



USAID
FROM THE AMERICAN PEOPLE



Using EGRA to Report on SDG 4.1.1a

10th Meeting of GAML
2023 - Paris, France
Melissa Chiappetta
USAID

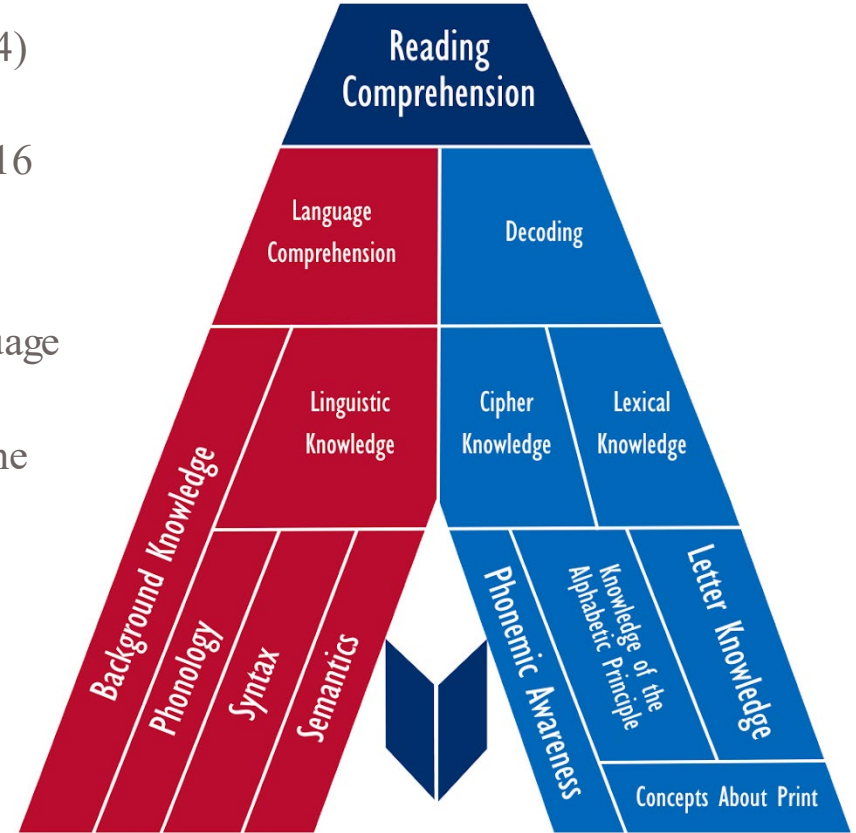
AGENDA

- Overview of Early Grade Reading and Math Assessments
- EGRA Coverage
- Updates to the EGRA Toolkit- Coming in February 2024
- EGRA Adherence to Eligibility Criteria for Reporting on SDG 4.1.1
- EGRA's Potential to Expand SDG 4.1.1a Coverage
- Other USAID Assessment and Benchmarking Guidance and Work
- Recommendations and Questions



EARLY GRADE READING ASSESSMENT (EGRA) OVERVIEW

- Designed in 2008 for young learners (grades 1-4) based on the Simple View of Reading
- Refined through piloting, Toolkit updated in 2016
- Administered orally and 1:1 between a trained assessor and a child
- Adapted to and typically administered in a language that the child can use and understand
- Statistically linked to allow comparison over time
- Subtasks measure:
 - Listening comprehension and vocabulary
 - Letter naming, letter sounds,
 - Familiar words and non-words
 - Decoding
 - Decoding accuracy and fluency
 - Reading comprehension



EARLY GRADE MATHEMATICS ASSESSMENT (EGMA) OVERVIEW



- Designed for grades 1-4 in 2008
- Piloted and refined over time-Toolkit published in 2014
- Criteria for inclusion of subtasks:
 - Represent a progression of foundational skills that support proficiency
 - Research indicating predictive power—i.e., test skills related to students' future performance
 - Common in many early grade curricula
- Administered orally and 1:1
- Adapted to local contexts and languages
- Subtasks measure:
 - Number identification
 - Number discrimination
 - Missing Number
 - Addition and Subtraction
 - Word Problems

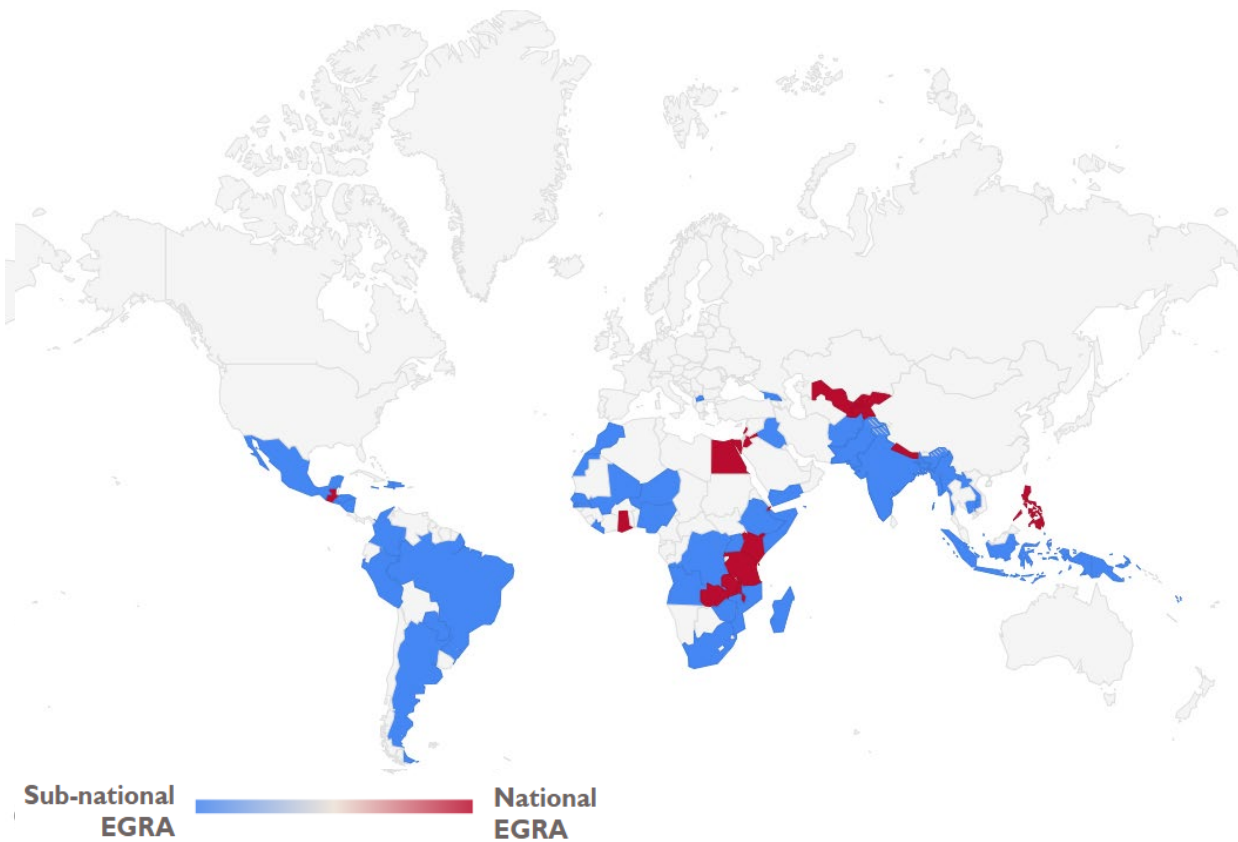
EGRA AND EGMA ADDITIONAL QUESTIONNAIRES

- Include learner questionnaires to identify key demographic information
- Optional but common:
 - Teacher survey
 - Head teacher survey
 - School environment protocol
 - Classroom observations
- Occasionally include household surveys
- Allow countries to identify which skills learners lack, which groups are most effected, and why



EGRA DATA CAN SIGNIFICANTLY INCREASE SDG 4.1.1A COVERAGE

- 75+ countries have conducted EGRAs using more than 120 languages
- **At least 23 countries have conducted nationally - representative EGRAs with Grade 2 or 3 data since 2014**
- At least 3 more nationally-representative EGRAs are expected in 2024-25
- Also have 3 nationally representative EGMAs



2023/2024 EGRA CONTENT UPDATES

- New Receptive and Expressive Language Module (up to 45 items)
- Move to Decoding Accuracy at Grade 2 (about 50 words)
- Decoupling Decoding Fluency and Comprehension at Grade 3
- Allowing learners to use lookbacks for the reading comprehension questions
- Improving and Expanding Reading Comprehension (15 items)
- Ensuring EGRA Adheres to Universal Design for Assessment Practices





2023/2024 EGRA TECHNICAL UPDATES

- Clarifying the need for nationally representative data for countries that wish to report to SDG 4.1.1a
- Ensuring reporting of reliability and validity measures by domain and overall assessment
- Updating assessment confidentiality requirements
- Clarifying guidance on use of conjunctive or compensatory scoring model
- Updating benchmark setting guidance
- Updating documentation and reporting requirements to include details on the adaptation workshop and item development/appropriateness, tool piloting, assessment psychometric properties, administration incidents, etc.

EGRA ADHERENCE TO THE
ELIGIBILITY CRITERIA FOR
SDG 4.1.1A REPORTING



CRITERION 1: IS THE ASSESSMENT SUFFICIENTLY ALIGNED WITH THE MPL/GPF?

Level of Alignment	Category	SDG 4.1.1(a) GPF Grade 2	EGRA 2016-2023	EGRA 3.0 (2024)
Minimally Aligned	Test Length	Min. score of 20 if setting "meets" level Min. score of 45 if setting three levels		
	Domains (depth)	Decoding (min. 10 points) Listening Comprehension (min. 5 points)		
	Subconstructs (breadth)	4 of the 7 decoding and listening comprehension subconstructs		
Additionally Aligned	Test Length	Min. score of 20 if setting "meets" level Min. score of 45 if setting three levels	Min. 60 items across listening comprehension, decoding, and reading comprehension	
	Domains (depth)	Decoding (min. 10 points) Reading Comprehension (min. 5 points)	Decoding - 50 items Reading Comprehension - 5 items	
	Subconstructs (breadth)	3 of the 5 decoding and listening comprehension subconstructs	4 of the 5 subconstructs covered	
Strongly Aligned	Test Length	Min. score of 20 if setting "meets" level Min. score of 45 if setting three levels		~105 items across listening comprehension, decoding, and reading comprehension
	Domains (depth)	Reading Comprehension (min. 10 points)		Reading Comprehension - 10-15 items
	Subconstructs (breadth)	1 of the 2 reading comprehension subconstructs		At least 1 of the 2 subconstructs covered

CRITERION 1: IS THE ASSESSMENT SUFFICIENTLY ALIGNED WITH THE MPL/GPF?

MPL requirements

EGRA Toolkit Guidance

01

MPL: When listening to slightly longer texts, they make simple inferences

- **Task:** Listening comprehension subtask/Receptive and expressive language module
- **Guidance:** 2024 revision to include between 12-45 questions (5 items in the 2016 guidance)

02

MPL: Accurately read aloud

- **Task:** Accuracy subtask
- **Guidance:** 2024 revision to focus on accuracy with ~50 words (fluency was linked to comprehension in 2016 guidance)

03

MPL: Retrieve explicit information

- **Task:** Reading comprehension subtask
- **Guidance:** 2024 revision to require 10-15 questions (5 items in 2016 EGRA guidance)

CRITERION 2: HAVE THE ITEMS BEEN REVIEWED QUALITATIVELY AND QUANTITATIVELY?

	Criteria for Reporting on SDG 4.1.1	EGRA Toolkit Guidance
01	Is there evidence that the items in the assessment have been reviewed qualitatively?	<ul style="list-style-type: none">Teachers and curriculum, language, and assessment experts are engaged in the tool adaptation processPilots conducted by teachers & trained enumerators
02	Is there evidence that the items in the assessment have been reviewed quantitatively?	<ul style="list-style-type: none">Items difficulty/facility and item discrimination will be required to be reported in EGRA 3.0 ToolkitItem discrimination must be greater than 0.2
03	Have items been reviewed to ensure appropriateness for relevant population subgroups?	<ul style="list-style-type: none">Assessment follows UDA principles and can be adapted for learners with disabilitiesAdaptations consider contextual realitiesPilots are conducted with relevant subgroups
04	Test development process is documented in detail	<ul style="list-style-type: none">EGRA guidance will provide a job aide aligned with Standards 4.7 - 4.10 to meet this criterion

CRITERION 3: IS THE SAMPLE OF LEARNERS THAT TOOK THE ASSESSMENT REPRESENTATIVE?

	Criteria for Reporting on SDG 4.1.1	EGRA Toolkit Guidance
01	Was the assessment administered to the whole cohort or a sample?	<ul style="list-style-type: none">• Nationally-representative sampling guidance will be provided in the job aide in the EGRA toolkit• Will include sampling when multiple assessments are used to cover different areas/languages
02	Were any subgroups of the population excluded?	<ul style="list-style-type: none">• EGRA generally does not include OOS children/youth• Goal: to ensure inclusion of learners with disabilities• Guidelines around age of students?
03	Is the margin of error 5% or less - 95% confidence level?	<ul style="list-style-type: none">• Sampling requirements do include this guidance in the EGRA Toolkit
04	Was the MDES calculated to ensure comparison over time	<ul style="list-style-type: none">• EGRA Toolkit guidance has always required power analysis be conducted and MDES be appropriate to allow comparisons over time

CRITERION 4: IS THERE EVIDENCE THE ASSESSMENT WAS ADMINISTERED IN A STANDARDIZED WAY?

	Criteria for Reporting on SDG 4.1.1	EGRA Toolkit Guidance
01	Were administration guides clear on the administration process?	<ul style="list-style-type: none">• Enumerator & supervisor training and reporting requirements are standardized - EGRA Toolkit and Guidance Note for Planning/ Implementing EGRA• IRR of .80 required; supervisor checks during admin.
02	Was the assessment administered in appropriate and standardized way?	<ul style="list-style-type: none">• Each supervisor and assessor team conducts daily dual-assessor checks (i.e., score the same child) to monitor scoring discrepancies• Supervisors have QA checklists
03	Were significant incidents recorded & relevant results excluded from the outcomes?	<ul style="list-style-type: none">• EGRA reporting guidance requires standard sections in the report methodology• Updates will provide specific documentation guidance around administration and incidents

CRITERION 5: ARE THE OUTCOMES OF THE ASSESSMENT SUFFICIENTLY RELIABLE?

Criteria for Reporting on SDG 4.1.1

EGRA Toolkit Guidance

01

Is the value of coefficient alpha for the assessment greater than or equal to 0.7?

- Cronbach's alpha is required with a value of at least 0.7, though most EGRAs have a value of 0.8
- Past EGRAs have not measured alpha by subtask due some subtasks being timed; updates will require this

CRITERION 5: ARE THE OUTCOMES OF THE ASSESSMENT SUFFICIENTLY RELIABLE?

Country	Grade	Year	Language	Sample	Population	Alpha
Asia 1	2	2019	Local	2114	168201	0.94
MENA 1	2	2015	Arabic	1760	1738785	0.91
MENA 1	3	2015	Arabic	1760	1682710	0.91
LAC 1	2	2018	Spanish	949	92217	0.89
LAC 1	3	2018	Spanish	956	91860	0.86
Africa 1	2	2015	English	7311	436102	0.86
Africa 1	2	2015	Local	7311	438594	0.89
MENA 2	2	2019	Arabic	2967	129878	0.90
MENA 2	2	2021	Arabic	1039	49757	0.91
MENA 2	3	2019	Arabic	3011	137267	0.89
MENA 2	3	2021	Arabic	1058	53243	0.90
Africa 2	2	2012	English	4262	18089	0.84
Africa 2	2	2013	English	2112	17807	0.82
Africa 2	2	2013	Kiswahili	2112	17521	0.90
Asia 2	2	2021	Local	1335	13135	0.90
Asia 2	2	2021	Russian	857	175780	0.89
Asia 2	4	2021	Local	1361	11255	0.82
Asia 2	4	2021	Russian	945	202939	0.85

Country	Grade	Year	Language	Sample	Population	Alpha
Africa 3	2	2015	English	1258	676167	0.82
Africa 3	3	2015	English	1224	627224	0.84
Asia 3	2	2020	Local	2692	177000	0.95
Asia 3	3	2020	Local	2718	166358	0.95
Asia 4	3	2013	English	2463	2300615	0.90
Asia 4	3	2013	Local	2463	2296396	0.91
Asia 4	3	2019	English	2385	2022639	0.86
Asia 4	3	2019	Local	2385	2022639	0.92
Africa 4	2	2016	Kiswahili	2647	246376	0.92
Africa 4	3	2016	Kiswahili	670	186958	0.92
Africa 5	2	2013-2018	Local	1701	25812	0.78
Africa 5	3	2013-2018	Local	960	24375	0.79
MENA 3	2	2018	Arabic	1465	48509	0.87

*RTI International, 2023

CRITERION 5: ARE THE OUTCOMES OF THE ASSESSMENT SUFFICIENTLY RELIABLE?

Criteria for Reporting on SDG 4.1.1

EGRA Toolkit Guidance

01

Is the value of coefficient alpha for the assessment greater than or equal to 0.7?

- Cronbach's alpha is required with a value of at least 0.7, though most EGRAs have a value of 0.8
- Past EGRAs have not measured alpha by subtask due to some subtasks being timed; required in update

02

Are QA arrangements appropriate for human-scored items with agreement >80%

- QA described on the last slide
- IRR of at least 80% required in piloting

**EGRA DATA CAN
SIGNIFICANTLY
INCREASE SDG 4.1.1A
COVERAGE BY
ADDRESSING
REPORTING CRITERIA
FOR EXISTING DATA
AND CODIFYING
CRITERIA IN UPDATED
TOOLKIT**



SOME DATA MAY ALREADY MEET SDG 4.1.1A CRITERIA

Criterion	Evidence required	Select country 1 data (2020)	Select country 2 data (2021)
1. Is the assessment sufficiently aligned with MPL/GPF?	MPL/GPF alignment and # items	Decoding (57 words), reading comprehension (5 items), and listening comprehension (5 items)	Decoding (60 & 70 word passages), reading comprehension (5 items), and listening comprehension (5 items) in local language & language of instruction (also EGMA)
2. Have the items been reviewed to determine suitability for inclusion?	Adaptation process and participants	Curriculum, language experts, statisticians, teachers, EGRA trainers conducted adaptation workshop; pilot examined item difficulty, item discrimination, and bias by language, SES and disability (item & subtask)	Use of past adaptation, same pilot statistics conducted, and pilot and assessment examined correlations across local language & language of instruction, all subtasks equated & comparable to past years
3. Is the sample representative?	Sampling frame (2 stage)	All government schools in country with Grade 2 stratified by region, language; students sampled by section, classroom, and sex	Sampled from all public and non-state schools listed in EMIS stratified by county; students sampled by grade, section, class, sex
4. Was assessment administered in a standardized way?	Kappa coefficient for IRR	Listening: 0.83; reading: 0.80; reading comprehension: 0.83	Listening: 0.97; reading: 0.95; reading comprehension: 0.86 (comparable for 2nd language)
5. Are the outcomes sufficiently reliable?	Cronbach's alpha	Listening: 0.85, reading comprehension: 0.81	Listening: 0.87, decoding passage: 0.9, reading comprehension: 0.91

USAID IS MAKING OTHER
RELEVANT UPDATES TO
ITS INDICATOR
REPORTING GUIDANCE



USAID IS ENHANCING ACCESS TO AND USE OF ASSESSMENTS AND BENCHMARKING METHODS

- **Assessment decision tree** and **job aides** to support decision-makers to identify the right assessment for each country's purpose - **expected February 2024**
- **US Foreign Assistance “F” Indicators allow use of ANY assessments and all standardized benchmarking methods**
- **EGRA specific:**
 - Expanding data summaries on the [EarlyGradeReadingBarometer.org](https://www.earlygradebarometer.org)
 - Permanent repository of USAID data with a growing catalogue of EGR data sets
 - data.usaid.gov → **Basic Education**
 - edu-links.org/resources/find-and-access-usaid-education-data
 - Templates to standardize EGRA reporting with STATA/R code and dataset templates - **expected February 2024**

RECOMMENDATIONS AND OUTSTANDING QUESTIONS

Recommendations

- **Stop Rules:** Accept, at a minimum, past data (but preferably future data as well) where stop rules have been used if administering agencies are able to demonstrate that the stop rules do not significantly affect results
- **Nationally -Representative Data:** Accept data that are nationally representative as a whole, even if that means different assessments are used for different populations (e.g., stratified by language) and linked through policy linking or another method
- **Age of Students :** Students enrolled in the grade of instruction are included in results as long as we demonstrate their inclusion does significantly affect results.
- **Acceptance of Assessment that use Classical Test Theory:** Accept use of assessments that use CTT or IRT (reporting guidance says only IRT allowed).
- **Alignment for Individually Administered Assessments:** Criteria for alignment should be the same for individually administered assessments as what is already being accepted for group-administered assessments.

RECOMMENDATIONS AND OUTSTANDING QUESTIONS

Recommendations Continued

- PAL, USAID, and UNICEF collectively recommend constituting a Technical Advisory Group (TAG) to support EGRA, MICS-FLM, and PAL-N assessment design decisions for reporting on SDG 4.1.1a (could be a reinstatement of the 4.1.1 Review Panel)
- Agree to finalizing outstanding questions by end of January 2024 for inclusion of any required changes in forthcoming EGRA Toolkit revisions

Discussion Questions

- **Number of Additional Countries Needed:** In 2024, how many more countries need to report on SDG 4.1.1a from which regions to ensure adequate coverage to upgrade SDG 4.1.1a back to Tier 1?
- **Household versus Classroom -Based Assessments:** Will UIS accept data that are nationally representative of the population of students who are enrolled in school rather than for all children nationwide?
- **Test Confidentiality:** Will UIS accept the results from tests when used versions are released publicly and not reused to report results?

NEXT STEPS

- Continue examining impact of the use of stop rules
- Check domain-level alphas across
- Finalize EGRA Toolkit (with guidance from GAML, TCG, UIS)
- Continue examining past EGRA data to determine which datasets might be sufficient to include, and discuss inclusion with relevant countries
- Support UIS and other stakeholders in establishing a TAG







USAID
FROM THE AMERICAN PEOPLE

(1) Albertka is my name
When I grow up I want
to be a lawyer.
(2) I love to be in school.
(3) I best su

STAY CONNECTED

usaid.gov/education | edu-links.org 

education@usaid.gov | edulinks@usaid.gov 

twitter.com/USAIDEducation | facebook.com/USAIDEducation 