the full 12 weeks. This delivery model applies to both C/U and W/C. The intended delivery of LRF is across 14 weeks, 12 weeks of core material and 2 weeks of revision sessions. Only the 12 weeks of core material is being delivered for the pilot study, as the EGRA tests reduce the number of weeks available in each semester,

Tailoring

Teachers have some flexibility over how they facilitate the sessions; however, the content of each session is set. Teachers are asked to allow for 15-20 minutes of individual independent practice after each lesson.

Teachers will be encouraged to draw on their professional judgement about tailoring instruction according to students' needs. Teachers will be expected to ensure they adequately progress through the content of the practice book, while at the same time ensuring pupils are able to adequately master each 'lesson' as they do so.

Control condition

Pupils in control schools will receive teaching as usual. No schools (intervention or control) will be offered financial incentives to participate.

Logic model

A logic model was developed for each strand of the intervention (C/U and W/C) in advance of the evaluation. Both LRF logic models were updated by the evaluation team in collaboration with QRF (see Figure 1 and Figure 2 below). The updating process included a logic model workshop on 27 July 2021 attended by key members of the QRF and evaluation teams.

Moderating factors

The implementation and process evaluation (IPE) will explore a range of likely moderating factors that will affect the outcomes of the intervention. These are likely to include;

- Teachers' understanding and commitment to the values of the intervention
- The quality of resources and materials (e.g. the student practice book)

- Feedback given to students during independent reading time
- Parents' understanding of LRF and the quality of their communication with schools.

Figure 1: LRF logic model – Whole Class

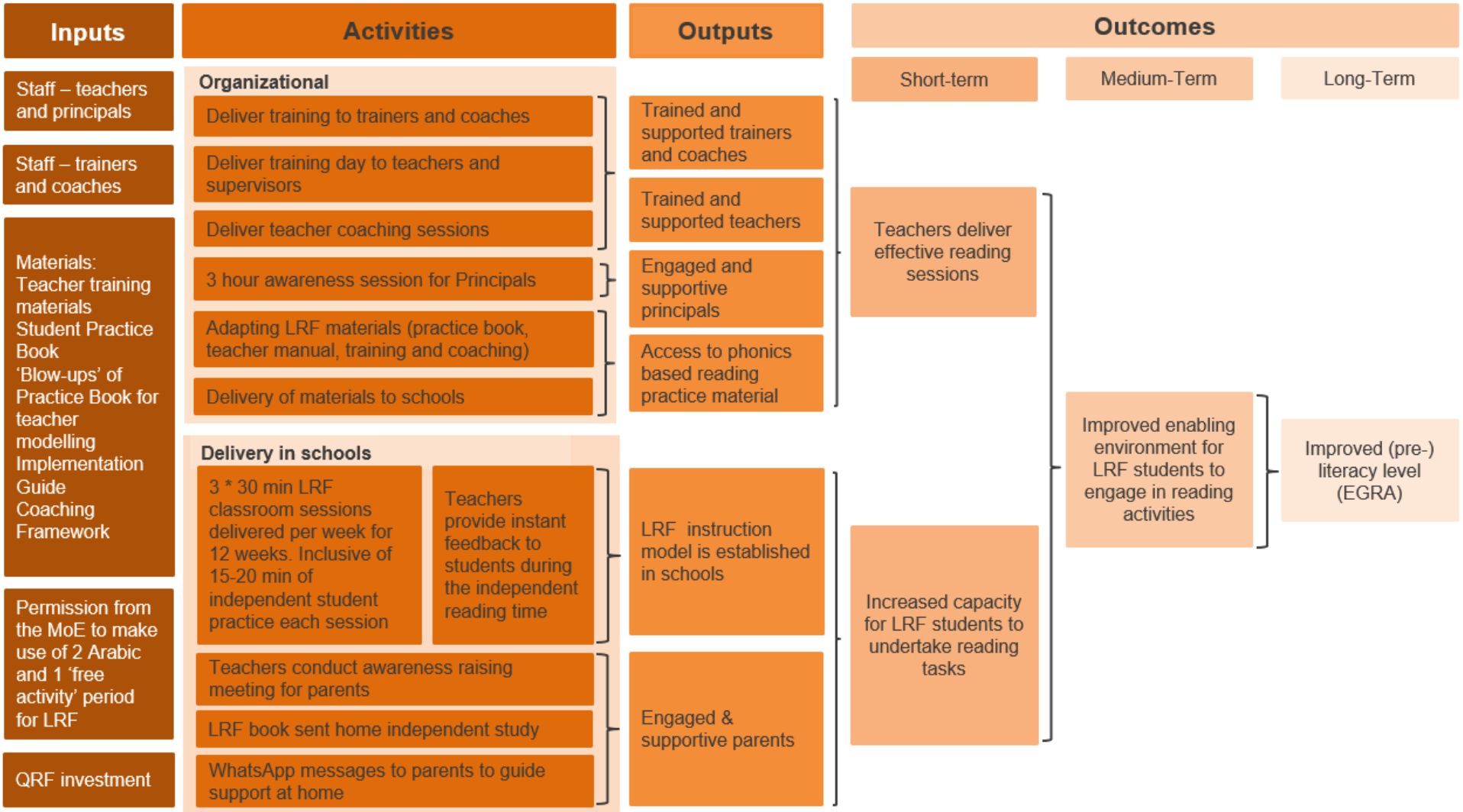
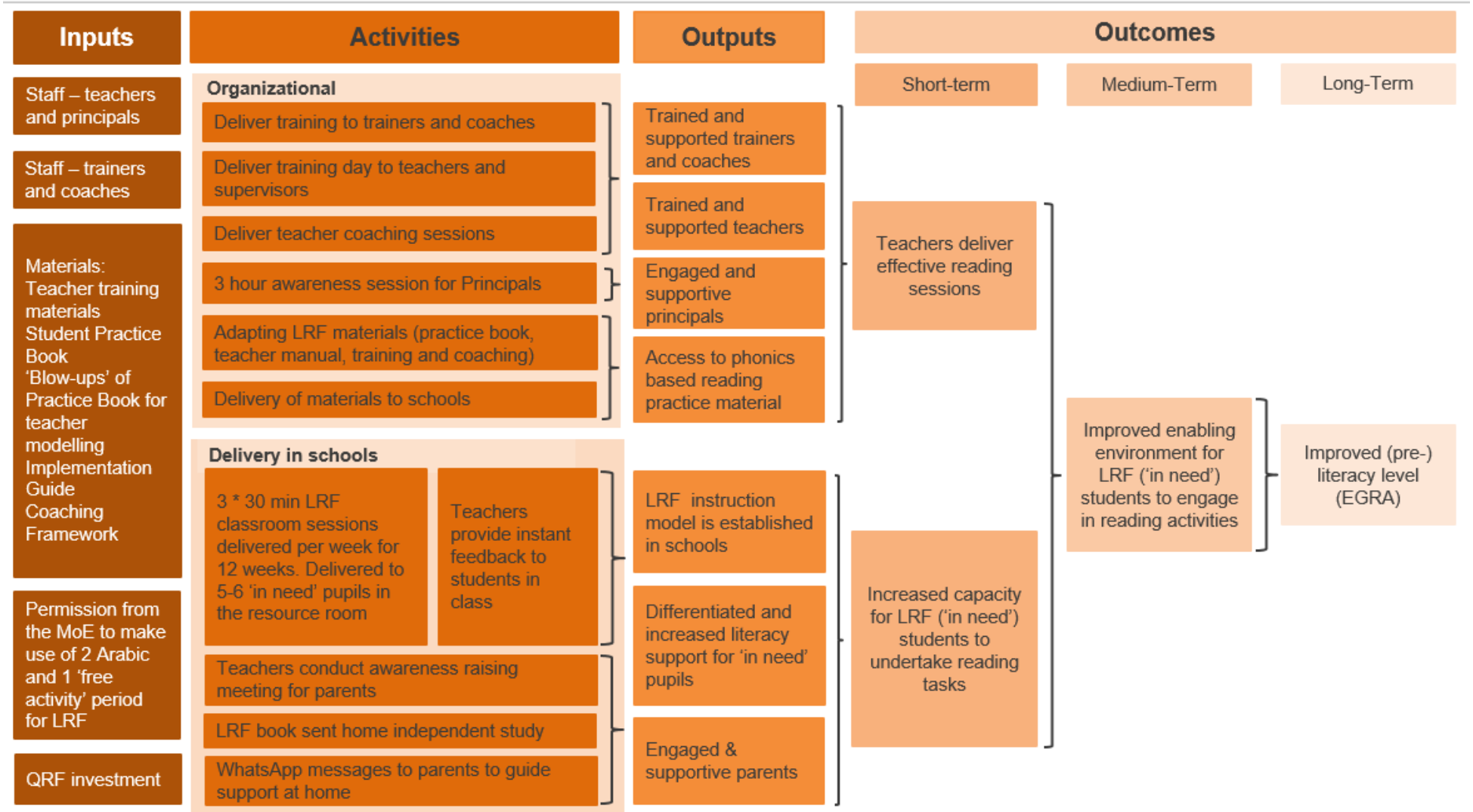


Figure 2: LRF logic model – Catch Up



Research questions

The pilot aims to answer the following research questions within the following domains:

Evidence of promise

1. In what ways, and to what extent, does 'Let's Read Fluently!' (LRF) affect school, teacher, student, and parental practice as compared to business-as-usual teaching and learning?

2. How do principals, teachers, parents and students perceive the intervention and any changes that it has delivered?

3. Is there evidence to support the Logic Models?

4. Is there any evidence of unintended consequences (negative or positive) as a result of the implementation of LRF?

Feasibility of intervention

5. Was LRF delivered as intended in terms of dosage, nature and quality? What modifications were made, with what implications?

6. What is the learning about teacher's use of the Coarse-Grained Diagnostic (RAMP) tool? How successful is, it in use, at identifying the most appropriate students for the Literacy Catch-Up implementation model (C/U model only)?

7. What were the facilitators and barriers to engagement in the teacher training, teacher coaching and Supervisor and Principal orientation sessions?

8. To what extent do teachers develop sufficient skills and confidence through the training and coaching?

9. What do we know about how teachers need to be supported (coached) during delivery?

10. Are there any key contextual factors that appear to facilitate or impede successful implementation of LRF?

Assessing feasibility of the efficacy trial(s)

11. What does the Pilot tell us about the feasibility of the process components of an Efficacy Trial, e.g., school recruitment, retention, or data collection in both intervention and control groups?

12. What does the Pilot tell us about the feasibility of the resources of an Efficacy Trial, e.g. measurement instruments or specific equipment used?

13. What does the Pilot tell us about the feasibility of the management components of an Efficacy Trial, e.g. problems with data collection or variability of collected data?

Assessing readiness for trial

14. What changes, if any, are needed to the Logic Models?

15. What changes to the intervention, implementation models, support or materials need to be made?

16. What can we learn from the Pilots about minimal detectable effect size estimates, intra-cluster correlations, pre-and-post correlations and sample sizes?

17. Is there any evidence of contamination between the control and treatment groups? For example, from the Supervisors who attend the training alongside teachers in their District spreading, or promoting, aspects of LRF to other schools under their Supervision.

Table 3 links the pilot research activities to the research questions and evaluation domains. Each of these activities is discussed in the Methods section, below.

Table 3: Research activities against evaluation domains and research questions

Research Questions	Research activity							
	EGRA tests	Focus Group Discussions (FGDs) with teachers	FGDs with parents	FGDs with pupils	Key Informant Interviews (KIIs) with school stakeholders	Classroom observations	Teacher survey	Pupil survey
In what ways, and to what extent, does 'Let's Read Fluently!' (LRF) affect school, teacher, student, and parental practice as compared to business-as-usual teaching and learning?		♦	♦	♦	♦		♦	♦
How do principals, teachers, parents and students perceive the intervention and any changes that it has delivered?		♦	♦	♦	♦		♦	
Is there evidence to support the Logic Models?	♦	♦	♦	♦		♦	♦	♦
Is there any evidence of unintended consequences (negative or positive) as a result of the implementation of LRF?		♦	♦	♦	♦		♦	♦

Feasibility of intervention	Research activity							
	EGRA tests	FGDs with teachers	FGDs with parents	FGDs with pupils	Klls with school stakeholders	Classroom observations	Teacher survey	Pupil survey
Was LRF delivered as intended in terms of dosage, nature and quality? What modifications were made, with what implications?		♦			♦	♦	♦	
What is the learning about teacher's use of the Coarse-Grained Diagnostic (RAMP) tool? How successful is it, in use, at identifying the most appropriate students for the Literacy Catch-Up implementation model (C/U model only)?		♦			♦			
What were the facilitators and barriers to engagement in the teacher training, teacher coaching and Supervisor and Principal orientation sessions?		♦			♦		♦	
To what extent do teachers develop sufficient skills and confidence through the training and coaching?		♦			♦	♦	♦	

What do we know about how teachers need to be supported (coached) during delivery?		♦			♦		♦	
Are there any key contextual factors that appear to facilitate or impede successful implementation of LRF?		♦	♦		♦	♦	♦	♦

Assessing feasibility of the efficacy trial(s)	Research activity							
	EGRA tests	FGDs with teachers	FGDs with parents	FGDs with pupils	Klls with school stakeholders	Classroom observations	Teacher survey	Pupil survey
What does the Pilot tell us about the feasibility of the process components of an Efficacy Trial, e.g., school recruitment, retention, or data collection in both intervention and control groups?	♦	♦	♦	♦	♦	♦	♦	♦
What does the Pilot tell us about the feasibility of the resources of an Efficacy Trial, e.g. measurement instruments or specific equipment used?	♦	♦	♦	♦	♦	♦	♦	♦
What does the Pilot tell us about the feasibility of the management components of an Efficacy Trial, e.g. problems with data collection or variability of collected data?	♦	♦	♦	♦	♦	♦	♦	♦

Assessing readiness for trial	Research activity							
	EGRA tests	Qualitative FGDs with teachers	Qualitative FGDs with parents	Qualitative FGDs with pupils	KIIs with school stakeholders	Classroom observations	Teacher surveys	Pupil surveys
What changes, if any, are needed to the Logic Models?		♦	♦	♦	♦		♦	♦
What changes to the intervention, implementation models, support or materials need to be made?		♦	♦	♦	♦	♦	♦	♦
What can we learn from the Pilots about minimal detectable effect size estimates, intra-cluster correlations, pre-and-post correlations and sample sizes?	♦							
Is there any evidence of contamination between the control and treatment groups?					♦	♦		

Methods

Trial design - pilot

The pilot evaluation will be conducted as a three-arm cluster randomised controlled trial (RCT). It will have one group that acts as a control group to two different treatment groups, which are Whole Class (W/C) and Literacy Catch-up (C/U). This pilot design is intended to inform, and test the feasibility of, a future efficacy trial. Even though the focus of the pilot is not on estimating an effect size, as it will be highly imprecise in a pilot evaluation of this size, this pilot will involve the analysis of the effects of each programme on Arabic literacy attainment to quality assure the data pipeline and statistical framework.

'Let's Read Fluently!' (LRF) is an intervention that aims to successfully improve foundational literacy skills of pupils by supporting them in overcoming obstacles they face while learning language. The primary outcome of interest is Arabic literacy attainment among Grade 1, Grade 2, and Grade 3 pupils in Jordan. The secondary outcomes are the specific sub-domains of Arabic literacy attainment for the same grades. These sub-domains are letter sound identification, speed and accuracy of word decoding, and reading comprehension. We will measure Arabic literacy attainment and its sub-domains by administering Early Grade Reading Assessment (EGRA) and a set of pre-literacy items which is explained in detail on p. 22. The same primary and secondary outcomes will be measured in the baseline using the same tools.

Schools will be randomly assigned into one of the three pilot arms (i.e. W/C, C/U and Control). Schools will be stratified by region and urban/rural classification prior to randomisation to ensure balance across pilot arms across strata after randomisation. Grade 1 pupils in schools assigned to the W/C programme and Grades 1, 2 and 3 pupils in schools assigned to the C/U model will be eligible to participate in the programme. 50% of pupils in Grade 1 in W/C schools will be tested in Arabic literacy attainment, whereas 50% of Grade 1, 2 and 3 pupils in the lowest 20% of their class in C/U schools will be tested in Arabic literacy attainment.

As indicated above, the pilot will explore the intervention's feasibility and evidence of promise and assess readiness for trial. We have also designed the same model for the efficacy trial. If we progress to an efficacy trial, it will provide evidence on the impact of LRF, mechanisms of change and lessons to inform future scale-up. Full details of the RCT design will be set out in a separate protocol, and accompanying analysis plan, for the efficacy trial (following completion of the pilot).

Table 4: Study design - pilot

Trial design, including number of arms		Pilot Evaluation
Unit of randomisation		School level
Stratification variables (if applicable)		Regions and urban/rural classification
Primary outcome	Variable	Arabic literacy attainment
	measure (instrument, scale, source)	**Source:** EGRA Grade 2 assessment with the addition of a set of pre-literacy items **Instrument**: Early Grade Reading Assessment (EGRA) + pre-literacy tool **Scale:** scale scores with a mean of 500 and a standard deviation of 100.

Secondary outcome(s)	variable(s)	Specific sub-domains of Arabic literacy attainment
	measure(s) (instrument, scale, source)	**Source:** EGRA Grade 2 assessment with the addition of a set of preliteracy items **Instrument:** Early Grade Reading Assessment (EGRA) + pre-literacy tool **Sub-domains:** Letter sound identification; speed and accuracy of word decoding; reading comprehension
Baseline for primary outcome	**variable**	Arabic literacy attainment
	measure (instrument, scale, source)	**Source:** EGRA Grade 2 assessment with the addition of a set of preliteracy items **Instrument:** Early Grade Reading Assessment (EGRA) + pre-literacy tool **Scale:** scale scores with a mean of 500 and a standard deviation of 100.
Baseline for secondary outcome	**variable**	Specific sub-domains of Arabic literacy attainment
	measure (instrument, scale, source)	**Source:** EGRA Grade 2 assessment with the addition of a set of preliteracy items **Instrument:** Early Grade Reading Assessment (EGRA) + pre-literacy tool **Sub-domains:** Letter sound identification; speed and accuracy of word decoding; reading comprehension

Randomisation

Every school recruited for the pilot will be randomly allocated to one of the pilot arms (Whole Class [W/C], Literacy Catch-up [C/U] or control group). Schools will be randomly allocated to groups at the beginning of the 2021/22 academic year.

Schools will be stratified by region and urban/rural classification prior to randomisation to ensure balance across pilot arms across strata after randomisation. Jordan has three geographical regions (i.e. middle, south, and north). Amman, the country's capital and the biggest city in Jordan, is located in the middle region. To equally represent schools in the middle region but not in Amman, we will divide the middle region into two geographical regions: Amman and the remainder of the middle region. Furthermore, due to logistical reasons, near north and near south regions will be formed. The near north region is formed of Jerash and Ajloun governorates and near south region is formed of Karak governorate only. Therefore, we will have four geographical regions (i.e. Amman, middle excluding Amman, near south and near north). Given that we have four geographical regions, and a school could be either in a rural or urban area, we will have eight strata.[23]

[23] We will have the following strata: Amman urban, Amman rural, Middle excluding Amman urban, Middle excluding Amman rural, North urban, North rural, South urban and South rural.

Randomisation will be carried out by an analyst at NatCen in September 2021. Randomisation will be undertaken in Stata and both the 'do' and 'log' files saved as a record of the randomisation process.

Recruitment

All primary schools in Jordan will be eligible for the pilot as long as they satisfy the following conditions:

- having students in Grades 1, 2 and 3
- being a single shift school[24]
- not being part of any other literacy interventions, other than the Reading and Writing Project which has been in grades 1-3 in all the Ministry of Education school in Jordan since 2015.
- having a resource room teacher
- not delivering blended teaching[25]
- not being in Syrian refugee camps[26]

Only primary schools from middle, near north and near south[27] will be eligible for the pilot evaluation. QRF will be responsible for recruiting schools satisfying the eligibility criteria listed above in September 2021. They will also liaise with the MoE, who will provide written permission for the intervention and evaluation to take place, as well as for use of the EGRA. QRF will explain to prospective schools what participation in the evaluation will involve during the recruitment process.

In total, QRF will recruit 24 primary schools. An equal number of recruited schools will be randomly assigned to each pilot arm (completely randomised). This will mean 8 primary schools in the W/C intervention, 8 in the C/U intervention, and 8 in the (shared) control group.

Participants

All Grade 1 pupils in classes assigned to the W/C model will be eligible for the pilot evaluation. We expect, on average, 22.6 Grade 1 pupils per class.[28] With an expected pupil attrition rate of 10%, we assume, on average, 20.4 Grade 1 pupils per class to be eligible. We will randomly choose 50% of eligible Grade 1 pupils from each class to take part in EGRA testing at baseline and test the same pupils again at endline of semester 2.[29] This will give us 10.2 Grade 1 pupils per class on average to take part in testing for the W/C intervention.

---

[24] In Jordan, schools may operate on one shift or two shifts. Schools operating on two shifts (morning and afternoon shift) have different group pupils in morning and afternoon shift, while schools operating on one shift have one group of pupils during whole day.

[25] In September 2021, the Ministry of Education of Jordan has decided to implement blended teaching in some schools due to the COVID-19 pandemic. Schools who are required to implement blended teaching will have 2/3 days face-to-face teaching and 2/3 days pre-recorded online teaching in a week. This form of blended teaching would have negative effect on the implementation of the intervention and would not reflect business-as-usual in the control schools. Therefore, the schools that are part of the blended teaching will not be eligible to the pilot trial.

[26] When schools in the Syrian refugee camps and single shift schools were excluded from the list of eligible schools, we would be excluding all schools in the refugee camps and the Syrian evening schools from the list of eligible schools. Our final list of eligible schools would include schools that have either Syrian refugee students who are integrated to the Jordanian Educational System or no Syrian refugee students. The term "integrated schools" is used by the MoE and indicates schools where Syrian refugee children and Jordanian children are taught together in the same classrooms

[27] The North and South regions were narrowed to near north and near south for logistical reasons.

[28] This figure has been extracted from Data from the School Management Information System (MIS) in March 2020.

[29] A limited sample of pupils from each class was chosen to reduce burden on schools and pupils; also looking to avoid unnecessary additional assessment costs.

The 20% of lowest achieving students in a class in either the second semester of Grade 1, or in the first semester of Grades 2 and 3 are eligible for the C/U implementation model.[30] Teachers in the C/U and control groups will administer the RAMP coarse-grained reading assessment to identify pupils eligible for C/U. We expect, on average, 4.8 pupils across grades.[31] With an expected pupil attrition of 10%, we assume, on average, 4.3 pupils per class across grades to be eligible for the trial. From the sample of trial participants, we will randomly choose 50% of the eligible pupils from each class to take part in testing at the beginning and the same pupils will again be tested at the end of the trial.[32] This will give us 2.15 pupils per class across grades to take part in testing for the C/U model.

The pilot evaluation includes a total of 24 primary schools (8 in W/C, 8 in C/U and 8 in the control group). We also expect to have 2.2 classes for each grade per school.[33] Based on these figures, we expect, on average, 180 pupils in W/C, 114 in C/U and 294 in the control group.

The following procedure will be followed for the recruitment of pupils for testing. One enumerator will be assigned to one school. Upon their arrival at the school, the enumerator will ask for the list of students. For the C/U model, the bottom 20% of the students of the class will be identified by the teacher. 50% of those students will be selected at random for EGRA testing. For the W/C model, 50% of the class will be selected at random for testing. For randomly selecting pupils for testing, the students will be numbered and then will be chosen using a random number generator. In the case of absences on the testing day, the enumerator will select the next student on the list present on the day of testing.

Detecting unforeseen problems

Since this is a pilot study, it is not powered to estimate the impact of LRF. (See Appendix A for calculations to illustrate.) The primary purpose of this study is to identify unforeseen problems in the interventions, implementation, and how they are evaluated so that we can best prepare for the full trial. Problems can be identified in a variety of ways through all the evidence that is available to us, such as continuous feedback on pilot study process, IPE, and quantitative analyses of item responses.

It is possible to calculate how rare a problem we are likely to be able to observe at least once in our sample (Viechtbauer et al., 2015; Fugard and Potts, 2015). We are not attempting to estimate the true prevalence of any problems – again we would be underpowered to do so with any precision. Rather, the idea is that different problems will occur at different rates in the population of schools and pupils from which we are sampling. The calculation begins by specifying the probability with which we want to be confident that our sample includes a school or data from a pupil demonstrating the problem, which we set to 80%.

Across 24 schools, there is an 80% chance that we can observe at least one school-level problem (e.g. implementing the intervention or withdrawing from the study), which occurs at a rate of 6.4% or more in the population from which schools were sampled. At pupil level, difficulties (e.g. intervention acceptability or measurement issues) within each arm are likely to be observable if they occur at a rate of 1.4% or more in the population of pupils sampled. Note how we are likely to become aware of rarer problems at the pupil level than at the school level. This is simply because there are more pupils than schools, so there is more of a chance of a rarer problem to occur. Analogously, the more times you roll a pair of six-sided dice, the more likely it is that you will observe two sixes at least once.

[30] The rationale for selecting students in the second term of Year 1 is that it is felt by QRF that the first semester of Year 1 is too early to identify struggling readers. The rationale for selecting students in the first term of Years 2 and 3 is simply project planning.
[31] This figure has been extracted from Data from the School Management Information System (MIS) in March 2020.
[32] As with the W/C model, a limited sample of pupils from each class was chosen to reduce burden on schools and pupils as well as to avoid unnecessary additional assessment costs. In order to select the pupils, enumerators assign each pupil a number and a number generator selects the pupils at random.
[33] This figure has been extracted from Data from the School Management Information System (MIS) in March 2020.

Baseline measures

To accurately measure the impact of LRF, the Early Grade Reading Assessment (EGRA) must be able to detect the full range of learner proficiencies. The pilot study combines a set of pre-literacy items (see Appendix B) and the Jordanian EGRA at both baseline and endline (referred to as the EGRA+pre-lit from hereon in). Baseline EGRA+pre-lit will help explain variation in outcomes due to prior ability, increasing the power of the pilot to speak to 'evidence of promise', inform power calculations and possible baseline testing as part of the efficacy trial.

Outhred and Rolleston's (2014) review of EGRA in low- and middle-income countries (LMICs) for the UK Foreign, Commonwealth and Development Office found EGRA's theoretical foundations are sound.[34] However, substantial missing data from floor effects make off-the-shelf EGRA tools sub-optimal for impact evaluations.

Development of the Pre-literacy items

In order to ensure the literacy measure is robust and fit-for-purpose we conducted a thorough desk-based review of the Jordanian EGRA. The review considered the suitability of the EGRA for students participating in LRF. On the basis of the desk review, the tool was refined, and additional pre-literacy items were added to ensure it will robustly measure learners' proficiency.

The administration of the EGRA+prelit tool retained the same protocols as the EGRA assessment previously administered in Jordan. The assessment begins with the pre-literacy items and if learners are able to answer the easiest items correctly, they move through the pre-literacy items onto more difficult items. This process leads seamlessly into the EGRA Jordan items. The same protocols are also used regarding stop rules (when a learner has reached the section of the test that exceeds their ability level, the learner is moved on to the next item or the test administration is ended).

The EGRA+pre-lit was pre-tested with 57 Grade 1 learners from 5 schools in August 2021. We conducted psychometric analysis on the pre-test data and found the adapted tool:

- Is well targeted to the proficiency levels of the learners, while leaving room for growth at endline.
- Has items that predominantly met international item discrimination standards. Where items did not meet discrimination standards, these items were at the easiest or most difficult level of the scale and were the result of a lack of data (due to the small sample) and standards would be achieved at pilot.
- Achieved outstanding reliability, at between .941 (Raju's Beta) through to .959 (Guttman's L2).[35]

The tool remained unchanged for use in the pilot study on this basis.

EGRA+pre-lit will be administered at both baseline and endline. Baseline primary and secondary measures will be identical to the endline primary and secondary measures, respectively. After administering EGRA+pre-lit, Arabic literacy attainment score with a mean of 500 and a standard deviation of 100 calculated from the Item Response Theory (IRT) will be used as the baseline measure for the primary outcome. The baseline measures of the secondary outcomes will include specific sub-domains from the EGRA+pre-lit, which are letter sound identification, speed and accuracy of word decoding, and reading comprehension. The formation of these measures is explained in detail in the outcome measures section.

[34] Outhred and Rolleston (2014) *Review of in low-and middle-income countries (LMICs) for the Department of International Development*, Sierra Leone. Unpublished report. Oxford Policy Management, Oxford, UK.
[35] Reliability refers the overall consistency of a measure. Reliability is one of the key aspects of measurement quality (along with validity and fairness). For population level estimates, a reliability score of .95 is concerned by experts (Cicchetta &Sparrow, 1981; Fleiss, 1981, Landis and Koch, 1977 and Regier et al, 2012) as ranging between excellent and almost perfect.

Outcome measures

Analysis of the EGRA+pre-lit tool outcome data undertaken as part of the pilot evaluation will inform the final choice of EGRA question banks and questions for the efficacy trial. The primary outcome measure is a single learning metric for reading, which will be obtained by administering the EGRA+pre-lit explained in detail in the baseline measures section. We will use Item Response Theory (IRT) to produce an interval-scaled measure that takes the difficulty level of individual items into account.[36]

A key principle underlying the IRT model is that of seeking to measure a latent unidimensional trait. This simply means an underlying construct (i.e. one that cannot be measured directly) that can be thought of in terms of more or less. In this impact evaluation, the latent unidimensional traits that we are seeking to measure using the pupil tests are abilities/performance in Arabic literacy. The Rasch model is the simplest latent trait model. It is based on a mathematical model of a person's (in our case, a pupil's) response to an item. The latent trait is conceived as a single dimension along which items can be located in terms of their difficulty and persons can be located in terms of their ability. The Rasch model is probabilistic and is a special case of an item-response theory (IRT) model.[37] The model estimates the probability of answering the item correctly as a logistic function of the difference between the person's ability and the item's difficulty. This can be seen in the formula below, which shows the form of the Rasch model for dichotomous responses:

$$\mathrm{P}\{x_{vi} = 1 \mid \beta_v, \delta_i\} = \frac{e^{(\beta_v - \delta_i)}}{1 + e^{(\beta_v - \delta_i)}}$$

Where $P\{x_{vi} = 1 \mid \beta_v, \delta_i\}$ is the probability of a correct response for a particular person and item combination, β_v and δ_i are, respectively, the ability of pupil v and difficulty of item i.[38] Items 1-41 will be administered, including print media familiarity items, vocabulary, letter identification, letter sounds, syllable sounds, high frequency word identification, reading and reading comprehension (see Appendix B). All administered items will be included in the IRT model, which will produce a scale score per participant. Therefore, the overall Arabic literacy attainment score calculated in the IRT model above will come from β_v rescaled to have a mean of 500 and a standard deviation of 100.

The secondary outcomes are the specific sub-domains of Arabic literacy attainment. These sub-domains are letter sound identification, speed and accuracy of word decoding, and reading comprehension. We will also produce separate metrics for these sub-domains from the EGRA+pre-lit. Letter sound identification will be computed into letter sounds per minute from the EGRA letter sound identification word task. Speed and accuracy of word decoding will be computed into words per minute from the EGRA word tasks. Reading comprehension will be the five reading comprehension items at the end of the EGRA tool.

Endline data will be collected from Grade 2 and 3 pupils in March 2022 and from Grade 1 pupils in May 2022.[39]

Analysis

Primary outcome pilot analysis

Even though the focus of the pilot is not on estimating an effect size, we will develop and quality assure the data pipeline and statistical framework. Estimates of efficacy will be highly imprecise in these pilot analyses. Primary analyses will estimate the difference between intervention and control on the EGRA+pre-lit measure. In line with the EEF analysis guidance,[40] the primary analyses will

[36] More details on IRT can be found in the Appendix C.
[37] More details on the Rasch model can be found in the Appendix C.
[38] If someone's ability is equal to an item's difficulty, then the probability that they obtain the correct answer is 0.5. If the person's ability is greater than item difficulty (or, equivalently, if the item is easier than participant ability), then the probability of obtaining the correct answer is above 0.5.
[39] Testing of grades 2 and 3 will take place in March 2022 due to the early closure of schools in December 2021. This will enable the remainder of the LRF content to be taught once schools return in February 2022;
[40] EEF (2018) Statistical analysis guidance for EEF evaluations.

use the EGRA+pre-lit (total) and follow an intention-to-treat (ITT) approach. The analyses will use a three-level multilevel model with treatment assignment at the school level (level 3). This model will include school and class-level random effects and will account for the baseline EGRA+pre-lit.

A separate model will be formed for each intervention type (The W/C approach and the C/U approach). The basic form of the model is:

$$Outcome_{ijk} = \beta_0 + \beta_1 Baseline_{ijk} + \beta_2 Intervention_k + \beta_3 Strata_k + u_{jk} + e_{ijk}$$

Where pupils (i) are clustered in classes (j) within schools (k). The intervention effect is estimated by β_2, β_3 represents strata fixed effects for the schools (i.e., their geographical location), β_0 is the intercept, β_1 is the slope for baseline scores, u_{jk} a classroom-level random intercept, and e_{ijk} the residual term. Note that in these pilot analyses there is no random intercept by school since all school level variance will be explained by intervention condition and strata, so we know that a school-level random effect would have zero variance. In line with the EEF analysis guidance, other additional covariates will not be considered. The analysis will be carried out using R statistical software (R Core Team, 2021) and code will be published.

The difference between the intervention and control groups at endline will be expressed as a standardized effects size using Hedges' *g* with 95% confidence intervals. Following EEF guidelines, the numerator will come from the unstandardised effect estimate given by β_2 in the multilevel model specified above, which is adjusted for baseline score and strata. The denominator is the unconditional pooled standard deviation in the primary outcome at endline. The formula is provided below:

$$g = \frac{\beta_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Where n_1 and n_2 are the number of pupils in both groups and s_1^2 and s_2^2 are the within-group variances in outcomes at endline.

The primary analysis will also include detailed descriptive analysis: histograms, means, quartiles, and SDs, for all measures, groups, and time points.

We will also report school and class-level Intracluster Correlation Coefficients (ICCs[41]) (alongside 95% confidence intervals) in analyses. Again, since the study has a small sample size, these estimates will be imprecise. We will use the following model with a random intercept by school, v_k, and no fixed effects for strata so it is arithmetically possible that the variance of $v_k > 0$:

$$Outcome_{ijk} = \beta_0 + \beta_1 Baseline_{ijk} + \beta_2 Intervention_k + u_{jk} + v_k + e_{ijk}$$

Secondary outcome pilot analyses

EGRA Sub-domains

For all defined secondary outcomes (i.e. letter sound identification; speed and accuracy of word decoding; reading comprehension), we will follow an intention-to-treat approach using a basic model for each intervention type (the W/C approach and the C/U approach) similar to that of the primary outcome analysis:

$$Outcome_{ijk} = \beta_0 + \beta_1 Baseline_{ijk} + \beta_2 Intervention_k + \beta_3 Strata_k + u_{jk} + e_{ijk}$$

[41] The ICC measures similarity between units in the same cluster; in this case, pupils within the same classroom. Units within the same cluster may exhibit similarities due to being exposed to similar environmental characteristics. This must be accounted for when conducting sample size calculations, since similarity between units reduces the amount of unique information each new observation contributes to the sample.

Where pupils (i) are clustered in classes (j) within schools (k). The intervention effect is estimated by β_2, β_3 represents strata fixed effects for the schools (i.e., their geographical location), u_{jk} a classroom-level random intercept, and e_{ijk} the residual term. For these measures we will also be using standardized scores (z-scores), reporting confidence intervals at 95% level, and the effect size using Hedges' formula previously described.

Implementation and Process evaluation design - pilot

An implementation and process evaluation (IPE) will be carried out to address the evaluation domains set out in Table 5. Our IPE domains of interest are informed by the EEF's framework for implementation and process evaluations.

Table 5: Connecting research domains with IPE data collection

Domain	Research activity			
	Qualitative FGDs and KIIs	**Classroom observation**	**Teacher survey**	**Training attendance data**
Fidelity				
Dosage			♦	♦
Quality	♦	♦	♦	
Adaptation	♦		♦	
Reach			♦	♦
Responsiveness	♦	♦	♦	
Usual practice	♦			

There are several different strands of work within the IPE:

- Analysis of training attendance data
- Focus group discussions (FGDs) with teachers, parents and pupils
- Key Informant Interviews (KIIs) with QRTA stakeholders and school supervisors
- Classroom observations
- A teacher survey
- A pupil survey

Table 5 above, shows how the different IPE research methods map onto the research domains of interest. Each is described in more detail below.

We will gather information about **the number of participants attending LRF training** and **coaching sessions** from attendance registers collected during training. This information will be supplied by QRTA and offer an indication of the reach and take-up of training.

Focus Group Discussions will be carried out with groups of teachers, parents and pupils. FGDs will help explore perspectives around the intervention and its context and deepen and triangulate findings. Each FGD with parents and teachers will last around 45-60 minutes, while FGDs with pupils will last around 20 minutes.

The aims of the Student FGDs include: finding out whether pupils use the practice book, how they use it, what they like or find difficult about it, whether they feel they are making progress with reading or not, and why they think this is.

The aims of the Teacher FGDs include: understanding perceptions around the challenges and opportunities presented by the LRF models, opinions about the suitability of the LRF approach and resources, perceptions of pupil use of the practice book, any adaptations they have made to the LRF model and any perceived changes in learning outcomes or behaviours among students due to by the intervention.

Finally, the aims of the parent FGDs include: assessing the level of parental engagement with the practice book and pupil literacy progress, collaboration with the school, levels of resources or support to literacy provided at home, challenges and opportunities associated with sending the practice book home, and any perceived changes in learning behaviours or outcomes observed by parents.

We will carry out a small number of **Key Informant Interviews (KIIs)** with all QRTA coaches. One group interview will take place in January 2022 followed by a second at the end of semester 2 (in May 2022). The first group interview will focus on the C/U model and the second on W/C. It is expected that QRTA trainers/ coaches will be able to observe challenges, strengths and opportunities across multiple schools. School supervisors and principals will be expected to discuss the same but at their specific school.

Classroom observations will assess pupil engagement, use of the student practice book and learning processes in intervention schools. Observations will be conducted two times each semester for each class and will be guided by a Fidelity of Implementation and Pupil Engagement tool.

A **survey of teachers** implementing LRF will be undertaken alongside the second classroom observations. This will collect data on classroom (W/C) and resource room (C/U) teachers' perspectives on training and coaching, delivery to pupils, dosage, and resources needed to implement the intervention.

We will also carry out a **survey of pupils** in both the intervention and control schools. The survey will be designed for young children and will explore pupils' access to (physical and digital) story books and reading engagement at home (both with family members and alone).

IPE numbers and sampling

Numbers and sampling for FGD

Sampling for the focus groups will be carried out with attention to equity considerations, and inclusion of a diversity of viewpoints. Specifically, in our sampling we will ensure representation of the three main geographic areas (middle – including Amman and non-urban areas outside of the capital, near north and near south of Jordan), as well as rural and urban schools. As shown in Table 6, our sampling will also achieve coverage of both LRF implementation models, all three Grades, and control schools (in order to assess usual practice).

We will also look to ensure that focus groups target differentials in terms of schooling, separating out by Grade. The same would also be done with any single-sex schools (separating boys' and girls' schools accordingly).

With parents, we will aim to ensure adequate representation in terms of implementation model, Grade, geography and child's gender.

The schools selected for teacher FGDs will form the sample frame for the parent and pupil FGDs – which will allow for triangulation of views across the stakeholder groups.

Table 6: Numbers and distribution - FGDs

	Grade	No. W/C	No. C/U	No. control	Total
Teacher FGD	1	3 FGDs, across 6 schools (semester 2)	1 FGDs. across ~2schools (semester 2)	1 FGD, across 2 schools	**5**
	2 and 3		2 FGD, across all schools (semester 1)	1 FGD, across 2 schools	**2**
Parent FGD	1	3 FGDs, across 3 schools (semester 2)	1 FGD in 1 school (semester 2)		**4**
	2 and 3		2 FGD in 2 schools (semester 1)		**2**
Pupil FGD	1	1 FGD in 1 school (semester 2)	1 FGD in 1 school (semester 2)		**2**
	2 and 3		1 FGD in 1 school (semester 1)		**1**

The purpose of the FGDs, namely, to understand use of and acceptability of the LRF model, explore the efficacy of the approach and materials, as well as any perceived impacts for pupils, means they will be required to take place towards the end of the implementation period in each semester.

By virtue of the design we intend to undertake the seven W/C FGDs (three with teachers, three with parents and one with students) towards the end of the semester 2 intervention period, i.e. in April 2022. One FGD with control group teachers, and one with control group parents will take place in the same time period.

The C/U model FGDs will be split between semester 2 (April 2022) and semester 1 (December 2021) according to Grade. Specifically, resource room teachers and parents of children in Grades 2/3, and Grade 2/3 pupils will be invited to FGDs in semester 1, and those in Grade 1 to FGDs in semester 2 (see Table 6).

Principals will be asked to help recruit parents to the FGDs from one school in the middle, near north and near south Jordan geographical areas for each implementation model (i.e. 3 C/U and 3 W/C FGDs).[42] This will be an open invitation to all parents of pupils participating in LRF.

We will take a census approach, inviting all teachers in each implementation model to the FGDs. The teacher FGDs will be held online or in a central location in Amman.

For both the parent and teacher FDGs, verbal consent will be given by participants at the start of each session. Those who no longer wish to take part or do not want to be recorded will have the ability to leave at this point. However, the option to not record the session will also be given.

For the control group FGDs, teachers will be recruited from two schools in the middle region and two schools in the near north region, with the discussion taking place within one school in each region (dependent on their ability to host the event). Parents will be recruited from one school in each of the same regions, with the FGD being held at the same two schools as the teacher FGDs. We will target the parents/ carers of children who are participating in the LRF program.

Numbers and sampling for classroom observations and teacher survey

Classroom observations will be conducted in all intervention schools (16 of the 24 schools taking part in the pilot), negating the need for sampling at the school level.

Classroom observations will be conducted with schools in both the W/C and C/U implementation models (see Table 7). They will be conducted twice during each semester – W/C, and C/U for Grade

[42] Small numbers of pupils in the C/U model mean that the parent FGDs in C/U schools will include just 3-5 parents.

1 in semester 2 (in February and April), C/U for pupils in Grades 2 and 3 in semester 1 (in October and December).

Table 7: Numbers and distribution - classroom observations and teacher survey

	Grade	No. W/C	No. C/U	Total
Classroom observations with FOI	1	22 in 8 schools	4 in 3 schools	26 in 11 schools
	2		3 in 3 schools	3 in 3 schools
	3		3 in 2 schools	3 in 2 schools

The teacher survey will be conducted alongside the second classroom observation (see timings above). Classroom (W/C) and resource room (C/U) teachers will be asked to complete a short survey about the experiences and perceptions of the intervention. It is likely to include:

- Perceptions and usefulness of training
- Whether coaching sessions used and how many
- Perceptions of intervention from the teacher and school perspectives
- Whether delivery was as intended (i.e. whether dosage was as expected, or sessions were missed)

The survey will also collect data on costs of the intervention (particularly in terms of teacher time). This cost data will not be reported at the pilot stage but will be reviewed in order to identify any potential issues ahead of the efficacy trial.

The teacher survey will be prepared in October for sign-off in November 2021.

Pupil Survey

In addition to the above, the evaluation includes a survey with all pupils in both intervention and control schools (i.e. all pupils taking part in the pilot). The pupil survey will be administered alongside the EGRA at endline in semester 1 and at baseline and endline in semester 2 (as set out in Table 8) to provide process measures for the LRF evaluation.

Whilst the pilot timeline did not allow for a baseline pupil survey, semester 2 offers an opportunity to test before and after survey design ahead of any efficacy trial in 2022-23.

Table 8: Pupil survey – data collection

	2021-22 Academic Year		
Implementation model (Grade)	Semester 1 (Dec 21)	Semester 2 (Feb 22)	Semester 2 (May 22)
C/U (Grade 1)		**Baseline**	**Endline**
C/U (Grade 2, 3)	**Endline**		
W/C (Grade 1)		**Baseline**	**Endline**

Based on the draft LRF logic models and theory about literacy acquisition, we anticipate that children's reading habits may change before final impacts on Oral Reading Fluency are observed. We therefore intend to explore potential effects of LRF on the incidence of reading using descriptive analyses. The ability to measure whether children have books at home, and whether and how often they read at home will help improve understanding of the LRF logic model. Therefore, collecting this information provides a valuable addition to the evidence-base around LRF.

The survey also provides a valuable opportunity to collect information about reading habits at home that is not currently collected in the Jordanian context. This may bring a wider benefit in terms of

understanding the context in which LRF is implemented. (Proposed questions are included in Table 9).

The survey will take the form of a short (5 minute) questionnaire which will be conducted immediately after the EGRA assessment (administered by the same enumerators). The instrument will be finalised by the end November 2021, for endline testing in December.

Table 9: Pupil survey – proposed questions

Factor	Questions
Access	Do you have internet at home? Do you have a computer or smart phone at home? Are you allowed to use it with your family members?
Resources	Do you have story books at home? Can you borrow story books from school or library?
Behaviours	Do you borrow story books from school or library? Do you use the internet to read story books? Do you read story books at home? Alone? With a parent or sibling? How often do you read story books at home?
Let's Read Fluently (Endline testing only)	Do you take the practice book home? Do you read the practice book at home? How do you feel about reading from the practice book at home/school?

IPE analysis

The consortium will use NatCen's Framework approach, adapted for bilingual working[43], to manage data from observations, focus groups and key informant interviews. The framework approach is a type of thematic analysis which evidences the relationship between themes and anonymised cases.

Using themes covered in discussion guides and any other themes which emerge from the data, we will assemble a matrix in which each row represents an FGD or KII and each column a theme/ sub-themes. We will then summarise the qualitative data in the matrix, including illustrative verbatim quotes where appropriate. Once all data has been coded in this matrix, we will move onto analysis. This will involve a phase of 'detection', including studying what participants say about a particular phenomenon, listing these and then sorting them thematically. Once we had identified different themes in the data, we created higher-level categories that worked as meaningful conceptual groupings for participants' views and experiences within and across schools.

Summaries will be produced in English and shared with the wider team for review. Core members will then come together for an analytical planning meeting, where key themes, patterns and issues emerging from the qualitative data are discussed, and across-team verification of findings established. Findings will be reviewed in detail and mapped against the pilot RQs/LMs in preparation for reporting.

Data from the teacher and pupil surveys, collected using Alchemer survey software, will be exported to SPSS for descriptive analysis.

[43] NatCen will receive translated notes of the IPE activities from Integrated. FGDs and Interview notes will then be applied to frameworks.

Ethics and registration

Ethical approval

This project will be submitted to NatCen's Research Ethics Committee (REC), made up of senior NatCen staff and external experts where appropriate, for scrutiny in advance of data collection. NatCen's ethics procedure meets the requirements of the UK Economic and Social Research Council (ESRC) and the UK Government Social Research (GSR) Professional Guidance. The evaluation will be undertaken according to NatCen procedures designed to ensure our research is conducted in line with five principles outlined by the GSR guidance:

- Sound application and conduct of social research methods and appropriate dissemination and utilisation of the findings.
- Participation based on valid informed consent.
- Enabling participation.
- Avoidance of personal and social harm.
- We will ensure participants are not identifiable in the outputs.

Registration

The trial was registered on the Open Science Foundation on 26/11/2021 (URL). The trial registry will be updated with outcomes at the end of the project.

Data protection

We recognise the need for data security and operate to extremely high standards of confidentiality and anonymity.

NatCen is fully accredited to ISO 27001 and subject to annual external audits of procedures to maintain accreditation. We also hold Cyber Essentials Plus Certification. We were previously registered under the Data Protection Act and are now fully GDPR compliant.

This project will have its own data security plan. For transferring data securely between partners, we will use NatCen's Secure File Transfer Protocol (SFTP) server. Those requiring access will be given password-controlled access to the server.

EGRA will be undertaken by Integrated, with pupils assigned a unique identifier. Test results will be submitted to EGRA's Prodigy tool and a pseudonymised dataset transferred to NatCen and OM using the SFTP server. EGRA data will be stored with back-end provider Prodigy and sent directly to authorised Integrated personnel, and Integrated will store the IPE data within a dedicated drive with authorised personnel access only.

All information stored, processed and/or transmitted at Integrated is protected in a manner consistent with contractual and legal restrictions proportionate to the level of sensitivity, value and risk of that information to Integrated, its partners and/or clients. Sensitive information is secured against disclosure, modification, and access by unauthorised individuals while both holding and transferring it. Personnel with authorised access are obliged to maintain data confidentiality through provisions including contractual legally binding provisions in employment contracts, as well as a signed code of conduct for all employees.

Data shared with NatCen will be stored on NatCen's secure network, with access to the project folder restricted to authorised personnel only. The data will be backed up and NatCen will carry out regular testing to ensure this process is effective.

To ensure integrity and confidentiality, all data and files held by NatCen are classified to one of three different levels, with each level having its own specific requirements for how the data are stored, handled, and transmitted. Any data containing personal details is deemed to be 'Respondent Confidential'. For such data, protection against the disclosure of respondent identities – whether by

direct association with a name or address or by indirectly associating information disclosed – is built into all stages of the process.

OM will need access to EGRA data to conduct analyses. NatCen will transfer EGRA data to OM using the SFTP server. The data will be stored on OM's secure servers and only named team members will have access to this data.

The partnership will agree a data retention period with QRF. Once this period has expired and data has been archived with EEF, all partners will securely erase project data (with explicit permission from QRF) to US Department of Defense 7 standards.

Personnel

The team includes staff from NatCen who will lead the partnership, as well as Integrated, Oxford MeasurEd (OM) and School-to-School (STS).

NatCen will be the lead partner and accountable to QRF. NatCen will lead on project coordination and management, evaluation design, analysis and reporting.

Integrated will be a subcontracted partner. They will lead EGRA testing, IPE data collection and contribute to evaluation design, analysis and reporting.

Oxford MeasurEd will be a subcontracted partner. They will develop and refine the EGRA outcome measure and contribute to evaluation design, analysis and reporting.

School-to-School will be a subcontracted partner. They will bring expertise in child literacy and act as a 'critical friend' throughout the evaluation, contributing to evaluation design, analysis and reporting stages.

Table 10: Evaluation team

Name	Project role	Organisational role
Sashka Dimova	Principal Investigator and strategic lead	Research Director, Evaluation, NatCen
Hannah Woodbridge	Day-to-day project manager	Senior Researcher, Children and Families, NatCen
Julia Ruddick-Trentmann	Project management support	Researcher, Children and Families, NatCen
Andi Fugard	Impact Evaluation Lead	Research Director, Evaluation, NatCen
Enes Duysak	Impact Evaluation Support	Senior Researcher, Evaluation, NatCen
Nedjma Koval	Data collection lead	Integrated
Leen Al Refai	IPE Lead, project management	Integrated
Zaid Qiblawi	EGRA, IPE Data Collection Support	Integrated
Rachel Outhred	Outcome measure development lead	Oxford MeasurEd
Lydia Marshall	Support for outcome measure development, evaluation design, analysis and reporting	Oxford MeasurEd
Kaydi-Ann Newsome	Outcome measure development support	Oxford MeasurEd
Daniel Phillips	Quality assurance	Group Head, Evaluation, NatCen

Table 11: Delivery team

Name	Project role	Role and team
Lubna Dirini	Manager, Lead trainer	QRTA
Amani Alker	Coach	QRTA
Mohammad Salameh	Coach	QRTA
Robert Palmer	Quality Assurance	QRF, Executive Director, Research and Program Development
Maysoon Masoud	Project lead	QRF, Research & Program Development Manager
Rami Asses	Administrative & financial management	QRF, Project Management Specialist
Emilee Rauschenberger	Quality Assurance	QRF, Research Manager
Haneen Alabed	Material development lead and trainer of trainers	QRF, Research & Program Development associate Manager
Hanif Pabani	Quality Assurance	Advisem, Impact Evaluation consultant
Louis-Pierre Michaud	Quality Assurance	Advisem, Impact Evaluation consultant
Julie Helson	Quality Assurance	Advisem, Impact Evaluation consultant

Risks

We take a proactive approach to the management of risks, considering the likelihood and potential impact of key risk factors, as well as mitigations and contingencies (see Table 12).

Table 12: Key risks

Potential Risk	Likelihood	Impact	Mitigation and contingencies
Slippage in research timetable	H	M	Detailed project timetables developed and regularly reviewed to identify problems early and focus activity on addressing them. There are options to expand the project team to catch up and to reschedule work if necessary.
Partnership does not work together successfully	M/H	H	We will work closely with QRF to assess the degree to which LRF implementation, evaluation delivery and data collection are affected. We will work flexibly with QRF to adjust our analysis, approach, and timetable accordingly to ensure impacts can be measured in accordance with implementation. The corresponding caveats for interpretation, if present, will be added.
Major disruption to evaluation fieldwork resulting from public health emergency, extreme weather, or other national emergencies	M/H	M	We will work flexibly with QRF and participants to ensure data is collected safely. We will draw on our extensive experience of delivering research during major disruptions, such as the pandemic, which has involved collaborating with participants remotely, offering flexibility with data collection mode (online and phone), and being receptive to last-minute cancellations.
Fieldwork staff illness/ unavailability/ turnover due to reasons including Covid-19	M	M	We have a large team of researchers and fieldwork assistants to draw on should availability unexpectedly change (changes will be made in agreement from QRF). We will take Covid-19 precautions through the provision of PPE and institute a containment and testing policy to our fieldwork staff.
Difficulties engaging participants leading to failure to meet target recruitment rate	M	H	We will discuss strategies with QRF and work closely to complete recruitment on time during the inception stage. We recommend an early decision be made on whether to progress to efficacy trial so recruitment begins early.
School and/or pupil level attrition	L	M/H	This will be addressed by setting out the requirements for the trial at the outset and providing schools with clear instructions at the start of the project on what needs to be done and when.

			Our experienced field team will ensure as many children as possible are tested in each school at baseline and endline. Power calculations will account for expected attrition.
Outcome measure having floor effects	M	M	Our experienced team will ensure outcome measures are robust and fit-for-purpose through a desk-based review. This includes pre-testing the EGRA tool, informing the instrument development process to enable high-quality, well-targeted data to be collected during the pilot and helping mitigate the danger of multiple pupils receiving zero scores due to likely low learning level following school closures.

Timeline

Table 13: Timeline - Pilot

	School Calendar	Proposed dates	Notes
	Schools open	15-Aug	Catch-up program
	Official start of the new school year	15-Sep	
	Semester 1	15-Sep-21 - 6-Jan-22	
	Semester 2	6-Feb - 15-Jun-22	
	Key milestones	**Indicative deadline**	**Notes**
July - Oct 2021	**Pilot Phase Set up**		
	Logic model workshop	27-Jul-21	
	Ethics approval for EGRA pre-test	11-Aug-21	
	EGRA pre-testing completed	26-Aug-21	
	School sample selection criteria identified	09-Sep-21	
	Schools recruited	13-Sep-21	
	MoE approval to use tools	25-Sep-21	
	Logic model agreed	30-Sep-21	
	Ethics approval complete for all pilot components	15-Oct-21	
	Pilot Catch-up Model Implementation G2, G3/ semester 1		
Sept - Dec 2021	Teacher training delivered	22-Sept - 02-Oct-21	
	RAMP diagnostic testing	22-Sept – 26-Sep-21	
	EGRA baseline testing	27-Sep - 30-Sep-21	
	School resources delivered	27-Sep - 30-Sep-21	
	Coaching delivered	03-Oct – 23-Dec-21	
	Classroom observations (1)	Oct-21	
	Classroom observations (2)	Dec-21	
	FGDs with teachers, parents and pupils	Dec-21	
	EGRA endline testing & pupil survey	26-Dec - 30-Dec-21	

	Pilot Whole classroom & Catch up Model Implementation for G1 /Semester 2		
Jan - May 2022	MoE approvals for training secured	15-Jan-22	
	KIIs with coaches	23-Jan-22	
	Teacher training delivered	02-Feb-22	
	EGRA baseline testing & pupil survey	06-Feb - 10-Feb-22	
	School resources delivered	13-Feb-22	
	Coaching delivered	6-Feb - 26-May-22	
	Intervention delivered (12 weeks)	13-Feb - 19-May-22	Inc. Eid break
	Classroom observations (1)	Feb-22	
	Classroom observations (2)	Apr-22	
	FGDs with teachers, parents and pupils	Apr-22	
	KIIs with coaches	1 May-22	
	EGRA endline testing & pupil survey	15-May - 19-May-22	
	Efficacy trial decision made	May-22	

	Pilot Phase Finalisation	
June – Oct 2022	Analysis workshop	Jun-22
	Draft pilot report delivered	Aug-22
	Efficacy trial protocol prepared	Sept-22
	Final report signed off	Oct-22

Appendix A

Power calculations

For power calculations, we assume 24 primary schools will be recruited and there will be 8 schools in each intervention arm (8 in the W/C intervention, 8 in the C/U intervention and 8 in the control group). We also assume, on average after accounting for 10% pupil attrition, 22.4 pupils from 2.2 classes (Grade 1 only) per recruited school will be randomly selected for testing at baseline and endline for W/C programme. Similarly, on average, 4.3 pupils from 6.6 classes (Grades 1,2, and 3) per recruited school will be randomly selected for testing at baseline and endline for C/U programme.

Power calculations for the primary outcome are estimated using the following assumptions: baseline Arabic literacy attainment measured by the Early Grade Reading Assessment (EGRA) is used as individual-level covariate. The year group specific EGRA is used as the primary outcome measure in both intervention programmes. The correlation between pre-test and post-test is assumed to be zero since the relevant correlations could not be identified from the literature or earlier studies.

The Intracluster Correlation Coefficient (ICC[44]) is assumed to be 0.22 based on the literature on Arabic literacy (QITABI, 2017, p. 43). Class-level ICCs are assumed to be larger than school-level ICCs. Therefore, the school-level ICC is assumed to be 0.1 and class-level ICC is assumed to be 0.12 for this primary outcome.

The calculations were undertaken using 'PowerUp!'[45] and indicate that the pilot trial of each intervention has statistical power of 0.12 to detect an effect of 0.2 standard deviations for the primary analysis for Arabic literacy attainment.[46] Since this is a pilot study, it is underpowered, and so there is a high probability that no statistically significant effect will be found.

Table A1: Power calculations

		Whole Class (W/C)	Catch-up (C/U)
MDES		0.2	0.2
Pre-test/ post-test correlations	level 1 (pupil)	0.00	0.00
	level 2 (class)	0.00	0.00
	level 3 (school)	0.00	0.00
Intracluster correlations (ICCs)	level 2 (class)	0.12	0.12
	level 3 (school)	0.10	0.10
Alpha		0.05	0.05
Power		**0.12**	**0.12**

[44] The ICC measures similarity between units in the same cluster; in this case, pupils within the same classroom. Units within the same cluster may exhibit similarities due to being exposed to similar environmental characteristics. This must be accounted for when conducting sample size calculations, since similarity between units reduces the amount of unique information each new observation contributes to the sample.

[45] Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24–67.

[46] Statistical power indicates the probability that there will be a statistically significant difference between the treatment and control groups given characteristics of a study, e.g., sample size and minimal detectable effect size. 0.8 is commonly used as a benchmark for a statistical power in fully-powered efficacy trials.

One-sided or two-sided?		2	2
Average cluster size		22.4	14.25
Number of schools	Intervention	8	8
	Control	8	8
	Total	16	16
Number of pupils	Intervention	180	114
	Control	180	114
	Total	360	228

Given the low power and the fact that this is a pilot evaluation, we suggest placing most emphasis on identifying unforeseen problems in the interventions, implementation, and how they are evaluated, rather than on estimating effect sizes.

Appendix B

ADAPTED EGRA WITH ADDITIONAL PRELITERACY ITEMS

EGRA+prelit -Reading Assessment Tool

General instructions:

It is important that you create an atmosphere of fun with the child being evaluated by starting with them a simple conversation about topics that interest them (see example below)

Let them feel that this assessment is like a game so they will enjoy it and is not a difficult task.

It is very important that you ONLY read the content of the boxes, aloud clearly and slowly

Good morning. My name is ____ I live in ____. I want to talk to you about myself, I have ..... of children, their age .....; I have at home……. the sports I do……. etc.] 1. Tell me about yourself and your family? [wait for response; If the student is not excited to talk, ask him/her question number 2. If he/she speaks comfortably, move to the verbal consent paragraph]. 2. What game do you like?

• Allow me to tell you why I am with you today. I work for the Ministry of Education, and I try to understand how children learn to read. You have been randomly selected to do this test. • I would love for you to cooperate with me in this process. But if you don't want to share, you can. • We will play a reading game where I will ask you to read some letters, some words and a short story out loud. • I will use this watch to calculate the time you need to read. • This is not an exam, and it has no effect on your school scores. • I will ask you some other questions about your family. • I will not write your name on the test paper. No one will see your answers to them. • Again, you are under no obligation to participate if you don't want to, and if we start and you don't answer a question, that's fine. Do you have a question? Are you ready?

Verbal consent

If you get the child's oral consent, put an (X) in this box Yes .1
If you do not get approval, thank the child and move on to the next child and use the same for

1. Date of Assessment:	Day: _____ Month: _____ Year: _____
2. Governate:	
3. MOE Field Directorate	
4. School Name	

5. National ID for School	
6. Student's Shift	○ One shift ○ Morning Shift ○ Evening Shift
7. Name of Evaluator	
8. Evaluator Code	
9. Grade	○ Second Grade ○ Third Grade
10. Division	
11. Child Number	
12. Child's Date of Birth	Month: _____ Year: _____
13. Child's Gender	○ Girl ○ Boy
14. Exam Start Time:	_____:_____ Choose One time slot: ○ Morning ○ Evening

Section 1 Print awareness	**60 Seconds**
— This is a book. Can you take it from me and put it into your hands? Then I'm going to ask some questions about the book. Ensure the book is in the hands of the child.	
— Let's Begin — With the book in your hands, can you show me the front of the book? [Include here instructions to the enumerator to indicate if the child correctly or incorrectly identified the front of the book]. — Thank you. Now can you open the book to the first page and point to where we can begin reading the story? [Include here instructions to the enumerator to indicate if the child correctly or incorrectly identified (a) the first page of the book and (b) where to begin reading.	Take the book back from the child before moving on to the next section
Section 2 Oral vocabulary	**60 Seconds**
Let's play a few more naming games now. Think about the different things that you can eat. Name as many things that you can eat as you can. Clearly put a tick (/) in the box for each correct word. MULTI-SELECT 01 Word one is correct 02 Word two is correct 03 Word three is correct 04 Word four is correct 05 Word five is correct 06 Word six is correct 07 Word seven is correct 08 Word eight is correct 09 Word nine is correct 10 Word ten is correct 11 Child was unable to say any correct words	After 30 seconds, you will tell the child to 'stop'. ✋ The Early stop rule: If the child hesitates to name things you can eat after 5 seconds, say "thank you" and stop the exercise.

Now I would like you to name as many animals as you can. Clearly put a tick (/) in the box for each correct word. MULTI-SELECT 01 Word one is correct 02 Word two is correct 03 Word three is correct 04 Word four is correct 05 Word five is correct 06 Word six is correct 07 Word seven is correct 08 Word eight is correct 09 Word nine is correct 10 Word ten is correct 11 Child was unable to say any correct words	After 30 seconds, you will tell the child to 'stop'.  <u>The Early stop rule:</u> If the child hesitates to name animals after 5 seconds, say “thank you” and stop the exercise.
Section 3 Recognitise Letter Names	**30 Seconds**
— We will do some alphabet letter games now. Do you see these letters? I would like you to tell me the name of each letter. It’s ok if you don’t know all of them. — Now let's do this exercise: tell me the name of this letter [and point to the first letter]: If a child gets stuck for more than 5 seconds, mark as incorrect and encourage the child to continue, pointing to the next letter and say: “now let’s try this one.” [Insert 5 high frequency words here] Clearly put a tick (/) on any mistake the child makes. In the event that the child corrects him/herself, circle the sign (/) that you previously made for him/her. Put a tick (/) on the last letter the child identifies.	If the child hesitates to name the letter for more than 3 seconds, point to the next letter and say: “Let’s continue, please.”  The Early stop rule: If you mark the first three answers as

	wrong and the child does not correct any mistakes, say “thank you” and stop the exercise.
Section 4: Recognize Letter Sounds	**60 Seconds**
— This is a sheet of Arabic letters and movements. Read as many of them as you can (read the letter's sound, not its name). For example, the sound of this letter [indicate the letter “l”] is “for,” as in the word “playing.” — Now let's do this exercise: tell me the sound of this letter [and point to the letter K]: Good, the sound of this letter is "k." The sound of this letter is "K". — Let's try another example: Tell me the sound of this movement [point to the aperture]: Well done, the sound of this movement is “-” The sound of this movement is “—” — Did you understand what is required from you? When I tell you “Let’s get started," read the sound of the letters as accurately and as quickly as possible. We'll start from here and continue this way [point to the first letter on the first line, and trace it with your finger on the letters in the entire first line]. are you ready? — Let's Begin	After 60 seconds, you will tell the child to 'stop'. If the child hesitates to read the letter for more than 3 seconds, point to the next letter and say: “Let’s continue, please.”
Clearly put a tick (/) on any mistake the child makes. In the event that the child corrects him/herself, circle the sign (/) that you previously made for him/her. Put a tick (/) on the last letter the child reads. Example: for K	The Early stop rule: If you mark all the answers in the first line as wrong and the child does not correct

	any of his mistakes, say “thank you” and stop the exercise. Put an (X) in the box at the bottom of the page and go to the next exercise.
Remaining time of exercise time (number of seconds)	
Check this box (X) □ in case you have left this part of the assessment Because the child did not read any of the words in the first line correctly	

Section 5: Read the character syllable	**60 Seconds**
— This is a sheet that includes Arabic syllables and movements, read as many of them as you can (read the passage). For example, we read this passage [point to the syllable "a'a"]" as in the word "aa". — Now let's do this exercise: read this passage [point to the syllable "ra"]: Good, we read this passage like this "Ra" We read this passage "Ra" — Let's try another example: read this passage [point to the passage]: Well done, we read this passage like this "C" The sound of this movement is "Su" — Did you understand what is required from you? When I tell you “Let’s get started," read the syllable accurately and as quickly as possible. We'll start from here and continue this way [point to the syllable in the first line, and trace it with your finger on the syllable in the entire first line]. are you ready?	After 60 seconds, you will tell the child to 'stop'. If the child hesitates to read the letter for more than 3 seconds, point to the next letter and say: “Let’s continue, please.”

<table>
<tr><td>— Let's Begin</td><td rowspan="2">✋

The Early stop rule:

If you mark all the answers in the first line as wrong and the child does not correct any of his mistakes, say “thank you” and stop the exercise. Put an (X) in the box at the bottom of the page and go to the next exercise.</td></tr>
<tr><td>Clearly put a tick (/) on any mistake the child makes.

In the event that the child corrects him/herself, circle the sign (/) that you previously made for him/her.

Put a tick (/) on the last letter the child reads.

Example: for K</td></tr>
<tr><td>Remaining time of exercise time (number of seconds)</td><td colspan="2"></td></tr>
<tr><td>Check this box (X) □ in case you have left this part of the assessment

Because the child did not read any of the words in the first line correctly</td><td colspan="2"></td></tr>
</table>

Section 6: Read high frequency words	
— This is a sheet that includes words. I'd like you to read as many of them as you can. For example, we read this word [point to the word "[add example word]" as in the word "[example]". — Let's Begin — Can you try the next word? [add 5 high frequency words, use one high frequency word as the example]. Clearly put a tick (/) on any mistake the child makes. In the event that the child corrects him/herself, circle the sign (/) that you previously made for him/her.	

Section 7: Audible comprehension			
The evaluator reads out loud the following text only one time and carefully (about a word every second)			
— I'll read you a short story out loud, just once Then I will ask you some questions Listen carefully, please and answer them correctly, did you understand what is required of you? At the beginning of the summer vacation the family wanted to take a trip. The father suggested to go to the sea, to enjoy swimming, and boating. The mother said: Why don't we go to the countryside to breathe fresh air and collect useful plants As for Khaled, he wanted to go to the city of games, After the discussion was long, the grandfather intervened, saying: We can visit all the places by making a program for the trips, everyone was happy with the grandfather's proposal.			
	No Answer	Not Correct	Correct
When did the family want to take a trip? <u>at the beginning of the summer holidays</u>			
Why does the father want to go to the sea? <u>to enjoy swimming, and boating</u> <u>to enjoy swimming</u>			

riding boat			
Who prefers to collect useful plants? The mother			
Who is the youngest person in the story? Khalid			
What did the family agree on? To make a program for trips on a serious suggestion To go to all the places			
— This is a sheet of Arabic letters and movements. Read as many of them as you can (read the letter's sound, not its name). For example, the sound of this letter [indicate the letter “l”] is “for,” as in the word “playing.” — Now let's do this exercise: tell me the sound of this letter [and point to the letter K]: Good, the sound of this letter is "k." The sound of this letter is "K". — Let's try another example: Tell me the sound of this movement [point to the aperture]: Well done, the sound of this movement is “-” The sound of this movement is “—” — Did you understand what is required from you? When I tell you “Let’s get started," read the sound of the letters as accurately and as quickly as possible. We'll start from here and continue this way [point to the first letter on the first line, and trace it with your finger on the letters in the entire first line]. are you ready? — Let's Begin			After 60 seconds, you will tell the child to 'stop'. If the child hesitates to read the letter for more than 3 seconds, point to the next letter and say: “Let’s continue, please.” The Early stop rule: If you mark all the answers in the first line as wrong and the child does not correct any of his mistakes, say “thank you” and stop the exercise. Put an (X) in the box at the bottom of the page and go to the next exercise.
Clearly put a tick (/) on any mistake the child makes. In the event that the child corrects him/herself, circle the sign (/) that you previously made for him/her. Put a tick (/) on the last letter the child reads. Example: for K			

	10	9	8	7	6	5	4	3	2	1
(10)	نـ	ــه	حـ	ـة	ج	ـة	هـ	ف	تـ	بـ
(20)	ـِ	و	قـ	ز	س	خ	ـهـ	فـ	ة	ب
(30)	ق	ص	ب	ـغـ	ق	و	ز	خ	ـة	صـ
(40)	ذ	ظ	ي	نـ	و	ـهـ	ط	شـ	ض	ـعـ
(50)	ء	خـ	ش	ق	ن	ي	غ	د	بـ	غ
(60)	م	ن	ط	سـ	فـ	ث	ح	ذ	ض	ب
(70)	ـعـ	ط	ث	ق	ج	ــه	ضـ	ــه	ح	تـ
(80)	و	ص	جـ	د	ذ	خـ	ـهـ	جـ	ظ	ل
(90)	س	لـ	ذ	خـ	ث	ء	مـ	خ	عـ	ز
(100)	خـ	هـ	ـهـ	ضـ	ـعـ	ح	و	ذ	ب	ـُ

Remaining time of exercise time (number of seconds)	
Check this box (X) □ in case you have left this part of the assessment Because the child did not read any of the words in the first line correctly	

Section 8 Part A : Read a text orally	**Section 8 Part B : Reading comprehension**
— This is a short story, focus well and read it correctly, aloud and as quickly as possible. - When you're done, I'll ask you some questions about what you've read. Did you understand what is required of you? - When I tell you, "Let's begin." - Start reading. ready? Let's Begin	Pull the text of the story in front of the child and ask them the questions below. Leave the child maximum 15 seconds to answer each question. Ask the question corresponding to each line the child has read until you reach the line with the mark ( ]), which indicates where the child stopped reading.
After 60 seconds, you will tell the child to 'stop'. If the child hesitates to read the letter for more than 3 seconds, point to the next letter and say: “Let’s continue, please.” ✋<u>The Early stop rule:</u> If you mark all the answers in the first line as wrong and the child does not correct any of his	— I will now ask you some questions about the story that I read. Answer the questions correctly.

mistakes, say “thank you” and stop the exercise. Put an (X) in the box at the bottom of the page and go to the next exercise.				
Clearly put a tick (/) on any mistake the child makes. Tick the last word the child read	Put a tick (X) in the box that corresponds to the child's answer, and then move on to the next question.			
Tala came back from school at noon happy, so her mother asked her, 9	When did Tala come back from school? <u>At Noon</u>	No answer	Not Correct	Correct
Her mother about the reason for her joy Tala replied: During my way back, I heard voice of a cat meowing; 21	What is the sound that Tala heard? <u>A cat meowing</u>			
So, I looked for her until I found her stuck between a group of rocks looking at me sadly, 33	What is the sound that Tala heard? <u>Among a group of rocks</u>			
I helped her out; 39	What makes Tala so happy? <u>That she helped the cat get out</u>			
Her mother said: Bless you, my daughter, this is a wonderful deed, may God be pleased with it. 51	What did Tala's mother say to her daughter? <u>GOD</u> b<u>less my daughter, this is a wonderful job. This is a wonderful work (may God be pleased with him/her</u>)			
Remaining time of exercise time (number of seconds):				
Put a tick (X) in this box □ if you stop this part of the assessment because the child did not read any word in the first line correctly.				

Appendix C

Item Response Theory

Item Response Theory (IRT) is used by all leading international testing programmes, including the Programme for International Student Assessment (PISA), Trends in Mathematics and Science Study (TIMSS) and Progress in International reading Literacy Study (PIRLS). IRT has been used specifically for EGRA data in a range of scenarios, and is routinely used by Research Triangle International, who both originally developed EGRA and contextualised it to Jordan.[47]

The key advantage of using IRT to analyse pupil test scores for this impact evaluation, is that, under certain assumptions, this generates estimates of pupil 'ability' in Arabic literacy on an interval scale which can be directly linked to criterion-referenced competencies found in the curriculum. On an interval scale, equal differences between numbers (in this case, pupil ability estimates) reflect equal differences in the amount of the underlying attribute being measured. Since the key objective of the impact evaluation is to measure change in learning achievement over time, an interval measurement scale allows for more accurate estimation of change. Using raw scores and traditional test analysis for this purpose can be substantially misleading (Wright and Stone 1979).[48]

IRT removes the need to "weight" items as the model ranks items according to difficulty and creates an interval scale (where there are equal differences between the number). This is something traditional raw scoring is unable to accomplish.

Rash Model

The Rasch model enables the creation of an interval scale of scores for both the item's difficulty and the person's ability, and these scores are scaled in logits. The Rasch model has the property of specific objectivity, which may be interpreted as sample independence because it places item difficulty and proficiency onto the same scale.[49] This is its advantage over other IRT models.

As Cueto et al. (2009) state:

> "Rasch Models have statistics to evaluate the fit of the item into the model. The idea underlying these statistics is that correct answers in more difficult items are accomplished by people with higher ability. At the same time, these people will have a greater probability of attaining higher scores on easier items than on more difficult ones."[50]

Rasch model statistics can be used to determine whether to keep all of the items in the analysis, and also to provide insights into how to improve the tests for the next round. A summary of the results from the diagnostics tests that were carried out on the Arabic literacy item responses, are included in the pre-pilot results summary.

---

[47] Thissen, D., Nelson, L., Rosa, K., & McLeod, L. (2001). Item response theory for items scored in more than two categories. In D. Thissen, & H. Wainer (Eds.), Test Scoring (pp. 141-186). Mahwah, NJ: Lawrence Erlbaum. Thissen D, Pommerich M, Billeaud K, Williams VSL. Item response theory for scores on tests including polychotomous items with ordered responses. Appl Psychol Meas. 1995;19(1):39-49.

[48] Wright, B. & Stone, M. (1979). Best test design. MESA Press: Chicago, IL

[49] This is not possible with raw scores as the difficulty of the items is fundamentally linked to the score.

[50] Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives. Cueto, Leon, Guerrero and Munoz. 2009. Oxford University. Oxford.

This analysis evidenced that the EGRA alone is too difficult for the children included in this intervention and that the selected items are well-targeted to the ability level of the pupils but also leaves room for growth during the intervention.

Between 2011 and 2020 the Department for International Development, UK commissioned twelve major impact evaluations of their large-scale education programmes. Of these, six included primary data collection of learning outcomes. Of these, five used item response theory applied to early grade reading assessments and early grade mathematics assessments. The psychometric analysis is conducted pooling all intervention and non-intervention groups and pupil proficiency levels are scaled blind to intervention group or school.

Programme	**Primary data collection on learning outcomes?**	**Use IRT on early grade assessments?**
EQUIP-T Endline report	✓	✓
Developing Effective Private Education - Nigeria (DEEPEN)	✓	✓
Developing Effective Private Education - Nigeria (DEEPEN)	✓	✓
Girls Education Project (GEP) Phase 3	✓	✓
Teacher Development Programme	✓	✓
Complementary Basic Education (CBE) in Ghana	✓	X
Punjab Education Support Programme II	X	
Ilm Ideas Education Innovation – Phase II	X	
Girls Education Challenge Fund (Step Change Window, Innovation Window, Strategic Partnerships Window)	X	
EQUIP-T Endline report	X	
Developing Effective Private Education - Nigeria (DEEPEN)	X	
Developing Effective Private Education - Nigeria (DEEPEN)	X	