

Early Grade Reading Assessment (EGRA) Toolkit

Updated: June 19, 2026

Final draft* completed by: Hetal Thukral, sustainedED, LLC; Aimee Reeves, Seneca Consulting. Original draft by the USAID DECODE team (partially completed at the time of USAID's closure of the DECODE activity in February 2025)

This is the third edition (V1) of the Early Grade Reading Assessment (EGRA) Toolkit. This Toolkit is structured differently than previous versions and is intended to be accessible to a wider audience, specifically those overseeing the EGRA process and managing a team of experts. The current version does not include annexes for phases 3, 4 or 5, which will be made available in future versions.

CONTENTS *Five phases of an EGRA*



Click any phase above to jump straight to that section.

* Final draft completed with support from Gates Foundation and feedback from experts including: Amber Gove, Tangerine Central; Melissa Chiappetta, Sage Perspectives; Abdullah Ferdous, AIR; Luis Crouch, emeritus RTI; Cally Ardington, AFLearn; Brahm Fleisch, University of Wits; Rona Bronwin, FCDO; Marie-Helene Cloutier, World Bank; and Clio Dintilhac, Gates Foundation.

TABLE OF CONTENTS

Introduction

What is EGRA?

Who is the target audience for the EGRA Toolkit 3.0?

How is the EGRA Toolkit 3.0 structured?

01 Design

Why Choose EGRA

EGRA Among Assessment Options

Planning Your EGRA

Identify Your Team and Stakeholders

Determine Your Purpose

Identify the Languages You Will Assess

Determine the Grade(s) You Will Assess

Select Your Methodology

Plan your Research Ethics Review

Cost of EGRA

Decision set 1: Planning your EGRA

Selecting your EGRA Subtasks

Common Measures of Oral Language

Common Measures of Phonological Awareness

Additional Measures of Phonological Awareness

Common Measures of Alphabetic Principle (Letter-Sound Correspondence)

Additional Measures of Alphabetic Principle (Letter-Sound Correspondence)

Common Measures of Alphabetic Principles (Decoding)

Common Measures of Fluency

Common Measures of Reading Comprehension

Additional Measures of Reading Comprehension

Optional Subtasks–Vocabulary and Oral Language

Universal Design for Assessment Principles

Subtask and Language Considerations

Decision set 2: EGRA Tool Design

Determining your Data Collection and Analysis Considerations

Data Collection Planning
Determining Additional Survey Needs
Planning EGRA Data Use

02 Adapt

Conducting an Adaptation Workshop

Adaptation Workshop Agenda
Adaptation Workshop Participants and Materials
Adaptation Workshop and Validity

Identifying Accommodations and Modifications

Creation of Equivalent Test Forms

EGRA Learner Instructions and Protocols

EGRA Learner Instructions
Early Stop Rule
Pause Rule

Pretest and Pilot

Pretesting your EGRA
Piloting your EGRA

03 Implement

Selecting Your Assessment Team

Training the Assessment Team

Conducting an Operational Data Collection

Planning Your Data Collection Schedule
Materials Needed for Data Collection
Daily Tasks at the Data Collection Site
Engaging Learners throughout the Assessment
Uploading Tablet-Based Data

Developing A Quality Assurance and Quality Control Plan

04 Analyze

05 Report

Introduction

What is EGRA?

EGRA is a framework for assessing reading comprehension that is adapted to different contexts and languages. It has been used in over 70 countries and to assess hundreds of languages. The framework has been used by thousands of institutions, including Ministries of Education.

EGRA is unique because:

- it is orally administered and can be used with young learners prior to their development of writing skills.
- it provides diagnostic information on precursor reading skills that are necessary for reading fluency and comprehension.
- it can be developed and administered in multiple languages based on common development guidance (as provided in this toolkit).
- it can be used for grades 1, 2, 3, and in some cases, higher grades when appropriate, especially if performance on written tests is poor.

Who is the target audience for the EGRA Toolkit 3.0?

The EGRA Toolkit 3.0 provides high-level, accessible guidance for an EGRA team at a Ministry of Education, research institute, non-governmental organization, or civil society organization. In contrast with the EGRA Toolkit 2.0, this version is specifically written for a general audience (i.e., readers who are both familiar with the EGRA and those who are not).

How is the EGRA Toolkit 3.0 structured?

This Toolkit provides guidance on how to conduct an EGRA and use the evidence produced through an EGRA. The Toolkit is organized into five phases (Figure 1), beginning with Design and concluding with Report and Use. Each chapter of this toolkit provides the

information that an EGRA manager would need to oversee the EGRA process. Within each chapter, the Toolkit references technical annexes and supplementary materials that are appropriate for technical members of the EGRA team.

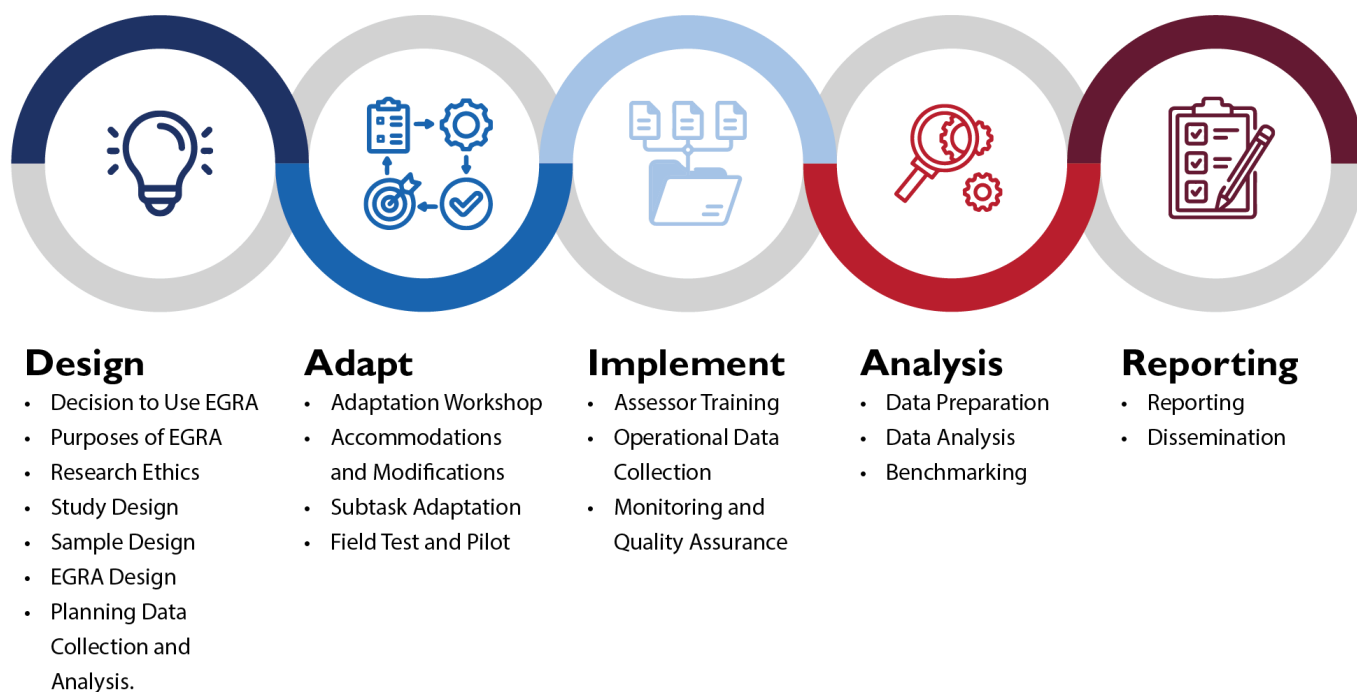


Figure 1: The Five EGRA Phases

01 Design

The first step in conducting an EGRA is Design. This is a foundational phase in which you make key decisions about why you are using EGRA and your specific purpose for the assessment. Your decisions in the Design phase impact all future phases of your EGRA.

The Design phase explores the following themes:

- Why conduct an EGRA, and why is EGRA unique among assessment options
- What specific decisions you must make when designing your EGRA
- What subtasks you should consider
- What specific decisions you must make when planning your EGRA data collection

The team members and technical capacities that will need to be involved in this phase include...

- EGRA Manager
- Reading Specialist
- Language Expert
- Experts in Sign languages and/or Braille (if relevant)
- UDA Expert
- Psychometrician and/or Statistician

Why Choose EGRA

UNICEF and the World Bank estimates that only about one-third of 10-year olds are able to read and understand a simple story. Since the COVID pandemic, while enrollment increased, average student learning in most low-income countries remained quite low. At the root of low learning levels is low levels of literacy. Basic literacy is the foundation children need to be successful in all other areas of education. Children first need to “learn to read” so that they can “read to learn.” That is, as children pass through the grade levels, more and more academic content is transmitted to them through text, and their ability to acquire new knowledge and skills depends largely on their ability to read and extract meaning from text.

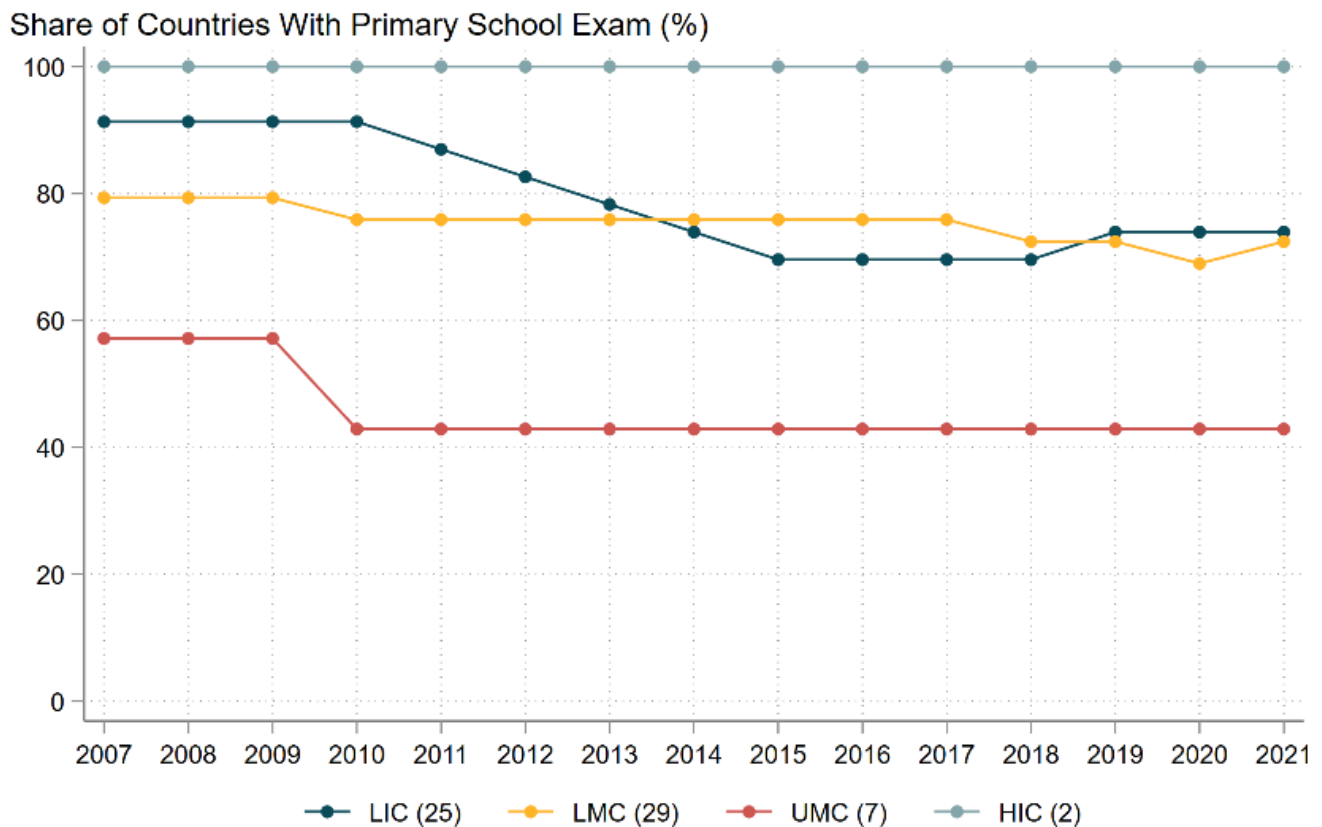


Figure 2: Share of Countries with Primary School Exam (%) (Center for Global Development, 2022)

Children who do not learn to read in the first few grades are more likely to repeat grades and eventually drop out of school; low literacy levels are linked to poor outcomes later in life. In order to address the growing number of young children unable to read, accurate information about their learning in the early grades is necessary. Until 2009, primary school leaving exams were prevalent in most low and middle income countries. Even when these leaving exams are administered, they are in written format and commonly administered in upper primary grades (Figure 2).

An assessment like the primary school exam – even if administered in earlier grades – is not designed or administered in a way that can provide accurate information on early literacy skills. First, the written format is unsuitable to measuring precursor skills that cannot be demonstrated through writing – like listening comprehension or decoding. If these skills are somehow tested in written format, the written test confuses important differences between learners – those who cannot write and those who cannot perform the required task. For example, if a student scores poorly on a written listening comprehension task, we

cannot determine if their low score is because they did not understand the story they heard (an important precursor skill for reading) or because they did not know how to mark their answer correctly on the test paper (because they do not have sufficient reading ability). The results obtained from a test like this will likely lead to a circular question – in order to take a written test, one first needs writing skills, which rely in turn on the ability to read. How does one then test reading ability accurately?

An orally-administered assessment can give us more information about what literacy skills a learner has – information that is not available when testing in a written format. Despite this, there are few standardized assessments that orally measure students' learning levels in the early grades. EGRA assesses reading comprehension, and the precursor skills needed for reading with comprehension. The EGRA measures a child's ability to read and comprehend simple text, a skill that is also measured by other standardized assessments.

What differentiates the EGRA is that it also measures the precursor skills that are critical to read and comprehend text, such as letter identification, decoding, and listening comprehension. The precursor skills that EGRA measures have been shown to be linked to reading comprehension across multiple languages with varying orthographies, meaning that the EGRA allows us to identify a child's progress on their way to becoming a reader.

EGRA makes it possible to modify or improve reading skills through adjusted instruction. The data obtained by administering an EGRA with learners provides meaningful insights – both in the learner's future performance and where they may benefit from instructional changes. For example, teachers can pinpoint the precursor skill or skills where a learner is getting stuck in addition to trends across groups of learners. EGRA data can also provide macro insights to inform changes in teacher professional development – identifying which reading skills are teachers in most need of teaching differently and incorporating tools, such as phonics-based instructional methods, into teacher professional development.

RESEARCH SPOTLIGHT

Research using EGRA data shows that a student's oral language, decoding, and fluency skills directly contribute to reading comprehension, while phonological awareness and alphabetic principle (i.e., letter-sound knowledge) contribute indirectly to reading comprehension through the ability to read whole words (Kim 2017; Kim 2020a; Kim 2020b).

The rate of reading words also matters. In order to comprehend the meaning of a series of words requires the learner to also read them at a pace that facilitates meaning-making. The EGRA measures these various skills necessary for a learner to read with fluency and comprehend the meaning.

— EGRA Among Assessment Options

The EGRA is one of 17 internationally recognized assessments that are used to measure reading, and each assessment serves different purposes and uses different methodologies. These 17 assessments are described in the Design Annex.

EGRA is defined by several key characteristics:

- It is orally administered and can be used with young learners prior to their development of writing skills
- It provides diagnostic information on precursor reading skills that are necessary for reading fluency and comprehension
- It can be developed and administered in multiple languages based on common development guidance (as provided in this toolkit)
- It can be used for grades 1, 2, 3, and in some cases, higher grades when appropriate, especially if performance on written tests is poor

FEATURES OF A HIGH-QUALITY EARLY GRADE ASSESSMENT: THE CASE OF EGRA

The UNESCO Institute for Statistics maintains a set of criteria for assessments to be used to report to SDG 4.1.1. The criteria ensure that assessments adequately measure reading ability and its underlying skills and that results are comparable. If you expect that the EGRA you are planning might be used for SDG 4.1.1 reporting, review the latest guidance available from UIS, including reporting templates. As you navigate the sections of this Toolkit, use the UIS criteria alongside to ensure the decisions you make in each step results in an EGRA will meet the reporting criteria: there are specific expectations regarding sampling and documentation that you will have to plan for.

How can you ensure that your EGRA meets quality standards for assessments?

- Follow the guidance in this Toolkit and the technical details in the Annexes.
- If you anticipate using EGRA results for SDG 4.1.1a reporting, consult the latest reporting and documentation requirements from UIS.

Planning Your EGRA

In this section you will learn about and identify the key parameters and decisions you will make during the Design phase to ensure that your EGRA meets your goals. The decisions you make during Design will inform actions in subsequent phases.

— Identify Your Team and Stakeholders

Successfully conducting an EGRA requires a team of experts (Table 1) and collaboration and communication with stakeholders (Table 2). In this case, stakeholders are individuals who have an interest in the results of the EGRA but are likely not involved in its implementation. Stakeholders include high-level officials such as government ministers, donor or funder representatives, school administrators, teachers, parents, and community members. Table 1 provides an overview of the members of the team, their typical tasks, and the recommended qualifications and Table 2 describes the role of stakeholders who are critical to a successful assessment. See Design Annex for detailed TORs.

TABLE 1: EGRA TEAM MEMBERS, TASKS, AND QUALIFICATIONS

EGRA Team Member Role	Typical EGRA Tasks	Recommended Qualifications
EGRA Manager	Manage the EGRA process and budget; coordinate EGRA team; ensure all technical tasks are completed	Project and/or research management; understanding of research processes
Reading Specialist	Co-facilitate EGRA adaptation workshops, support revisions, interpretation of results	Reading expertise in the language of the assessment, experience teaching reading with the learners being assessed
Language Expert	Support reading expert when administering EGRA in multiple languages	Expertise in the languages of the assessment, background in language policy and instruction
Curriculum Expert	Support and participate in EGRA adaptation workshops	Expertise in national and/ or regional curriculum, experience with the grade level being assessed
Experts in sign language, braille	Co-facilitate EGRA adaptation workshops for learners with difficulties, support revisions, administration and interpretation of results	Expertise in sign language and braille adaptations, classroom assessment using sign language and/ or braille
UDA Expert	Support design and development of the EGRA instrument, Co-facilitate EGRA adaptation workshop for learners with disabilities.	Expertise in UDA principles, disability inclusion and adapting assessments for children with disabilities, and experience in the region or country of assessment
Psychometrician (expert in measuring learning through assessments)	Ensure adherence to the 6 criteria for high-quality assessments, support subject-matter experts	Expertise in developing assessments, conducting item analysis and processing data
Statistician and/or analyst	Support item development, piloting, sample selection, data analysis and reporting, benchmarking	Expertise in data analysis, statistical testing, sampling, reporting technical results in easily accessible writing

TABLE 2: EGRA STAKEHOLDERS, TASKS, AND QUALIFICATIONS

EGRA Stakeholders	Typical EGRA Task	Recommended Qualifications
Teachers who teach reading in the grades being tested	Support the development of items during workshops and especially reading passages to ensure they are leveled and appropriate for the students, conducting pilot and interpreting results	Experience teaching reading in the grade levels being tested by EGRA, experience using data to inform instruction
Education officials involved in setting standards for learning	Support development of items to ensure adequacy of the EGRA to measure grade-level expectations, support establishing targets, review interpretation of findings and proportion of learners meeting standards	Have in-depth knowledge of children's learning levels either through classroom-based monitoring, teacher mentoring, or conducting formative or standardized assessments routinely

— Determine Your Purpose

The EGRA framework is made up of a set of subtasks that measure various skills that contribute to the higher-order skill of reading comprehension. Your choice of subtasks, approach to analysis and data use all require a clear understanding of your purpose for conducting an EGRA in the first place.

Note that your EGRA can be designed for one or more purposes. The more purposes you are trying to achieve with one assessment, the more resources may be required. For example, you may need a larger team to collect more data, you may need to conduct more analyses, and/or may need to produce multiple reports or products.

The most common purposes for conducting an EGRA are described below.

Using EGRA findings to inform classroom instruction.

When EGRA findings are used to inform classroom instruction, the assessment can also be called a formative assessment – to inform instruction based on progress the learner is making through the course of the year. By contrast, an end-of-year assessment is called a summative assessment – to determine whether the learner has mastered the required skills for that grade level.

EGRA as a formative assessment allows teachers to track learning. An EGRA can provide teachers with specific information on learners' skills and needs in phonological awareness, alphabetic principle (letter-sound or syllable-sound correspondence), decoding, fluency, reading comprehension, oral language, or vocabulary. Specifically, when used as a formative assessment, EGRA results can be used by teachers to make a variety of daily pedagogical decisions, such as how to provide individualized feedback to specific learners, modify instruction for the whole class, or place learners into differentiated small groups for instruction, based on learners' performance on EGRA subtasks.

When EGRA is used to inform classroom instruction, you should time its administration based on the academic calendar and the curriculum.

Using EGRA findings to report to the Sustainable Development Goals Indicator 4.1.1a.

When designed, administered, and analyzed according to the reporting criteria established by the UNESCO Institute for Statistics (UIS), EGRA can be used to report on Sustainable Development Goal (SDG) 4.1.1a.¹ Guidance for reporting to SDG 4.1.1A are included in UNESCO's SDG 4.1.1 Toolkit, available on the UIS website.

SDG INDICATOR 4.1.1A: THE PROPORTION OF CHILDREN AND YOUNG PEOPLE IN GRADES 2/3 WHO ACHIEVE AT LEAST A MINIMUM PROFICIENCY LEVEL IN (I) READING AND (II) MATHEMATICS, BY SEX.

If using the EGRA to report on SDG 4.1.1a, you should be aware of the following considerations:

- **Report using data collected from Grade 2 students.** The definition of SDG Indicator 4.1.1a is based on UIS's Minimum Proficiency Levels (MPLs) and the Global Proficiency Framework (GPF) detailed definition of skills, which coincide in Grade 2.
- **Ensure you have a nationally representative sample.** This means that the schools you include in your EGRA sample must be selected from a reliable list of all schools in a country. If you exclude schools, regions, or subgroups of students, you must provide thorough documentation and explanations for the exclusions.
- **Ensure you know what is required in the SDG submission package and that you have the technical staff capacity available to assemble the package.** The criteria for SDG 4.1.1 require documentation of all steps of the assessment including item development,

piloting, enumerator training, data collection, analysis and sustainability. In order to meet the criteria, the following staff expertise is required: a reading specialist, a language specialist, at least one psychometrician, at least one statistician, an EGRA trainer, field enumerators and field supervisors, and at least one analyst. When there are additional complexities, such as multiple languages, grades, more complex samples or language, additional team members and expertise may be required.

Using EGRA findings to measure reading progress over time.

EGRA can be used to track trends in a group of learners' progress over time and/or across contexts. When using an EGRA to track progress across time – such as evaluating a reading intervention at baseline, midline and endline – you may need to create equivalent test forms. Equivalent test forms are statistically the same in difficulty but slightly different in content – they have to be equated in order to achieve equivalence. Having equivalent test forms ensures that you are measuring knowledge and not recall and avoid the risk of test leakage. See 2.6. Creation of Equivalent Test Forms for more details.

When comparing EGRA results from one timepoint to another, remember that you can only compare reading outcomes in different contexts – such as for different languages, populations, and geographies – by comparing the percentage of learners meeting a common threshold or benchmark. To compare the percentage of learners meeting benchmarks, you will need to engage psychometricians and/or statisticians. The benchmarks are typically established during a workshop. Using tools like the GPF and policy linking, you will work with a team of reading and language experts, assessment experts and teachers over the course of 3-5 days.

If you will need equated forms and/or benchmarking workshops, note that these will require additional piloting and analyses before the operational EGRA data collection can begin. It will also require additional analyses to be completed so that you can compare appropriately across contexts. For more information, please see the Design Annex.

INAPPROPRIATE USES OF EGRA

EGRA should not be used as a high-stakes accountability measure to determine whether a child advances to the next grade. EGRA results should also not be used to evaluate an individual teacher's instructional ability. Finally, EGRA should not be used as an instructional program in and of itself.

RESOURCES

- 🔗 [Global Proficiency Framework](#)
 - 🔗 [What is the Global Proficiency Framework? YouTube explainer](#)
 - 🔗 [An overview of the framework and how it is used](#)
 - 🔗 [The Global Proficiency Framework – English \(pdf\), Spanish, French, Arabic](#)
- 🔗 [Policy linking Toolkit homepage](#)
- 🔗 [How Policy Linking and GPF can help education systems](#)
- 🔗 [ALIGN: Aligning Learning Inputs to Global Norms using the GPF YouTube explainer](#)
- 🔗 [SDG Reporting Toolkit](#)

— Identify the Languages You Will Assess

Deciding which language(s) your EGRA will assess is a critical step in the Design phase. You will also need to determine in which language(s) the assessor will provide assessment instructions to learners. The language(s) you assess will depend on the purpose of your EGRA and the language(s) of instruction in schools in your context.

It is important to know that assessing learners in more than one language will multiply amount of time that a child will need to sit for the assessment. It will also increase the overall time and budget required throughout your EGRA process. If you plan to assess more than one language, you should work closely with psychometricians/statisticians to create a right-fit sample and with reading experts on designing a right-fit EGRA, potentially considering reducing the number of total subtasks. Nationally-representative samples across multi-lingual contexts will likely mean that you will need to stratify – or organize – the population from which you will draw the sample, by language.

Oftentimes, a language mapping exercise will provide the information you need to understand the numbers and types of languages that the EGRA will need to accommodate. For more information on how to conduct a language mapping exercise please see the Design Annex.

RESOURCES

- [Language mapping exercise in Zambia](#)
- [Language mapping exercise in The Gambia](#)
- [What does school language mapping contribute to effective MLE programming?](#)

— Determine the Grade(s) You Will Assess

The EGRA framework measures foundational literacy in early primary years. Based on the GPF's descriptions of what children should be able to do in each grade, the EGRA reliably measures early reading skills up through Grade 4 in most countries. However, in contexts where actual learning levels or curriculum are lower than GPF's expectations, the EGRA may provide useful information for learners above Grade 4.

USING EGRA FOR OLDER LEARNERS

Using an EGRA to assess learners above Grade 3 should be based on a review of the curriculum, the proportion of learners able to read and comprehend text, and a comparison of the skills measured by the EGRA with the information needed for your purpose.

The following describe situations when using an EGRA for learners in Grade 4 or higher may be appropriate:

- Where upper primary learners perform below standards and many learners have low reading scores.
- When learning has suffered interruption, particularly due to conflict or crisis situations, such as natural disasters, generalized violence or armed conflict, forced migration and displacement, and disease outbreak or a pandemic.
- When learners are overage and/or learners are out of school and grade-based expectations are difficult to map to the learners' abilities.
- Learners may be learning in a language different from the languages they know and understand. For example, second language learners or migrants.
- When it is necessary to track the learning levels of a cohort of learners across grades, such as in a longitudinal study.

The grades you assess should correspond with your purpose for conducting an EGRA. For example, if you are using the EGRA for SDG 4.1.1a reporting, you should assess grade 2 learners. If you are using the EGRA to measure progress for a specific reading intervention, you should assess learners in the grades that are receiving that intervention. It is important to also be mindful of when in the school year you are using an EGRA, as a learner will likely have different reading skills at the beginning of the school year compared to the end.

— Select Your Methodology

It is useful to think of an EGRA administration as a study so that you can apply principles of study design. The methodology you select depends on the purpose of your EGRA and on your budget. Table 3 provides illustrative methodologies that are appropriate for different

purposes and major cost drivers to keep in mind as you determine your methodology. Since methodology selection and managing your budget will go hand in hand, you should collaborate closely with a psychometrician/statistician during this phase.

TABLE 3: ILLUSTRATIVE METHODOLOGIES AND COST DRIVERS BY EGRA PURPOSE

Purpose	Examples of study methodologies (illustrative, not exhaustive)	Major cost drivers to consider for the methodology
Using EGRA findings to inform classroom instruction	Census ² administration Sample-based cross-sectional study	Making results readily available to teachers to use following administration (dashboard or coaching session, for example) Frequency of administration (quarterly, monthly)
Using EGRA findings to report to SDG Indicator 4.1.1a	Nationally representative Grade 2 sample	Conducting all phases of the EGRA in accordance with the 6 criteria with adequate psychometric, statistical and analytical expertise
Using EGRA findings to measure reading progress over time	Longitudinal or semi-longitudinal study of a cohort across time Sample-based cross-sectional study of treatment and comparison schools across time Census-based study across time	Size of the sample Equating exercise across forms Cost of tracking individual learners for the duration of the study

— Plan your Research Ethics Review

Research guidelines, often referred to as research ethics, include a set of principles that are designed to protect the research subjects, data use, and fairness. It is particularly important to adhere to research ethics when conducting research with children, as is the case when conducting an EGRA. You will often need to apply for research approval from an Institutional Review Board (IRB) or a specific research ethics office in the country where you will be conducting your EGRA.

You should plan to have your methodology, study protocol, and all tools, including the EGRA, reviewed and approved by an IRB or equivalent entity. The time required for this review and approval varies depending on the entity, so it is important to research the approval process prior to finalizing a timeline for your EGRA.

You must obtain informed consent, for adults, and assent, for children, for all participants prior to administering any data collection tool. If conducting an EGRA in schools, this may not be necessary, however, as schools often have caregiver pre-approval to conduct assessments. You should be sure to understand what consent protocols apply to your context as you are planning your EGRA data collection.

You should also plan early for your data management protocols, as information such as where data will be stored, who will have access to data and for how long, will be part of the informed consent.

An IRB review typically involves submission of a package of documents, including the EGRA tools, consent forms, study design and data management plans. Remember that this review may take anywhere from a few days up to months and must be completed prior to the start of any data collection.

— Cost of EGRA

An EGRA study can be time- and resource-intensive. On average, USAID found that assessment costs accounted for 6%-13% of a total funded activity budget (Venetis, Harris-van Keuren, 2024). The key cost drivers for an EGRA study include:

- Data collection software and number of tablets required
- Salaries for a team of experts with specialized skills for designing the EGRA tool, the study methodology and analysis of the data; salaries for field teams and travel (international, regional or local)
- Number of days of data collection required to cover the sample - daily per diems, transport and lodging for data collection teams are often the largest cost for field work. Larger sample and more tools typically will require more days in the field (which in turn will require more cost)
- Number of EGRA tools being developed - when multiple assessments are planned over time or over multiple languages, it is often cost-effective to develop all the tools at once. Alternatively, you may plan to develop an item bank from which each tool will be

constructed. Psychometricians will help you navigate this decision process.

The key decision that influences these cost drivers is the size of the sample – more schools means more tablets, staff and days in the field. Therefore, ensure you have a statistician to support you in selecting a sample that best meets your study purpose without overshooting the sample size. For example, if you decide to disaggregate results to the regional level, the statistician will ask whether the decisions being made using the EGRA results are high stakes at the disaggregated level, or whether the disaggregations are simply to give you a sense of the trends. Depending on the use of the results at the disaggregated level, the sampling procedure may vary and can significantly reduce (or increase) the sample size.

— Decision set 1: Planning your EGRA

This section summarizes the key decisions for Planning your EGRA. These decisions should be made before continuing through the following sections to ensure that your EGRA is fit-for-purpose and can be implemented with quality. Once you have answers to the questions below, you should proceed to selecting your EGRA subtasks.

DECISION SET 1

- 1 Do I have the right expertise and sufficient time from experts to conduct the EGRA study appropriately? See 1.2.1 Identify Your Team and Stakeholders for details on the EGRA team and qualifications.
- 2 Have I clearly articulated the purpose for conducting an EGRA? Knowing your purpose clearly will provide important information to experts responsible for designing the EGRA tool, selecting samples, and conducting analyses of results. In other words, the purpose will drive decisions made in each subsequent step.
- 3 Will an EGRA meet my purpose? EGRA is ideal for assessing early grades in an oral format to understand reading comprehension as well as precursor skills to inform instruction, report to SDG 4.1.1a and/or track trends over time.
- 4 Which language(s) will I cover in the EGRA? What do the results of the language mapping exercise suggest and will data from multiple language assessments meet my purpose?
- 5 Do I need to conduct any pre-EGRA exercises such as a language mapping exercise, a benchmark workshop, and/or an equating exercise? These exercises will inform subsequent steps in the EGRA process and their results are needed before proceeding.
- 6 Which learners' reading abilities do I want to understand? The population of learners to whom you will generalize results is an important population to define. Which grades, communities, geographies are included or excluded?
- 7 What sample and study methodology best suits my purpose? The sample must be representative of the population whose reading abilities you seek to understand; the study methodology must suit the purpose, cost parameters and timeline for the study.
- 8 Do I have a plan for reviewing the research ethics guidelines for my study? This may be a review by an IRB or an equivalent entity in-country; note that this review must be completed and authorization to move forward received from the reviewing entity.
- 9 What is my budget? Many of your decisions will be contingent on the budget you have available. It is important to define an estimated budget early in your Design process so that you are making cost-informed decisions.

Selecting your EGRA Subtasks

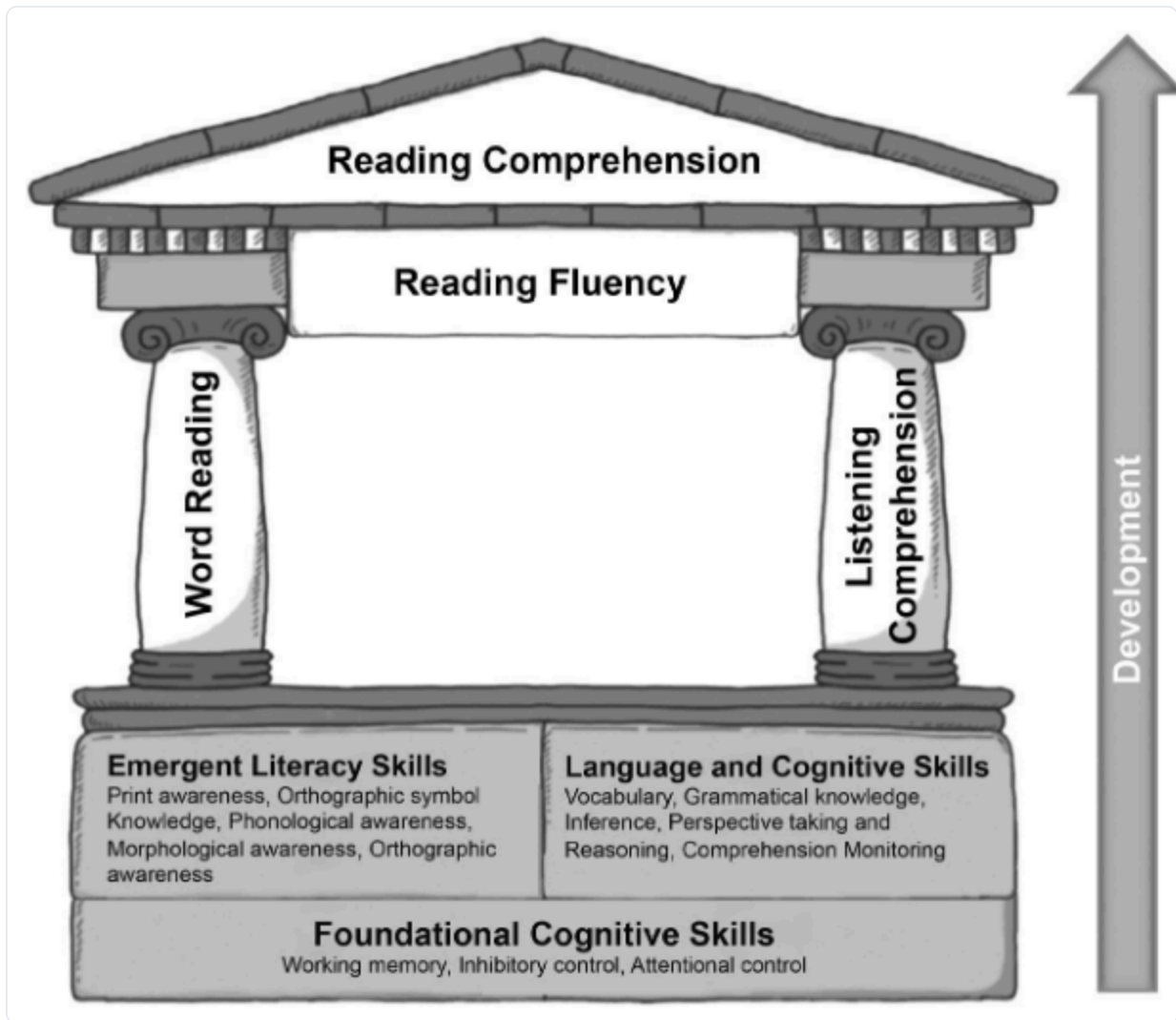
EGRA is a framework that must be customized for a specific context and specific purpose. You can think of this customization, formally called adaptation, as the process of taking the EGRA framework and using it to produce a specific assessment, with specific subtasks, for a specific population of learners. Prior to Adaptation, the second phase of your EGRA, you need to determine which subtasks are appropriate for your purpose. This selection is dependent on the decisions you made in the 1.2 Planning Your EGRA section of this toolkit.

Before selecting your subtasks and determining their order, you should have identified your purpose, learner population, and languages. Now, you will need to select the reading skills you want to measure and decide which EGRA subtasks will provide this information.

You should also keep in mind the total length of the EGRA. Adding too many subtasks to your EGRA will extend the time needed to administer the assessment, which can create fatigue among learners and has ethical implications. The average total administration time for an EGRA should fall within the range of 15 to 25 minutes; this means the EGRA should comprise roughly five to eight subtasks.

EGRA subtasks are grouped into five domains that correspond to the commonly recognized pillars of literacy³:

- Oral language
- Phonological awareness
- Alphabetic principle
- Fluency
- Reading comprehension



The Six Pillars of Literacy



The EGRA framework is designed to be progressive, starting with lower-order – or foundational – subtasks and moving to higher-order subtasks. This means that an EGRA starts with easier subtasks and progresses to a reading comprehension subtask. Oral language subtasks generally are the first or the last in an EGRA, and EGRA designers often put an oral language subtask, which tends to be easier and does not require reading, at the end to allow learners to end on a more positive note. You should determine the subtask order in coordination with your reading expert and psychometrician.

TABLE 4: EGRA SUBTASKS BY READING SKILL

Reading Skill (domain)	Subtask	Type	Number of Items	Skill Demonstrated by Learner's Ability To:
Oral language	Listening comprehension	Common	10	Respond to comprehension questions about a text the assessor reads out loud
Phonological awareness	Initial sound identification	Common	20	Identify the first sound in a common word
	Letter sound discrimination	Common	10	Identify the first sound different from set of three words
	Phoneme segmentation	Additional	10	Identify and sound out each sound present in a word
Alphabetic principle (phonics)	Letter name identification	Common	50	Provide the name of letters
	Letter sound identification	Common	50	Provide the sound of letters.
	Syllable identification	Common	50	Provide the sound of syllables
	Dictation	Additional	10	Write down letter sounds, words, and/or a short sentence read out loud by an assessor
Alphabetic principle (decoding)	Familiar word reading	Common	50	Read words which are randomly ordered and drawn from a list of frequent words
	Nonword reading	Common	50	Make letter-sound (grapheme-phoneme correspondences) through the reading of simple nonsense words

Fluency	Reading accuracy	Common	2 to 3 passages	Read a text out loud for accuracy
	Oral reading	Common	2 to 3 passages	Read a text out loud at a sufficient rate (timed, typically one minute)
Reading comprehension	Sentence verification	Common	10	Respond to comprehension questions about sentences they have read silently or out loud
	Passage reading comprehension	Common	At least 10 questions	Respond to comprehension questions about passages they have read silently or out loud
	Maze	Additional	10	Identify which of three words makes sense in a sentence

In this section, you can learn more about reading skills and EGRA subtasks to help guide your decision. You should consider this as a menu of options and should work with a reading expert to decide which subtasks should be included in your EGRA. More information on each subtask and how to construct it are included in the Subtask Design and Development Annex.

— Common Measures of Oral Language

Listening comprehension subtask assesses a child's ability to comprehend and understand oral discourse, and cover language comprehension domains such as receptive vocabulary, syntactic knowledge, and inference making. Historically, assessing listening comprehension has the assessor read children a short story on a familiar topic and then ask the children comprehension questions about what they heard. The listening comprehension subtask is used primarily in juxtaposition with the reading comprehension subtask to determine whether comprehension difficulties stem primarily from low reading skills or from low overall language comprehension.

To learn about how to construct the listening comprehension subtask, see the Subtask Design and Development Annex.

— Common Measures of Phonological Awareness

As Stanovich (2000) and others have indicated, “children who begin school with little phonological awareness have trouble acquiring alphabetic coding skill and thus have difficulty recognizing words” (393). Research has found that phonological awareness plays an important role in reading acquisition. Testing for and remediating phonological awareness deficits is thus important for later reading development.

EGRA most commonly measures phonemic awareness (one aspect of phonological awareness) through the identification or discrimination of initial or final sounds. These approaches are common in tests of early reading, including:

- Dynamic Indicators of Basic Early Literacy Skills (DIBELS);
- Test of Phonological Awareness, Second Edition Plus (TOPA-2+); and
- Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2).

EGRA instruments typically include one or two measures of phonological awareness, usually at the phonemic level (i.e., phonemic awareness). These include initial sound identification subtasks, in which children are presented with a word orally and asked to isolate and pronounce only the first sound of the word; or a letter sound discrimination task, in which the children are presented with three words and asked to pick out the word with a beginning sound that differs from the other two.

To learn about how to construct the initial sound identification and letter sound discrimination subtasks, see the Subtask Design and Development Annex.

— Additional Measures of Phonological Awareness

An optional (i.e., not core) phoneme (or syllable) segmentation subtask, in which the children are presented with a word orally and asked to pronounce its component phonemes (or syllables) individually, has also been used. Because segmentation is a more complex skill (Linan-Thompson & Vaughn 2007), this subtask has produced floor effects in many low-performing contexts. It can be an appropriate alternative, however, when the initial sound identification subtask produces a ceiling effect.

To learn about how to construct the syllable segmentation subtasks, see the Subtask Design and Development Annex.

— Common Measures of Alphabetic Principle (Letter-Sound Correspondence)

Knowledge of how letters correspond to names or sounds is another critical skill children must master to become successful readers. Letter-sound correspondences are typically taught through phonics-based approaches. Letter identification knowledge is a fairly common assessment approach and is used in several early reading assessments, including the Preschool Comprehensive Test of Phonological and Print Processing (Lonigan, Wagner, Torgesen & Rashotte 2002).

Research, especially from less transparent orthographies like English, has shown that knowing letter name identification is highly predictive of later reading achievement in those languages. This subtask tests children's ability to recognize the graphemic features of each letter and accurately map it to its corresponding name. Originally, letter name identification was the more widely used measure of alphabet knowledge within EGRA assessments, and it was shown to be a strong predictor of later reading achievement in English.

Letter sound identification tests the actual knowledge learners must have to be able to decode words—i.e., knowing the sound the letter represents allows learners to sound out a word. This subtask tests children's ability to recognize the graphemic features of each letter and accurately map it to its corresponding sound. Letter sound identification has become the more frequent option, because letter sound knowledge is more directly linked to children's ability to decode words, especially in transparent orthographies (Ehri 1998).

The syllable identification subtask has been used in contexts where the language has primarily open (i.e., vowel-final) syllables and/or where the reading pedagogy stresses syllabic combinations.

To learn about how to construct the letter sound identification, letter name identification, and syllable identification subtasks, see the Subtask Design and Development Annex.

— Additional Measures of Alphabetic

Principle (Letter-Sound Correspondence)

Finally, dictation asks learners to listen to letter sounds, words, and/or a short sentence and then write them down. The subtask measures learners' alphabet knowledge and ability to hear and distinguish the individual letter sounds in isolation or in words and to spell (encode) words correctly. If a sentence is presented, the task may also measure their ability to use correct sentence-writing conventions such as capital letters and punctuation. This subtask has proven challenging to score in a standardized way in some contexts. It is no longer part of the core instrument but has been used in some countries that have found it appropriate.

To learn about how to construct the dictation subtasks, see the Subtask Design and Development Annex.

— Common Measures of Alphabetic Principles (Decoding)

The next step up in skill difficulty is for readers to use their mastery of the letter-sound correspondences to decode words.

The familiar word reading subtask presents a list of words that children are expected to be able to read at their grade level and will probably have encountered before. According to the dual-route model, children are more likely to process familiar words—if they are indeed familiar, and especially if they have irregular spellings—directly by the lexical route. That is, they might recognize the words instantly, rather than attempting to decode them sound by sound.

The nonword reading subtask provides indirect insight into children's ability to decode unfamiliar words. The nonword reading subtask presents the children with a written list of pseudowords that follow the phonological and spelling rules of the language but are not actual words in the language. Children are asked to read as many of the nonwords out loud as they can, as quickly and carefully as they can. According to the dual-route model, this subtask requires children to apply their decoding skills based on their knowledge of the grapheme-phoneme mappings. Because the nonwords will not have any whole-word representation previously stored in long-term memory that could be accessed directly, learners must rely on decoding to identify them.

To learn about how to construct the familiar word reading and nonword reading subtasks, see the Subtask Design and Development Annex.

— Common Measures of Fluency

Reading accuracy is an important measure of learners' ability in grade 1 and grade 2. Learners receive a short, written passage on a familiar topic and asked to read it out loud carefully. The reading accuracy subtask is timed and measures the learners' accuracy and is reported in the number of correct words read per minute.

Oral reading is EGRA's most direct measurement of fluency. Children are given a short, written passage on a familiar topic and asked to read it out loud "quickly but carefully." Fluency comprises speed, accuracy, and expression (prosody). The oral reading fluency subtask is timed and measures the learners' reading rate and accuracy and is reported in the number of correct words read per minute. This subtask does not typically measure expression.

To learn about how to construct the reading accuracy and oral reading subtasks, see the Subtask Design and Development Annex.

— Common Measures of Reading Comprehension

Reading comprehension includes extracting meaning from text and constructing meaning from text. There are various approaches to assessing reading comprehension, such as sentence tasks (e.g., sentence verification, sentence comprehension), and passage reading with various formats such as multiple choice, open-ended, or retell.

Historically, EGRA reading comprehension asked learners questions connected to the portion of text they have read out loud in the available time on the oral reading fluency subtask. However, this may underestimate learners' comprehension skills, especially when assessing reading in agglutinated languages that have a higher frequency of longer words. In this revision of the EGRA Toolkit, reading comprehension is no longer systematically linked to the oral reading subtask to ensure all learners are able to attempt all reading comprehension items. This can be achieved either by administering separate fluency and comprehension tasks, or ensure learners have sufficient time to read the passage and attempt all items. The recommended common EGRA reading comprehension subtasks are listed in Table 5.

The number of reading comprehension items was also insufficient in the previous Toolkit. We now know that at least 10 comprehension questions, based on at least 2 to 3 different passages that are untimed, provides a more accurate assessment of comprehension levels.

TABLE 5: COMMON READING COMPREHENSION SUBTASKS AND RECOMMENDED GRADE LEVELS

Subtask	Number of Items	Grade 1	Grade 2	Grade 3	Grade 4
Sentence verification	5 to 10 items	✓	✓	✓	✓
Passage reading comprehension	4 to 5 items per passage and at least 2 passages, for a total of at least 10 comprehension questions		✓	✓	✓

ADMINISTRATION OF READING COMPREHENSION

When administering sentence verification or passage reading comprehension, it is important that how learners are permitted to read the sentences and/or passage as well as respond to questions are consistent across the administration. In general, all learners should have the choice to read out loud or silently when measuring reading comprehension; all learners respond to reading comprehension questions in the same language. If reading comprehension is based on a timed reading passage, ensure that learners are allowed sufficient time to complete the passage, even if it is beyond the 1 minute fluency measure.

Reading and understanding a sentence, as is done in the sentence verification subtask, requires less working memory than reading a passage. The sentence verification subtask is appropriate for all grade levels, but particularly well-suited for early readers in grades 1 and 2, who may not have the skills to read a passage. Passage reading comprehension is appropriate for learners in grades 2 to 4.

In passage reading comprehension, learners will be asked to read a text either silently or out loud and when they are finished, the assessor asks comprehension questions about the whole passage. It is possible to use sentences rather than a passage. In this case, the

assessment measures sentence verification, asking the learner to read a sentence and answer a true/false question related to that sentence.

To learn about how to construct the sentence verification and passage reading comprehension subtasks, see the Subtask Design and Development Annex.

READING COMPREHENSION FOR SDG 4.1.1A REPORTING

Reporting to SDG Indicator 4.1.1a per UIS criteria requires at least ten reading comprehension items (assuming 1 score point per item). An EGRA can use sentence verification and/or reading comprehension questions to meet this requirement. If used as the only subtask to measure reading comprehension, the sentence verification subtask may include up to ten items. If combined with a passage reading comprehension subtask, each subtask could include five items each. To achieve ten items using the passage reading comprehension subtask, the EGRA should include at least two comprehension passages. The two passages should contain different topics or stories appropriate to the grade level, as well as different genres of texts (e.g., narrative, expository, etc.). It is not recommended to develop one passage with 10 comprehension questions.

— Additional Measures of Reading Comprehension

Maze measures reading comprehension by asking learners to identify which of three words that makes sense in a sentence. To choose the appropriate word, the learner must understand how words are ordered in a sentence (syntax) and the meaning of the words (semantics) both in the text and in the options provided.

To learn about how to construct the maze subtask, see the Subtask Design and Development Annex.

— Optional Subtasks–Vocabulary and Oral Language

Vocabulary subtasks have not commonly been used in the EGRA framework. However, EGRA users have developed and tested new subtasks to measure vocabulary and oral language, both of which may be beneficial for your assessment purpose. Emerging results

from pilots on the Receptive and Expressive Language Module, or RELM, can be found [here](#) and are the basis for the descriptions provided below. Additional piloting will likely provide improvements on the descriptions below in future years.

TABLE 6: ADDITIONAL OPTIONAL SUBTASKS – VOCABULARY AND ORAL LANGUAGE

Reading Skill (domain)	Subtask	Skill Demonstrated by Learner's Ability To:
Oral language	Listening comprehension picture matching	Select an image that matches a sentence or sentences read out loud by the assessor
	Listening comprehension short story	Respond to comprehension questions about a text the assessor reads out loud
Vocabulary	Expressive vocabulary	Identify a noun or verb that corresponds to an image
	Receptive vocabulary	Identify a single word in an image among the four images
	Semantic fluency	Say as many words as possible in response to an image given by the assessor.

These subtasks were developed under the Receptive and Expressive Language Module (RELM), with the goal of measuring the extent to which students can speak and understand the language in which they are learning to read, measuring key competencies that are building blocks for reading acquisition.⁴ You should work closely with reading and language experts and psychometricians to adapt the RELM to your specific context.

The RELM vocabulary subtasks can be included to assess receptive vocabulary and expressive vocabulary. The first subtask assesses receptive language by asking learners to identify the picture of the target word. Expressive vocabulary assesses the learners' ability to name the object or action in a picture. The semantic fluency subtask for RELM measures a learner's semantic or vocabulary knowledge. Learners are shown a picture and given a minute to name objects or words associated with the picture in the target language. The subtask is optional within RELM and is designed for learners in multilingual environments as a way to determine how well learners speak different languages.

Another way of assessing oral language is RELM’s listening comprehension picture mapping, which has the assessor read out loud one item at a time. Item difficulty increases from a single sentence, to two or three sentences, to a short story on a familiar topic. After reading the item, the assessor shows the learner a stimulus containing four pictures. The assessor asks the learner to select the image that matches the prompt. In addition, RELM’s listening comprehension short story has the assessor read out loud a short story and ask the learner multiple choice questions for comprehension and understanding of oral discourse.

To learn about how to construct the RELM oral language or vocabulary subtasks, see the Subtask Design and Development Annex.

— Universal Design for Assessment Principles

Universal Design for Assessment (UDA) is important to consider when designing an EGRA. UDA is the concept that all assessments should be designed to allow participation of the widest range of learners (Thompson, Johnstone, & Thurlow, 2002). UDA should alleviate the challenges faced when trying to retrofit existing assessments to be more inclusive because it builds inclusivity into the design of the assessment from the start. The seven elements of UDA are described in Table 7. Some of these elements are addressed in the EGRA adaptation process, such as ensuring simple, clear, and intuitive instructions and procedures, and maximizing readability and comprehensibility.

TABLE 7: UNIVERSAL DESIGN FOR ASSESSMENT ELEMENTS

Element	Description
Inclusive assessment population	Tests are designed to allow the participation of all learners in the target population regardless of disability status, language, gender, etc.
Precisely defined constructs	Test constructs are clearly designed and remove irrelevant cognitive, sensory, emotional, and physical barriers
Accessible, non-biased items	Test items are free of content that may be biased against certain groups; all learners in the target population have equal probability of answering correctly
Amenable to accommodations	Tests allow for accommodations without changing constructs
Simple, clear, and intuitive instructions and procedures	Test instructions and protocols are simple, clear, and easy to understand, regardless of learners' experience, knowledge, or language skills
Maximum readability and comprehensibility	Tests use plain language, well-constructed sentences, and contain minimal organizational complexity (e.g., limited sentence length, avoid unnecessary or difficult words)
Maximum legibility	Test items, instructions, tables, figures, and illustrations should be easy to decipher and interpret (e.g., large font, color contrast, sufficient spacing)

Sources: Thompson, Johnstone, & Thurlow 2002; World Bank 2023; Reeves, Sarton, Giulivi, & Oetman 2023.

Integrating UDA principles into your EGRA is good practice in all cases, and especially necessary when:

- Comprehensive disability screenings and individualized learning supports are not available for all learners, which means many learners' disabilities or support needs are not known when designing the EGRA. By incorporating UDA principles and elements into an EGRA, researchers are allowing for more accessibility for all learners, including those with unidentified support needs.

- Learners in the sample are likely to include learners who use different home languages and those from marginalized populations (Lazarus et al. 2022). UDA is closely linked to Universal Design for Learning (UDL), which is based on the idea that learners should have multiple means of engagement, of representation, and of action and expression (World Bank 2023). UDL and UDA should ideally be integrated into an education system simultaneously.

ACCOMMODATIONS AND MODIFICATIONS

Assessment or testing accommodations change how the learner is tested but do not change what is tested.

Assessment or testing modifications are changes to what an assessment measures—the content or construct being tested on a specific subtask.

Although UDA elements can maximize accessibility, it is not possible to make one assessment work for every type of learner. Accommodations and modifications may still be needed to ensure the EGRA reliably measures the true reading abilities of learners. You should engage reading experts with an understanding of UDA or a UDA expert to support the Design and Adaptation phase of your EGRA.

RESOURCES

- World Bank. (2023). Using principles of Universal Design for Assessment (UDA) to design accessible learning assessments. World Bank Group.
- Lazarus, S., Johnstone, C., Liu, K., Thurlow, M., Hinkle, A., & Burden, K. Lazarus, S. S., Johnstone, C. J., Liu, K. K., Thurlow, M. L., Hinkle, A. R., & Burden, K. (2022). An updated state guide to universally designed assessments (NCEO Report #431). National Center on Educational Outcomes.

— Subtask and Language Considerations

In section 1.2.3 Identify the Languages You Will Assess, you determined which languages to consider for the EGRA. In this section, you will begin operationalizing those decisions. Language experts will be important for ensuring the subtasks selected are valid measures of the language abilities of learners. Language experts can help ensure the following considerations are addressed:

- Multilingual learners are accommodated in administration: learners' language should not hinder their demonstration of content knowledge (i.e. literacy skills). This means assessors should speak the language that learners speak and understand to provide accurate oral instructions for EGRA subtasks, the instructions should be designed during the language analysis process.
- Interpretation of assessment data using a multilingual lens: EGRA results should be interpreted in conjunction with linguistic characteristics. For example, learners who are tested in phonological awareness—a foundational literacy skill—may perform better if there are similar sounds in the language they speak at home and the language of instruction. Likewise, learners who are tested in a transparent orthography, or a language that has regular grapheme-phoneme correspondences, are likely to master early reading skills such as decoding more quickly than learners who are tested in an opaque or irregular orthography. This is relevant for both interpretation of results and for benchmarking.
- Communicating assessment results to multilingual populations: How you present results should adhere to UDA, ensuring that audiences with complex linguistic portfolios can comprehend and speak to assessment results. Data should be presented simply and clearly, in a language that the audience speaks and understands, and with visual support. Utilizing presenters who speak and understand the language of the audience will ensure greater understanding.

These considerations should be balanced with the overall purpose of the study as any additional study requirements will likely cost more time and/or resources. Moreover, ensuring that all learners—regardless of the language they speak and understand—are able to fully participate in taking the EGRA is an important element of UDA.

RESOURCES

- [The EGRA: Its theoretical foundation, purpose and limitations](#) (Dubeck, Gove, 2015)
- Example of multilingual EGRA (forthcoming)

— Decision set 2: EGRA Tool Design

This section summarizes the key decisions from the preceding sections. These decisions should be made before continuing through the following sections to avoid confusion.

DECISION SET 2

- 1 Which skills do you need to assess to meet the purpose of the EGRA? There are a variety of subtasks that can be included in the EGRA depending on purpose and population. Prior data might also indicate the skills that learners have mastered (in which case, there is less reason to reassess unless there has been a change in curriculum or the environment); prior data may also indicate where learners have most struggled. With new subtasks being developed by the broader community each year, the options in subtasks continue to grow. The subtasks selected should have a clear use case and appropriate for the context.
- 2 How much time per learner should your assessment take? On balance with selecting the subtasks is the decision on how long a learner can be expected to sit for the assessment. Oftentimes, we want more data than we actually need. Starting with the purpose of the EGRA – and selecting subtasks that meet that purpose – will help avoid overly lengthy assessments without a clear purpose for the information being collected.
- 3 How can your EGRA incorporate principles of Universal Design for Assessment? Are comprehensive disability screenings and individualized supports for learners already available and indicate the appropriate accommodations required when conducting an EGRA? When little information is available, UDA can ensure the EGRA is accessible to all learners including those with unidentified needs.
- 4 In which language(s) will you administer subtasks? Based on the language mapping you may have clarity on which languages you will need to assess. Now you will need to decide which languages will you administer the EGRA subtasks in – based on the purpose of the EGRA study. Consider the following factors:

 - 1 For each language, you will need a language expert, data collectors and analysts comfortable in that language.
 - 2 Additional steps will be required during piloting to ensure the EGRA in each language is of comparable difficulty; whether translations were adequate, and how to aggregate results across multiple languages
 - 3 Additional time for assessments will be required if you decide that an individual learner will be assessed in more than one language, and time for analyzing student-level data across languages

- 4 The language of the instructions for each subtask may be different from the language of the assessment, and the language of the request for consent. This will depend on the language assumptions of children, their parents, teachers and in which language they will most feel comfortable understanding the purpose of the assessment.

If you do not expect to be able to adequately manage multiple language tools, it is better to work with fewer languages and obtain higher-quality data. Remember that exclusions based on language can unintentionally limit the population to whom you want to generalize to.

Determining your Data Collection and Analysis Considerations

— Data Collection Planning

Planning data collection involves several steps. During the Design phase, your data collection planning is high-level – ensuring you have the right team, a schedule for each phase and decisions that will incur costs and time. The aspects of data collection planning that will need to be decided on first include: electronic versus paper-based data collection, timing in the school year and other factors that might dictate when it is safe and appropriate to travel to schools, and planning for recruitment of field teams.

Planning for electronic and/or paper-based data collection

The use of tablets, and to a lesser extent smartphones, has become the norm for EGRA data collection. Electronic data collection improves data quality by reducing errors, increasing real-time quality control and verification of data, and reducing time and costs associated with data entry. Commonly used data collection software includes Tangerine, Open Data Kit (ODK), SurveyCTO, and Myna, among others. In most cases, EGRA data collectors record responses on the tablet while learners use paper stimuli, although some EGRAs have incorporated stimuli on tablets.

There may be circumstances when paper-based data collection is necessary – if tablets cannot be procured, may be seen as intimidating to learners who are not familiar with the devices, they present a security risk for the field teams, or they are simply not allowed by local authorities. You should understand early in your planning if tablet-based data collection is appropriate for your context and if you will need to integrate paper-based data collection, as this will necessitate different planning, training, data entry, and analysis requirements.

Developing your workplan

You should begin planning your entire EGRA workplan, including a schedule for operational data collection, during the Design phase. You should develop a draft workplan that includes a timeline for data collection to make sure that all steps required for your EGRA are sequenced and planned, and that your timeline can accommodate all steps, including potential delays.⁵ An example field work timeline is available forthcoming.

Some considerations when developing your workplan include:

- Academic calendar - at which point in the academic calendar and curriculum is it appropriate to conduct the assessments, based on what your EGRA purpose is? If end-of-year learning levels are desired, then the EGRA data collection should happen as late in the school year as possible, but avoiding times when learners are likely to be absent or when other exams may be administered.
- Length of time in the field - the longer the data collection period is planned, the greater variability in learner performance will be captured in the data. For example, learners assessed within 2 weeks of the second trimester are likely comparable; if the sample consists of learners assessed as early as the end of the first trimester and as late as the end of the second trimester, then the range of learning abilities captured in the data may not be comparable. Length of time in the field is a tradeoff with the size of the team (and subsequently, cost).
- Climate - the effects of climate on both the data collection team's ability to reach schools safely and on learners' attendance should be considered. For example, conducting the assessment during the wet season (i.e., time of year where the majority of a country's or region's annual precipitation occurs) will affect sampling decisions as some schools will become unreachable.

- Availability of assessors and timeline for recruiting – data collection teams, consisting of assessors, supervisors and quality control officers, are a significant factor in the timing of the data collection. Oftentimes, individuals who will work as data collectors will have other employment or commitments. When government staff are working as assessors, their preexisting commitments may dictate their availability to conduct EGRAs. Planning recruitment and selection of the data collection team is an important marker in the timeline.
- Timeline for completion of steps prior to data collection - prior to data collection, you must complete the Adapt phase of your EGRA, including any piloting of tools, revisions, and trainings. Ensure that there is adequate time to complete these steps before the scheduled data collection.
- Additional considerations when planning field work:
- Identify government officials, social workers, counselors, or other key figures in the community for any necessary referrals or approvals to visit schools.
- Allow time needed to obtain informed consent (i.e., if parents or guardians are not accessible)
- Safety and security protocols, including any coordination with local security authorities

These planning considerations should inform components of the assessor training and practice in schools.

RESOURCES

- Tools for field planning
- Workplan template and example
- **RERA Toolkit**, pages 108 to 109, for guidance on school community stakeholders to consult and topics to discuss during site planning.
- **Safer Learning Environments Toolkit**, page 38, for field visit planning considerations, including acquiring research permits.
- Field Planning Checklist: checklist for teams to help ensure they have completed necessary steps for conducting primary research (**Safer Learning Environments Toolkit, Annexes**, p. 127)

— Determining Additional Survey Needs

To understand how the EGRA results vary across different learner, school and community characteristics, there are often surveys conducted alongside the EGRA. Surveys with learners, teachers, parents, community members, and government officials often included as part of an EGRA data collection. Based on the disaggregations and trends that you are interested in examining, you will likely need at least a learner survey.

A learner survey typically captures the following information:

- **Gender/sex:** To understand gender disparities, if any.
- **Mother tongue/Language spoken at home:** To understand a learner's home language environment.
- **Age:** Assess whether the learner is over-age or under-age for the grade they are attending.
- **Family Composition:** Understand the structure of the learner's home life and learning supports
- **Previous Schooling Experience:** Understand the learner's educational history, including any gaps or interruptions in schooling.
- **Access to Learning Materials:** Identify disparities in access to learning materials and technology, which can affect remote learning and homework assignments. Teachers and school administrators can answer questions about the availability of materials in the school, while learners can answer questions about the availability of materials at home.
- **Socioeconomic Status/Home Background:** Obtain information about family assets, parental education, and other indicators of socioeconomic status to identify economic disparities.
- **Parental Involvement:** Assess the level of parental involvement in education.

— Planning EGRA Data Use

You should have already identified your EGRA purpose, which is the first step in planning EGRA data use. Now, you should decide whether benchmarks are needed and how to obtain/establish them, the format and structure of the results for the stakeholders, and the team that will be responsible for writing and presenting the EGRA results.

Benchmarks

Benchmarks are quantitative values representing a minimum score needed by a learner to be considered to have reached a minimum proficiency level. The benchmark operationally defines what it means to have ‘minimum proficiency’ in the SDG 4.1.1a indicator.

Benchmarks can be used for all 3 proposes of an EGRA since the benchmark provides an externally defined threshold or criteria. It helps answer the question ‘what level of performance is good enough?’

When designing an EGRA with plans to set benchmarks, the following issues will determine whether a benchmark is feasible:

- Who is in the sample: if learners across geographies and languages are included, the benchmarks should be reflective of these variations in the sample. A statistician will be important to help determine the appropriate benchmarking method to use. In the process of setting benchmarks along with EGRA implementation, researchers should consider when and if it is sensible to adjust benchmarks for sub populations (e.g. learners of different linguistic groups, learners of different ages or years of formal schooling, learners with disabilities, or learners living in crisis and conflict).
- Which skills do we want to know if learners are meeting expectations: Not all subtasks in EGRA are relevant for benchmarking. SDG 4.1.1a requires minimum proficiency in reading comprehension to be reported. Depending on the purpose, benchmarks on precursor skills may be warranted if they are helpful for teachers, curriculum experts, or other stakeholders to know if learners are meeting the minimum expectations.

Presentation of findings: format and structure

While it seems too early to plan for reporting results at this stage, it is important to have a general plan so that the breadth and depth of the data you collect is aligned with the depth and breadth of what you want to – or are required to – report. The following are some considerations to make at this stage:

- What are the technical skillsets of the stakeholders to whom I have to present the findings: policy experts, researchers, teachers, curriculum designers, teacher training experts all have varying levels of expertise understanding and using data, graphs, and technical language. To ensure the format and structure you use to present the results is accessible to your target audience, ensure you first understand what their expectations and skillsets are.

- What types of formats will meet the purpose of the study and stakeholders needs: consider a range of options to meet stakeholders needs, including written reports, summaries, infographics, slides, videos, and more. The level of details, breadth and depth of the findings presented, and the level of technical details included will vary based on the format you select and the audience's needs.
- If writing a report, what is the appropriate writing level and length: previous EGRA reports have ranged from 40 pages to over 200 pages. Understanding what your audience(s) needs are and how to best present the findings to meet their needs, may involve a technical report in addition to other formats. Consider who the specific audience is for a report, what level of writing (technical language, accessibility, reading level, language) and length is appropriate. The majority of EGRA reports in the past have tended to be written for a research audience but presented to a policy audience.
- Will we present findings in a live (or virtual) session: live presentation of findings to stakeholders often is the only time they may engage with the findings. In this case, the report writing team may spend more time and attention on the slides or handouts or talking points for the presentation than on a written report.
- Are stakeholders open to engaging with the findings as they emerge, by participating in a Data Dive: a best practice in sense-making results is to hold data dives, or sessions where a range of stakeholders sit with the EGRA team (analyst, statistician, designer, manager) and engage with the raw results. Results may be as simple as a table of descriptive statistics like the average score and the number of learners; they may be more complex such as a series of statistical tests and their results. During a data dive, the analyst presents the results and asks stakeholders to engage in a meaning-making conversation: what makes sense about this result? What new questions does this result raise for you? What other factors does the result not take into account? What other ways should we look at this question? Data dives will surface questions early – and before results have been written or presented formally – and ensure that the final analysis and presentation of findings include a thorough analysis.
- What technical documentation to plan for: If you plan to report to SDG 4.1.1a, follow the guidance for the submission package. For all EGRAs, the following technical documentation should be planned for (more details in the Reporting and Use section):
 - A test development report explaining the procedure used to develop, review, and select items
 - A sampling report detailing the sample size calculation and the process of sample selection

- Technical annex tables presenting the descriptive and statistical results of the analyses
- A benchmarking report that summarizes the benchmarking methods and results

Data presentation and application team

You should determine who will be responsible for writing or presenting results. Oftentimes, the analyst who conducted the analysis of the data is closest to the results; however, they will need a partner who can help ask critical questions, identify the level of details that are appropriate and necessary to share with the audience, and ensure that you narrate a complete story rather than compile many isolated facts. At this juncture, ensure there is clarity – and expertise – among the team to present the results in a usable format and structure to stakeholders.

DECISION SET 3

The following questions should be answered before moving ahead in the EGRA process:

- 1 Will tablets be feasible to procure? How many and what are the requirements?
- 2 Which data collection software is best suited to our needs?
- 3 Is field work scheduled based on school calendar, climate, timing of pilot studies and ethics approvals?
- 4 Who is responsible for recruiting and hiring field teams and what is the process for vetting, training and selection?
- 5 What approvals are needed for field teams to enter schools?
- 6 What is the right window for field work considering academic calendar, curriculum pace, climate, availability of field teams?
- 7 What are the safety and security protocols to follow during field work? Who is defining these protocols, training, and ensuring any concerns are identified and addressed?
- 8 Which skills and for which learners do I need benchmarks so that the data meet the purpose?
- 9 What do the stakeholders want to know about the EGRA study and to what level of detail?
- 10 In what formats and structures will we share findings to meet the needs of our audiences?

02 Adapt

Adapt is the second step in an EGRA. Adaptation refers to the process of contextualizing the EGRA framework for a specific population of learners in a specific location while retaining the principles of reading skills measurement. The Adapt phase results in an EGRA tool that is field tested, piloted, and finalized for data collection. This chapter includes information on:

- The adaptation workshop
- Development of accommodations and modifications
- Adaptation of subtasks
- Field tests and pilots

The team members and technical capacities that will need to be involved in this phase include...

- EGRA Manager
- Reading Specialist and Language Expert
- Curriculum Expert
- Braille and/ or Sign Language Specialist (If relevant)
- UDA expert
- Psychometrician and/or Statistician

DEFINITIONS FOR ADAPTATION, MODIFICATION, AND ACCOMMODATION

It is important to differentiate between three terms used in Sections 2.1 and 2.2.

Adaptation refers to the process of taking the EGRA framework and contextualizing it for a specific population of learners in a specific location. An assessment accommodation is a change to how the learner is tested, while a modification is a change to what is tested. Accommodations and modifications are generally introduced during the EGRA adaptation process. Accommodations should not change the validity of the assessment or its comparability. Because modifications change the content or constructs being tested, they may change the validity of the assessment and may make an assessment incomparable.

EGRA is a framework that can be customized for specific contexts, populations, and purposes, and therefore must be adapted to each purpose and context. Adaptation refers to the process of contextualizing the EGRA framework for a specific population of learners in a specific location. Adaptation can involve developing an entirely new EGRA tool or reviewing and revising an existing tool to be fit-for-purpose. Unless an EGRA is being used with the exact same population under conditions identical to a previous administration, researchers should go through the adaptation process to review and validate existing items, develop new items, and consider the addition of any accommodations or modifications relevant for the learner population that will be assessed.

When adapting the EGRA, the selected subtasks must be appropriately leveled - meaning the tasks are of appropriate difficulty; the subtasks must be relevant for the purpose; and the whole assessment must be of appropriate length for the learner's developmental age.

Table 8 defines the differences between adaptation of new instruments and adaptation of existing instruments. If a country-specific EGRA is being developed for the first time, it is considered an adaptation of new instruments; if an EGRA has already been conducted in-country, then the workshop is an adaptation of existing instruments.

TABLE 8: DIFFERENTIATING BETWEEN ADAPTATION OF NEW AND EXISTING EGRAS

Adaptation steps when using a new EGRA	Adaptation steps when using an existing EGRA
Language analysis	Language analysis (optional)
Item selection	Item reordering/randomization
Verification of instructions	Verification of instructions
Pretesting	Pretesting

EGRA ADAPTATION IN PRACTICE

A project seeking to implement an English-medium EGRA in Zimbabwe in 2024 may choose to use an English-medium EGRA from Zimbabwe, developed in 2015, as a starting point. Because the existing tool is nine years old, likely outdated, and potentially only relevant for certain provinces in the country, the project should go through an adaptation process and hold an adaptation workshop to (a) determine the relevance of each subtask on the existing tool to the specific project's purpose and goals, and select additional subtasks if needed; (b) review, develop, modify, and translate instructions on the existing tool for the location-specific context, including developing instructions in different regional languages; (c) review specific items on the existing tool to align with the project's or current national curriculum, (d) review any semantic, idiomatic, or conceptual themes and make adjustments based on the context; and (e) review and update administration protocols, and introduce accommodations if relevant, to ensure that administration protocols are appropriate for the learner population that is part of the project. If the project is reaching out-of-school youth who have never been in school or been assessed, and are likely unfamiliar with a testing environment, the project team may determine that, instead of one minute on timed subtasks, learners should receive three minutes to give them ample time to demonstrate their knowledge. This would be an accommodation that is added during the adaptation process.

Conducting an Adaptation Workshop

The first task under Adapt is to conduct an adaptation workshop. The purpose of the adaptation workshop is to make the EGRA framework fit for its intended purpose(s) and context, such as study population and language. There are four main goals for an adaptation workshop:

- To introduce government officials and local curriculum and assessment specialists to the research that underpins the EGRA framework.
- To adapt the EGRA framework to the contexts, populations, and purposes, with specific tasks such as:
 - Selecting subtasks to include in the tool;
 - Translating instructions for the entire tool and for specific subtasks;

- Selecting and/or reviewing items based on specific curriculum, including reviewing items and passages to ensure they reflect locally and culturally appropriate words and concepts.
- To verify the face and content validity of the EGRA tool;
- To review discuss the ethics of research and working with human subjects, especially children, and to establish procedures for informed consent (adults) or assent (children)

At the end of your adaptation workshop, you should have an EGRA tool ready for piloting.

— Adaptation Workshop Agenda

Typically, an EGRA manager plans the adaptation workshop and engages others in the EGRA Team to conduct the workshop. Adaptation workshops are often 5-6 days and include different plenary and small group working sessions. The duration of the adaptation workshop and specific sessions will depend on several factors, including existence of a previously used EGRA for the given language/country/grade; number of subtasks to be tested; number of languages to be tested; need for additional questionnaires and instruments; and purpose and audience of the workshop.

The most difficult part of adaptation is usually story writing, so it is important not to leave this subtask until the last day. This step involves asking local experts to write or review short stories using grade-level appropriate words, grade-level appropriate sentence complexity, and grade-level appropriate grammar. Local experts also need to write appropriate comprehension questions to accompany the stories. Both the stories and the questions often need to be translated into English or another language for review by additional early grade reading experts and revised multiple times in the language of assessment before they can be finalized.

Additionally, the adaptation workshop agenda should include time for pretesting some or all of the EGRA tool. This generally means that workshop participants will visit nearby schools to use the draft EGRA tool content with learners. This pretesting provides valuable feedback to the workshop participants and helps identify obvious changes that may be needed, such as revisions to ambiguous answer choices or overly difficult vocabulary. If the adaptation workshop cannot be held in an area where pretesting is possible, you should consider arranging for a pretest after the workshop. The pretest provides important

feedback to guide the finalizing of an EGRA tool for piloting and can help the pilot be more successful. For more information on how to conduct an adaption workshop, please see the Adaption Annex.

— Adaptation Workshop Participants and Materials

Having the right people participating in the workshop is critical – they are going to both provide content based on their expertise and ensure buy-in based on their role. Whether you are creating an entirely new EGRA tool or reviewing an existing EGRA tool, you will need broad expertise on the technical content and the context. Keep in mind that some participants may be part of multiple groups, such as a government official that is also a literacy subject-matter expert.

Who to include in the Workshop

- **Government officials from national and local levels:** These participants are experts in curriculum development and assessment and can offer guidance on whether the content is consistent with the national curriculum. Government officials are also important participants, as they will be responsible for evidence uptake in policymaking, if relevant for your purpose. Also, including government officials in an adaptation workshop strengthens their capacity to design assessments, which is important for sustainability.
- **Subject-matter experts:** This group includes local, national, and international language and literacy experts. These participants can advise on subtask and item selection, provide translations for specific languages, and support story writing and development. Members of this group should understand reading research and pedagogical/instruction processes, and at least one should have a background in reading assessment tools and instruction. Sometimes a member of this group is also a workshop facilitator.
- **Teachers:** Academics (reading specialists, in particular) and current or former teachers (with a preference for reading teachers) are also important participants. They can verify that the content of reading subtasks are grade-level appropriate and bring their context-specific and classroom-level experiences to the adaptation process.
- **Psychometricians, statisticians, or assessment development experts:** You should also include at least one technical participant who can advise on the psychometric considerations for each subtask and for the overall EGRA tool. Members of this group should have a background in education survey research and in the design of

assessments/tests and should be experienced in statistical programs, such as SPSS or Stata. Sometimes a member of this group is also a facilitator, and often a member of this group will later lead assessor training and data collection.

How many people to include in the workshop

The number of participants will depend, in part, on the number of languages involved in the adaptation process for a given study, but you should expect to have about 30 participants. Depending on the languages that will be used during the workshop, you should also plan to have interpreters attend. You should provide them with sufficient background content on EGRA so they can effectively communicate technical concepts.

ADDITIONAL WORKSHOP PARTICIPANTS FOR ADAPTING EGRA FOR LEARNERS WITH DISABILITIES

When adapting EGRA for marginalized learners or learners with disabilities, additional specialists should participate in the adaptation workshop. These include:

- Civil society organizations for marginalized groups or key constituencies (e.g., Organizations of Persons with Disabilities or Indigenous Peoples groups);
- Government officials who oversee education for specific populations, such as learners with disabilities, girls, and refugees
- Teachers of learners with disabilities
- Subject-matter experts in inclusive education, sign languages, braille, etc.

Who facilitates the workshop

At least two experts will facilitate the workshop. Workshop facilitators generally come from one of the workshop participant groups listed above. Facilitators should have a solid understanding of the EGRA framework, be well-versed in the pre-adaptation decision-making, and be skilled at leading trainings in different contexts and with different types of participants. Ideally, facilitators should also have proficiency in the language(s) that will be used during the adaptation workshop.

Materials for the workshop

The materials you will need for your adaptation workshop are similar to those you would need for other types of training. A recommended materials list is provided in Table 9.

TABLE 9: RECOMMENDED EGRA WORKSHOP MATERIALS⁶

Workshop Materials
Paper and pencils with erasers for participants
An LCD projector, whiteboard, and flip chart (if possible, the LCD projector should be able to project onto the whiteboard for simulated scoring exercises)
Current national or local reading texts, appropriate for the grade levels and languages to be assessed (these texts will inform the vocabulary used in story writing and establish the level of difficulty)
Paper copies of presentations and draft instruments
A presentation on the EGRA-related reading research, development process, purpose, uses, and research background
Samples of EGRA oral reading fluency passages, comprehension questions, and listening comprehension questions from other countries or, for modification, copies of the previous in-country EGRA instrument

— Adaptation Workshop and Validity

TYPES OF VALIDITY

Face validity: How well a tool overall addresses the concepts and domains it is intended to measure.

Construct validity: How well the skills measured on the assessment relate to each other based on existing research.

Content validity: How well items measure the intended skills.

When adapting an EGRA, it is important to ensure its validity. Validity refers to the correctness of measures and the appropriateness of inferences or decisions based on the test results for a given purpose. There are three types of validity: face, content, and construct. You should check the validity of your EGRA tool during the adaptation workshop and piloting.

During the adaptation workshop, participants will verify the face and content validity of the EGRA tool. Participants ensure face validity by selecting subtasks and items that are appropriate for the learner population and context.⁷ Government officials, curriculum experts, and other relevant groups build content validity by reviewing proposed subtasks and making judgments about the appropriateness of each item type for measuring the early reading skills of their learners. To do this, they consider the intended purpose of the EGRA subtasks and make judgments about the appropriateness of each item type for measuring the early reading skills of their learners, as specified in curriculum statements or other guidelines for learning expectations or standards.⁸ After the adaptation workshop, you pilot tests the EGRA with a small subset of learners. Statisticians and psychometricians analyze the pilot data to assess the EGRA's construct validity. See Adaptation Annex for more information.

Identifying Accommodations and Modifications

Accommodations and modifications are changes introduced to an EGRA to ensure a test is accessible and fair for all learners. You should engage appropriate experts and take care to differentiate between changes that are accommodations and changes that are modifications, because this can affect the validity of the tool and comparability of learners' results. You should have a clear understanding of whether changes are accommodations or modifications in the context of the assessment, particularly because accommodations and modifications can interchange depending on how they are applied and affect the generalizability of results to a wider population or sample group. Additionally, you should take care to document any accommodations or modifications introduced to an EGRA during the adaptation process.

Accommodations

Assessment or testing accommodations change how the learner is tested but do not change what is tested. Accommodations are meant to increase accessibility and eliminate obstacles that learners may encounter, but accommodations should not make the assessment easier or harder. During the adaptation workshop, you should work with experts to identify the accommodations that will be provided to learners; ideally, you will also allow enumerators to capture which accommodations were used by a learner, if any, so that this information shows up in the dataset.

Table 10 provides examples of reasonable accommodations, by category. Though these accommodations should not change the validity or comparability of an assessment, you should work closely with psychometricians and subject-matter experts when introducing any accommodations.

TABLE 10: EXAMPLES OF EGRA ACCOMMODATIONS FOR LEARNERS WITH DISABILITIES

Category	Example EGRA Accommodations
Presentation	Larger font (16, 24, or 32 point) or braille stimuli (uncontracted or contracted) Magnification devices, braille mats, or typoscopes Instructions provided in sign languages Audio amplification devices Cues on stimuli, such as arrows or stop signs, to know where to go for the next item Increased spacing between items or sentences, or fewer items on each stimuli page Repeating instructions
Response	Communication devices (language board, speech synthesizer)
Setting	Low-distraction environment (far from windows or doors) Adaptive lighting or acoustics Adaptive furniture
Timing/scheduling	Extended time for subtasks (2 or 3 minutes) and for pauses (6-15 seconds) Longer breaks between subtasks Shorter testing sessions
<i>Sources. IRIS Center 2023; University of Kansas 2023; Reeves, Sarton, Giulivi, & Oetman 2023; World Bank 2023.</i>	

ACCOMMODATIONS FOR LEARNERS IN CONFLICT AND CRISIS CONTEXTS

Learners who are affected by trauma, particularly as a consequence of crisis or conflict, may benefit from assessment accommodations. Traumatic events affect people in different ways. In some situations, trauma may limit a learner's focus or attention span. Longer assessments may also result in more learner errors later in the assessment session. Low performance may be more an artifact of stamina or perseverance instead of reading ability. For these reasons, it may be helpful to create a procedure that allows trauma-affected learners to complete a modified assessment with fewer items during an assessment session. These learners' responses on the assessment can then be flagged so information can be incorporated into analysis. Decisions about such administration procedures should be established with the expert panel during the adaptation workshop.

In other situations, learners affected by trauma may not need to have shorter EGRA sessions but instead, have more time to complete the assessment. For example, it may take these learners longer to focus on the assessment activity or feel comfortable with the assessment process. This may be particularly true for girls and for children facing malnutrition, who may be silent at the start of assessment activities. It may take a few minutes for a learner affected by trauma or facing severe attention issues because of malnutrition to attempt assessment tasks. The accommodation here would be to give children additional time and, potentially, more support during the introductory discussion. Again, these learners' assessment responses can then be flagged so information can be incorporated into analysis.

Modifications

While assessment accommodations are defined as changes in how a learner is tested, modifications are changes to what an assessment measures—the content or construct being tested on a specific subtask. For EGRA, this may mean changing items, administration protocols,⁹ or potentially, instructions (World Bank 2023). It could also mean introducing an additional element into the subtask that changes the skill being tested. An example of a modification could be making items within a subtask easier or harder, such as including grade 1 items on a grade 2 EGRA; or changing a subtask's administration protocols, such as providing multiple-choice response options on a reading comprehension subtask. Another example could be placing a corresponding drawing in front of the learner during the listening comprehension subtask.

The modifications listed below are considered modifications only when provided to some, and not all, learners.

- Easier or harder content within subtasks
- Reduced number of items on subtasks
- Multiple choice options for listening comprehension and reading comprehension subtasks
- Changes in administration protocols, such as providing additional or unlimited time to measure only accuracy instead of fluency
- Providing visual aids, such as pictures, to accompany listening comprehension or reading comprehension subtasks
- Administering bilingual bimodal subtasks for learners who are deaf or hard of hearing

Because modifications change the measured construct, a modified assessment or subtask is no longer comparable to an unmodified assessment or subtask. Because of this, you should only use modifications when they are necessary to appropriately measure learners' skills and knowledge. It is important to also consider the EGRA's purpose when introducing modifications - if an EGRA is for a classroom-based assessment, modifications can be used for EGRAs meeting the purpose of a classroom-based assessment or project evaluation, but they must be carefully considered if an EGRA is being used for a national assessment since modifications make the EGRA results incomparable it is more challenging to develop modifications for an EGRA intended for a national assessment of foundational reading skills or donor indicator reporting. You should be careful to not create a parallel assessment system for learners with disabilities when introducing modifications, either intentionally or as an unintended consequence. Please see the Adaptation Annex for more information on accommodations and modifications for sign language, braille, and screening and identification.

Creation of Equivalent Test Forms

As mentioned earlier in this section, adaptation can involve modifying an existing instrument that was previously developed for a given language. If there is no concern about test leakage (i.e., if teachers have limited access to EGRA instruments and it is unlikely that learners will become familiar with a particular form of the assessment), the same

instrument can be used across multiple time points. However, if leakage is a concern, multiple assessments (or test forms) will be necessary to have measure changes in performance. To ensure that valid comparisons of results can be made across assessment forms/administrations, instruments must be modified in a way that creates new forms that are as equal as possible in difficulty to the original form. The term equivalent test forms refers to tests that are intended to be of equal difficulty (and thus directly substitutable for one another).

For more information on the creation of multiple test forms, see Adaptation Annex.

EGRA Learner Instructions and Protocols

The EGRA framework includes a standard set of instructions read to learners and recommended assessment protocols that you should consider when adapting an EGRA tool. This section briefly describes the standard EGRA instructions, the early stop - or autostop - rule, and the pause rule.

— EGRA Learner Instructions

The EGRA framework contains standard instructions that can be used for each subtask and that must be translated into language that will be used when administering the assessment with learners. This language might differ from the language being tested. It is important that instructions are not significantly modified during the adaptation process, because they were developed based on evidence and research.

See the Subtask Design and Development Annex for standard instructions that you can use in your adaptation workshop.

— Early Stop Rule

The subtask early stop rule, sometimes called the autostop rule, applies when a learner answers incorrectly a predetermined number of items at the start of a subtask. The early stop rule exists so that a learner is not required to continue a subtask if they have demonstrated that they do not have the required skill. The early stop rule reduces learner

frustration and also helps reduce the time it takes a learner to complete the EGRA. If a learner triggers the subtask early stop rule, a learner receives a zero score on the entire subtask, you end the subtask, and you move to the next one.

Table 11 shows the standard subtask early stop rule protocols by subtask.

TABLE 11: EARLY STOP RULES FOR EGRA SUBTASKS

Subtask	Stop Rule at Item Number
Letter sound identification	10
Letter name identification	10
Initial sound identification	5
Phoneme segmentation	5
Syllable identification	10
Syllable segmentation	5
Familiar word reading	5
Nonword reading	5
Oral reading (accuracy)	None
Sentence verification	None
Passage reading comprehension	None
Listening comprehension	None
Dictation	None

In some cases, you can use an assessment autostop rule. An early stop rule ends the administration of the EGRA after the learner has an early stop on the first two or three reading subtasks (e.g., letter sounds, syllable identification, or nonword reading). After applying the early stop rule, the succeeding subtasks are assumed to be a zero score. The

rationale for an assessment autostop rule is similar to the subtask autostop rule - it both improves the assessment experience for the learner by reducing potential frustration and it also reduces the time a learner needs to complete the EGRA.

For more background and evidence on the rationale for the early stop rule, see the Subtask Design and Development Annex.

— **Pause Rule**

The pause rule applies when a learner hesitates or is stuck on an item for a predetermined number of seconds, at which time the assessor marks the item incorrect or no response and tells the learner to move to the next item. The pause rule serves similar purposes to the early stop rule by reducing learner frustration and reducing the time it takes a learner to complete the EGRA. The pause rule also gives the learner the opportunity to attempt additional items, rather than remaining ‘stuck’ on an item they do not know. Table 12 provides standard pause rules by EGRA subtasks.

TABLE 12: PAUSE RULES FOR EGRA SUBTASKS

Subtask	Pause Rule
Letter sound identification	3 seconds
Letter name identification	3 seconds
Initial sound identification	3 seconds
Phoneme segmentation	3 seconds
Syllable identification	3 seconds
Syllable segmentation	3 seconds
Familiar word reading	3 seconds
Nonword reading	3 seconds
Oral reading	3 seconds
Reading comprehension	10 seconds
Listening comprehension	10 seconds
Dictation	15 seconds

Pretest and Pilot

Pretesting and piloting of EGRA subtasks and instruments are necessary steps to ensure that your EGRA is valid and reliable. Pretesting and piloting serve to:

- Test the validity and reliability of your EGRA tool
- Test the administration logistics, such as the cost, time, and protocols
- Provide capacity strengthening to enumerators that will participate in the implementation of the EGRA

Because you will need to train data collectors - also called assessors - to pretest and pilot parts or all of an EGRA, please refer to the Implementation Annex for guidance on how to develop and implement an assessor training.

— Pretesting your EGRA

During or after your adaptation workshop, you will pretest EGRA subtasks with learners. Pretesting allows you to validate components of the EGRA tool in your specific context.

The overall goal of pretesting is to determine whether the EGRA instruments function as intended. Specifically, you will learn:

- If EGRA subtasks and learner stimuli appear properly on the paper form or tablet-based data collection application;
- If learners understand the instructions for the subtasks and the included examples;
- If learners understand how to complete subtasks;
- How long it takes for a learner to complete each subtask and the entire EGRA;
- Whether the order of questions influences learners' answers on listening comprehension, listening comprehension short story, sentence verification, and passage reading comprehension.

You should plan to pretest with a small sample of learners (i.e., 12 to 50) (Ruel, Wagner, & Gillespie 2016) that are similar to the population that you will include in your operational data collection. It is important to use the data collection method (i.e., paper or electronic) that will be used in pilot and operational data collection.

— Piloting your EGRA

A pilot test is a small-scale preliminary study before conducting a full EGRA - often referred to as the operational data collection. A pilot can help you learn similar things to a pretest, and because a pilot is generally performed with a larger sample, it will also allow you to:

- Ensure the reliability and validity of the instrument through psychometric analysis;
- Obtain data on multiple forms of the instruments, for equating purposes;¹⁰
- Review data collection procedures, such as the functionality of the tablets and e-instruments, along with the procedures for uploading data from the field;
- Review the readiness of the materials; and
- Review logistical procedures, including transportation and communication among assessor teams, field coordinators, and other staff

Your pilot test logistics should be as similar as possible to what you anticipate for your operational data collection, though some considerations related to your sample, such as regions, districts, schools, learners per grade, will likely be different.

Table 13 outlines the key differences between the pilot test and the full data collection.

TABLE 13: DIFFERENCE BETWEEN EGRA PILOT AND OPERATIONAL DATA COLLECTION

Component	Pilot	Operational
Purpose	To test the reliability, validity, and readiness of instrument(s) and give assessors additional practice	To complete a full assessment of sampled schools and learners
Timing	Takes place after adaptation	Considers the time of year in relation to the academic calendar or seasonal factors (holidays, weather); also considers post-pilot adjustments and instrument revisions
Sample	Convenience sample based on target population for full data collection	Based on target population (grade, language, region, etc.)
Data	Analyzed to revise instrument(s) as needed	Backed up throughout the data collection process (e.g., uploaded to an external database) and analyzed after all data are collected
Instrument revisions	Can be made based on data analysis, with limited re-piloting after the changes	No revisions are made to the instrument during data collection

When you conduct a pilot study, you will need to make sure that the number of learners you assess is sufficient to establish test validity and reliability. For more information on pilot study sample requirements, see the Adaptation Annex.

Finally, after completing pilot data collection, the psychometrician or analyst will need to analyze the data to measure the test reliability and validity. For more information on analyzing your pilot EGRA data, see the Adaptation Annex.

You will also need to reflect on the data collection procedures, readiness of the materials, and logistics, and update your data collection plan based on what you learn. After your pilot study and analysis are complete, and after your team has finalized the EGRA tool and your data collection plan, you are ready to move to the implement step of the EGRA process.

03 Implement

Implement is the third step in an EGRA. Implement refers to using a piloted and finalized EGRA tool to collect data from learners. This chapter includes information three critical parts of the implement step:

- The assessor training
- Operational data collection
- Monitoring and quality assurance

The team members and technical capacities that will need to be involved in this phase include...

- EGRA Manager
- Reading Specialist
- Psychometrician and/or Statistician

Selecting Your Assessment Team

Before your assessor training and your operational data collection, you will need to recruit individuals to your assessment team. Although the composition of your assessment team will depend on your specific context, the most common roles on a team are assessors, supervisors, and quality control officers (QCOs). Note that sometimes these responsibilities are combined into one role, such as an assessor who is also a supervisor, or a supervisor who also conducts the responsibilities of a QCO.

- **Assessors** - also called data collectors or enumerators - are primarily responsible for administering the EGRA and any surveys to learners or other school staff. They are also often responsible for selecting learners to participate in the EGRA; setting up the assessment space; administering inter-rater reliability (IRR) assessments; reporting any required information to supervisors and QCOs; and any other instructions and responsibilities assigned by supervisors and QCOs.
- **Supervisors** are primarily responsible for coordinating school visits with school leaders or administrators; introducing the assessor team and the EGRA purpose and process to the school leaders and teachers; managing assessors' daily activities; providing on-site

or remote troubleshooting on tablets and assessments; ensuring assessment quotas are met at each school; holding daily debriefings with assessment teams; collecting and submitting any monitoring forms required; and reporting on progress to the overall assessment or project manager. Supervisors may also be responsible for ensuring the security of all assessment materials, such as tablets.

- **QCOs** monitor and ensure quality of data collected through the EGRA and student surveys; visit schools to observe assessors and conduct feedback meetings; track any challenges reported that affect data quality (for example, if a tablet ran out of power mid-assessment or a learner asked to stop an assessment early) for eventual use in data cleaning; visit schools to observe assessors and conduct feedback meetings; and report on progress to the overall assessment or project manager.

Assessors can be independent data collectors with previous experience in tablet data collection and/or with EGRA. They can also be education sector stakeholders, such as government education officials, teachers, or coaches. The profile of the assessors you recruit will depend on the specifics of your EGRA - such as the language of administration, the language that will be assessed, the length of the data collection, the time of the data collection, and your overarching goals of the data collection. For example, if your data collection will take part over a month and/or during the school year, it may not be feasible for government education officials or teachers to serve as assessors. Alternatively, if one of your goals is to strengthen the capacity of education sector officials to administer EGRAs, or if you believe that involvement of qualified government officials will increase the likelihood of results uptake, your recruitment will necessarily include these individuals. Regardless of their profile, you will want to ensure that all assessors have experience with or willingness to learn about how to collect data from learners. Critically, assessors must be fluent in the language in which the assessment will be administered.

For assessor team member candidates, as for any job candidate, you should consider any conflicts of interest, especially for education sector stakeholders like government education officials or teachers. You should be sure to discuss any potential conflicts of interest with candidates, train the assessor team on research ethics and the importance of neutrality in data collection, and conduct careful observation of assessors throughout the training and operational data collection to mitigate any bias.

Training the Assessment Team

After you have finished recruiting your assessment team members, you will need to train them to conduct the data collection. The training will prepare the team, including assessors, supervisors, and QCOs, to conduct the EGRA data collection - both to manage data collection logistics and to administer the EGRA and any additional survey tools.

Note that this section provides guidance on how to plan and conduct an EGRA assessor training, but this is not an assessor or supervisor manual and is not meant to be distributed directly to your assessor team. An illustrative assessor manual - sometimes referred to as a Test Administration Manual (TAM) - can be found in the Implementation Annex.

The main learning objectives of the assessment team training are:

- Participants understand the theory behind the EGRA framework, the purpose of each subtask and data collection tool, and the purpose and end use of the data that will be collected;
- Participants know how to accurately and effectively administer each component of the EGRA, including how to navigate the tablet and the data collection software; and
- Participants know what each team member's roles and responsibilities are, and how to implement the data collection logistics.

Before conducting training for your assessment team, you should ensure that the following tasks are completed:

- Finalize your EGRA tool and other surveys in paper form.
- Program your EGRA tool and other surveys into data collection software, if you will be administering them electronically.
- Identify your assessment team training facilitators. The facilitators should understand EGRA and its administration and have experience training data collectors. Ideally, facilitators should be proficient in the language being assessed and in the language that will be used during the training.
- Identify your training venue. Ensure that your training venue has enough space for small group and pair practice.

- Prepare your assessment team training agenda, including identifying who will facilitate each session and what materials you need for each session. (See the Implementation Annex for an illustrative training agenda)
- Create all required training materials. These could include PowerPoint presentations, student stimuli booklets, a Test Administration Manual, assessor cheat sheets, and supervisor assessment tracking and QC forms.
- Procure training materials, including tablets, flip charts, pens, notepads, and a projector. If using tablets, you should create a registration form that requires trainees to check in and check out their tablets each day; this is especially important for trainings with 10 or more trainees.
- Prepare tablets for data collection. You should make sure that all tablets have the necessary software and electronic versions of the EGRA and any additional surveys are correctly loaded.

TRAINING OF TRAINERS, SUPERVISORS, ASSESSORS, AND OTHER DATA COLLECTION TEAM MEMBERS

All members of your assessment team will need training. Depending on the scope of your data collection, you may need to conduct a Training of Trainers (ToT) for facilitators prior to your assessment team training. A ToT uses the assessment team training agenda as a guide, but focuses on preparing facilitators to conduct the training, and ensuring they understand the purpose and logistics of your EGRA and data collection.

You should also plan to provide supplementary training sessions to supervisors and QCOs Quality Control Officers (QCO) that are specific to their roles and responsibilities. For example, a supervisor/QCO training session should emphasize logistics of the data collection and specific daily reporting requirements. Often, these sessions take place during or after an assessment team training.

You should expect your assessment team training to last at least five to six days. The length of the training will depend on factors such as: the number of tools that will be used in the data collection, the previous experience of your assessors, the number of assessors to be trained, the number of training facilitators, and the budget and time you have available. For example, if your data collection includes an EGRA in one language and a short student survey, and your assessors already have experience using tablets for data collection and/or

administering EGRAs, five days is likely sufficient. However, if your data collection includes, for example, an EGRA in two languages, or if assessors will administer EGRAs and surveys to teachers, school leaders, and caregivers, you will likely need to add additional days to your training. See the Implementation ANNEX for additional information on how to prepare for an assessor training, including materials needed and training agenda.

Conducting an Operational Data Collection

After receiving training, the assessment team will conduct the operational data collection. An operational data collection, in contrast with a pre-test or a pilot data collection, is a full-scale data collection. During the operational data collection, the assessment team will administer the EGRA and any other survey tools based on your pre-determined sample and data collection plan. Often, you will split the entire assessment team into smaller data collection teams. It is important to arrange for a supervisor to accompany each team of assessors. Supervisors provide important oversight for assessors and the collection process. Supervisors are also able to manage relationships with the school staff; accompany learners to and from the testing location; replenish assessors' supplies; communicate with the support team; and fill in as an assessor if needed.

The logistics for an EGRA operational data collection are similar to other data collections. You should have a daily schedule and recommended daily protocols for each data collection team, and you should prepare each team with the necessary materials and information to complete their task effectively and efficiently.

This section details some of the most important logistical details that you will need to consider as you plan and enact your EGRA operational data collection. These considerations should be discussed with your assessment team during the training to ensure that they are well-equipped to conduct the operational data collection successfully

— Planning Your Data Collection Schedule

It is critical to have a clear plan for how your data collection will be completed within the expected timeframe. You should have a daily schedule per data collection team, so that they know what sites they will be visiting each day. You should also include contact information

for a person at the data collection site, like a school principal, so that the team can confirm the data collection visit and ask any questions.

How you plan your overall data collection schedule, and the schedule for each team, is linked to your sampling. See Design Annex for more information on sampling for EGRA.

When designing your data collection schedule, you should keep in mind the following details:

- When can you be at the data collection site: If you're collecting data at a school, you'll want to keep in mind the school day, any school shifts (i.e., morning and afternoon), and the school timetable (i.e., recess, lunch, etc.). You should also determine if data collection can take place outside of school hours, such as after school, on a weekend, or on holidays.
- How many assessments and tools can be completed per day: Experience to date has shown that it takes about 15-20 minutes per EGRA per learner. This means that a team of three assessors can complete about nine or ten instruments per hour, or about 30 children in three uninterrupted hours. If your assessment includes an EGRA in multiple languages, or if you have additional tools, such as teacher surveys, classroom observations, or school inventories, you should include these in your calculations. As much as possible, you should complete all assessments at a school or site within a school day.¹¹
- How long it takes to travel between data collection sites: The data collection team will need to travel between data collection sites and lodging (if applicable). Depending on your sample and location, this may include long distances. You should coordinate with individuals who know the data collection sites well and can provide any details on how long it takes to get there, keeping in mind that some sites may be difficult to reach.
- Are there any safety or security issues to consider: In some places, it is not advisable to travel after sundown, and there may be other considerations that can affect travel and your data collection schedule, such as strikes or protests.

There are likely other factors to consider when developing your data collection schedule, so you should work closely with individuals who have knowledge of field logistics and the context to make sure that your data collection schedule is realistic and achievable.

— Materials Needed for Data Collection

Before your assessment team begins data collection, you should ensure that all team supervisors have all the materials their team will need for their job. Consider providing bags, tablet cases, folders, and all printed materials to your supervisors. Depending on where the data collection will take place, you may also want to provide power banks. All assessors should have phones, both for communication within the team, and also, based on connectivity in the field, to use as a hotspot to upload daily EGRA data. Recommended materials include:

- Laptop or briefcase bags to organize and protect materials; one per assessor
- Tablets with cases that are fully charged and loaded with the data collection tools; one per assessor
- Spare tablets in case of technical malfunction; 1-2 per data collection team
- Learner stimuli booklets; one per assessor¹²
- Pencils or pens; several per assessor
- Clipboards; one per assessor
- Pencils or other small school materials to give to learners in appreciation for their participation
- Power Banks and extension cables

— Daily Tasks at the Data Collection Site

CONDUCTING INTER-RATER RELIABILITY CHECKS DURING DATA COLLECTION

During the first assessment each day, the supervisor arranges for assessors to work in pairs to simultaneously administer the EGRA to the first learner selected, with one actively administering and the other silently observing and marking. This dual assessment—which helps assure the quality of the data across assessors—is called inter-rater reliability.

Each data collection team should have a plan for how to arrive at the data collection site. Sometimes teams have a vehicle that transports them each day; sometimes, team members are responsible for securing their own transport to the site. Regardless, it is the

responsibility of the supervisor and data

Conducting Inter-rater Reliability Checks During Data Collection

During the first assessment each day, the supervisor arranges for assessors to work in pairs to simultaneously administer the EGRA to the first learner selected, with one actively administering and the other silently observing and marking. This dual assessment—which helps assure the quality of the data across assessors—is called inter-rater reliability.

Conducting Inter-rater Reliability Checks During Data Collection

During the first assessment each day, the supervisor arranges for assessors to work in pairs to simultaneously administer the EGRA to the first learner selected, with one actively administering and the other silently observing and marking. This dual assessment—which helps assure the quality of the data across assessors—is called inter-rater reliability.

collection team members to make sure they know what site they must visit each day, and to have a plan for how to arrive promptly at the agreed time.

When arriving at the data collection site, which is often, but not always, a school, the supervisor should visit the school principal or site representative to introduce the data collection team. In many contexts, the team will provide a signed letter from the government that permits them to conduct the data collection. The supervisor should explain the purpose and objectives of the data collection. The supervisor must emphasize that the purpose of the data collection is not to evaluate the school, the principal, or the teachers, and that all information will remain anonymous.

Supervisors and assessors should work with the school principal or site representative to identify an appropriate space for each assessor, such as an unused classroom, teacher room, or library. After identifying assessment spaces, each assessor will proceed to their assigned space and set up their assessment space. There are different possible seating configurations, but often the learner and assessor will sit on chairs facing each other, with a desk or table in the middle.

Data collection teams may need to sample classrooms and learners, depending on your sampling protocols. In many cases, the supervisor is responsible for randomly selecting classroom and learners. The data collection team will also need a standard protocol for where sampled learners wait before being assessed. To reduce test contamination, sampled learners will often wait in an empty classroom or space, so that learners returning to class after the assessment do not share information with them. This, however, has tradeoffs,

since learners may miss classroom instruction. Ultimately, you will need to determine a standard protocol that data collection teams follow. See the Implementation Annex for more details on classroom and learner sampling protocols.

The goal at each data collection site is to reach the daily quota of assessments. If a data collection team has not completed all expected assessments, the supervisor should determine if a.) the team should remain at the school to complete assessments after the school day; b.) the team should return to conduct the remaining assessments on the following day; c.) the team can move forward to the next data collection site without reaching the quota. Supervisors should determine the option that best fits the circumstances and communicate this recommendation to the EGRA Manager.

At the end of the day, the data collection team should make sure to leave the school or data collection site as they found it. This means removing any trash, rearranging desks and chairs, and ensuring that rooms are secured. Refer to the Implementation Annex for more information on characteristics of an assessment space.

— Engaging Learners throughout the Assessment

Ensuring that learners feel comfortable - both physically and emotionally- is critically important to making sure they are engaged and attentive throughout the assessment. Assessors should think about the following:

- Making the physical assessment space as calm and comfortable as possible. This includes things like making sure to have comfortable chairs or child-appropriate chairs, mitigating high temperatures by setting up an assessment space out of the direct sun, and making sure the assessment space is relatively free of distractions.
- Making the learner feel safe. The assessor should be mindful of culturally appropriate communication and maintaining appropriate physical interactions. The assessor should make sure the learner knows why they are being assessed, that their performance will be kept private and will not be used to reward or punish the learner, and that the results will be used to improve the learning experience in the classroom.
- Knowing how to refocus the learner on the assessment. If the learner seems anxious, the assessor could suggest that the learner take a few deep breaths or take a short break between subtasks. The assessor should monitor if or when learners seem distracted and gently guide them back to the assessment.

- Providing appropriate encouragement during the assessment. The assessor should be neutral during the assessment, especially during subtasks, but can provide gentle encouragement. The assessor can stress throughout that it is okay if the learner gets an answer wrong or does not know an answer. Generally, this information is part of the instructions that the assessor reads to the learner at the start of the assessment and at the start of each subtask.
- Providing the learner time to ask questions at the start of each subtask. In the instructions that the assessor reads at the start of the assessment, you should make sure there is a prompt to ask the learner if they have questions.

ENGAGING LEARNERS AFFECTED BY TRAUMA AND/OR IN CONFLICT AND CRISIS SETTINGS

Appropriately supporting and engaging learners affected by trauma and/or learners in conflict and crisis settings requires specific skills and potentially specific accommodations. You should include a specific session during your assessor team training on safeguarding considerations, recognizing signs of trauma, and how to assess learners who may be affected by trauma.

If you are conducting an EGRA within a conflict and crisis setting, you should work in collaboration with child psychologists and trauma experts to design specific guidance and protocols for assessors. The guidance should focus on:

Recognizing signs that a learner is affected by trauma;

- Supporting learners during introductory discussions and rapport-building;
- Knowing how to choose appropriate accommodations, either by shortening the session or offering more time to start and complete the assessment;
- Documenting any accommodations; and
- Accounting for the various accommodations in analysis and reporting.

During an assessment, recognizing signs that a learner is affected by trauma can also create an opportunity to refer learners for support services. This is a sensitive process and requires proper training. Your guidance and protocols should include any steps an assessor should take to refer a learner for support services; this likely will include the assessor engaging directly with a safeguarding/trauma expert or child psychologist and ensuring the confidentiality of the learner and their assessment data.

In some cases, a learner will ask to stop the assessment. Because of assent protocols, a learner can always decide to end the assessment early, and their request must be respected. After the assessor gently encourages the learner to continue, if the learner still insists, it is the assessor's responsibility to end the assessment and document why the assessment was terminated early.

Refer to the Implementation Annex for guidance on how to engage learners throughout the assessment.

EGRA FOR SELECTED INTEGRATED READING ACTIVITY IN MALI

The Selected Integrated Reading Activity (SIRA) was a five-year project (2016 to 2021), funded by USAID/Mali, to improve early grade reading instruction, improve service delivery systems, and increase parent, community, and private support for early literacy. The activity took place in the regions of Koulikoro, Sikasso, and Segou, and the District of Bamako.

The Mitchell Group, along with EdIntersect and Malian partner CEPROCIDE, conducted a mid-term performance evaluation of the activity between January and March 2020. The evaluation assessed 1,071 learners in 110 schools. The team contended with several contextual challenges: ongoing teacher strikes, insecurity, and the onset of COVID-19 (The Mitchell Group, 2020). For example, due to violence and threat of insurrection by local armed groups, 47 percent of schools originally identified were inaccessible, and the team had to adjust the sample (Giuliano Sarr and McMahon, 2020).

The evaluation team employed the following strategies to conduct a safe and successful evaluation.

- Emphasis on partnerships with local actors and systems, including CEPROCIDE and the SIRA project team. For example, during the assessor training, local education officials helped the team identify a school for piloting the assessment when the team encountered school closures due to teacher strikes. In addition, the evaluation team gathered and reviewed security information from a number of sources and relied on in-country team members' judgments to pivot the data collection plan when necessary. Key sources of information included USAID/Mali, the SIRA implementing team, CEPROCIDE, and in-country evaluation team members (Giuliano Sarr and McMahon, 2020). The team also relied on an evaluation manager who was a host-country national employed by the Mitchell Group, and who had a long history of playing logistical and operational roles. He applied his insights to assist with coordination and decision-making (K. Giuliano Sarr, personal communication, September 24, 2024).
- A flexible and responsive design that emphasized creative use of technology, including the creation of WhatsApp groups and regular telephone check-ins with assessors; initiatives to maintain flexibility, such as the development of a comprehensive back-up school list and daily field updates; and participatory relationships with stakeholders (Giuliano Sarr and McMahon, 2020).

The evaluation recommended that similar contexts consider the following (Giuliano Sarr and McMahon, 2020):

- Attention to the context and the potential need for conflict-sensitive mechanisms such as those discussed above,
- Learning and capacity-building at all levels, including with international team members, local team members, data collection teams, project implementing partners, and government stakeholders, and

Use of technology by international team members to provide remote support to in-country colleagues, and similarly between colleagues in Bamako and those conducting fieldwork in the regions.

— Uploading Tablet-Based Data

BENEFITS OF REGULARLY UPLOADING AND REVIEWING DATA

During data collection, regular data uploading and review can help catch any errors before the end of data collection, saving projects from sending data collectors back into the field after weeks of data collection. Additionally, daily uploads can help prevent loss of large amounts of data if a tablet is lost, is stolen, or breaks. Data can be checked to ensure that the correct grade is being evaluated, that assessors are going to the sampled schools, and that the correct numbers of learners are being assessed, as well as to verify any other inconsistencies. Constant communication and updates to let the project team know when data collection is proceeding, when the data analysts see uploaded data, and whether there are any delays or reasons that would prevent the uploading of data on a daily basis can help in reviewing the data as well as knowing what results to expect and when.

One of the most important tasks on each day of data collection, when internet connection is available, is to upload tablet-based data to a central server. Uploading data on a daily basis avoids potential data loss in the case of software or hardware malfunction, and it also allows for ongoing monitoring of data collection progress and data quality assurance. You should ensure that supervisors and assessors receive adequate training on how to upload their data, and what to do in case of any troubleshooting, during the assessor team training.

Assessors are responsible for uploading their data on a daily basis, and supervisors should ensure that this task is complete. Depending on location and connectivity, assessors can use wireless internet or mobile data - either directly through the tablet or by using a mobile phone hot-spot. You should work closely with individuals who have knowledge of field logistics to create the protocols for daily data upload that fit the context.

Depending on your specific quality assurance and quality control protocols, a member of your EGRA team, such as a QCO, will monitor daily upload and cross check that the reported data collected through daily supervisor reports are appearing in the server. See the Implementation Annex for more on quality assurance and quality control protocols.

Developing A Quality Assurance and Quality Control Plan

Quality assurance (QA) and quality control (QC) throughout the EGRA process and during data collection are necessary to ensure that your data are complete, precise, and accurate.

Quality assurance (QA) is the broader set of strategies and protocols that ensure data integrity, quality, and reliability throughout the EGRA process, including during EGRA tool design, pre-testing and piloting, assessor team training, data collection, and analysis.

Quality control (QC) refers to the specific set of protocols that guide the quality of your data. You should work closely with your entire EGRA team to design your Quality Assurance and Quality Control (QAQC) Plan, as everyone is responsible for the successful and quality implementation of the EGRA process. See the Implementation Annex for details on developing a QAQC plan.

04 Analyze

The fourth phase of EGRA is analysis. It is the phase when the raw EGRA data are processed, cleaned, and analyzed leading to initial findings about learners' performance in the reading skills measured by the EGRA.

The team members that will need to be involved in this phase include

- EGRA Manager
- Psychometrician or statistician
- Other team members to participate in an emerging results discussion

The Analyze phase of EGRA is guided by a three-level process, which increases in complexity and difficulty. These steps are generally conducted by a psychometrician or statistician/analyst, and completing these steps with quality requires a substantial level of technical expertise. This section is intentionally brief because it relies on a trained expert to complete the tasks. The trained expert will find the resources, examples and templates in the Analyze Annex. The EGRA Manager will also find guiding questions to help them liaise with the experts to successfully complete analysis to meet the stated purpose of the assessment.

To complete the analysis successfully, psychometricians or statisticians/analysts should be familiar with the following technical concepts:

- CTT vs. IRT and Rasch
- Deciding when you need to equate forms
- Deciding when you need to apply sampling weights
- Deciding when you need to drop cases or items

Using these concepts, the experts will work through a three-level analysis process. The three levels are briefly described below.

Level 1 contains steps to set up your dataset(s) and clean them. Level 2 contains steps to score your data and generate analysis variables, weights and prepare a final analytical file. Level 3 contains steps to conduct descriptive and statistical analyses and produce the output tables and results.

For each level, the following will be provided in a forthcoming release of this document:

- What does it mean?
- Which data sources are included in this step?
- What is produced at the end of this level?
- How much of this step can be automated?
- How do the outputs from this step get used?

See the forthcoming Analyze Annex for templates and worked examples of the three-level analysis process.

05 Report

The fifth phase of EGRA is reporting. It is the phase when the plans and data from previous phases are summarized and packaged for the intended audience. This phase draws directly from the Design phase, where decisions were made based on what will be reported, to whom, and for what purpose.

The team members that will need to be involved in this phase include:

- EGRA Manager
- Psychometrician or statistician
- Other team members to participate in an emerging results discussion

This chapter is forthcoming. It will include templates and worked examples in addition to narrative instructions. It will link to the "Presentation of findings" and "Data presentation and application team" sections at the end of the Design phase.

FOOTNOTES

- 1 See UIS criteria here: <https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2024/02/GAML-Criteria-for-reporting.pdf> ↩
- 2 A census is when all members of the population are included in a study, i.e., no sampling is conducted. ↩
- 3 Figure taken from page 9 of this Landscape Report that summarizes the skills assessed in EGRA through 2016 and the underpinnings of the assessment. The diagram of the Six Figures of Literacy is taken from here. ↩
- 4 The RELM is grounded on the Simple View of Reading (SVR) Framework. This model postulates that reading is supported by two pillars: language comprehension, or “the ability to construct meaning from spoken representations of language” and decoding, or “the ability to recognize written representations of words” - and their sub-skills (Hoover and Gough, 1990). While most existing Early Grade Reading Assessment (EGRA) sub-tasks focus on the decoding pillar of the SVR, the RELM focuses primarily on the language comprehension pillar. The RELM operationalizes linguistic comprehension by measuring listening comprehension and vocabulary, in its receptive and expressive form. The intent is that the RELM will help users to disentangle floor effects (or zero scores) for EGRA and, thus, drill down on the key components affecting reading comprehension. ↩
- 5 An example is obtaining ethics approval from an IRB. If IRB approval is delayed, operational data collection cannot begin. In this case, it is wise to ensure there is at least 1-2 weeks buffer before data collection is planned to begin in the event there is a delay in obtaining the IRB approvals. ↩
- 6 Workshop organizers should ask all participants if they need accommodations to participate, such as sign language interpretation, soft copy and screen readable versions of presentations, and braille print copies of any handouts. ↩
- 7 This includes ensuring that all items are accessible and non-biased for the assessment population. Accessible, non-biased items are free of content that may be biased against certain groups and give all learners in the target population equal probability of answering correctly. For example, if an EGRA is being conducted in a refugee camp context, special attention should be paid to ensure that an oral reading fluency passage includes cultural and contextual references that are familiar to all the different refugee groups in the camp. ↩
- 8 The degree to which the items on the EGRA test are representative of the construct being measured is known as test-content-related evidence (i.e., early reading skills in a particular country). ↩
- 9 Changes to administration protocols could be accommodations or modifications, depending on the changes. ↩
- 10 If multiple versions of an instrument will be needed for baseline/endline studies, preparing and piloting parallel forms at this stage helps determine, and has the potential to lessen, the need for equating the data after operational data collection. ↩
- 11 This is to reduce the chance that learners will share things about the assessment with their peers - also known as test contamination. ↩
- 12 Because learner stimulus sheets will be used with multiple learners, it is recommended that stimuli are laminated or put in plastic page protectors. ↩