



USAID
FROM THE AMERICAN PEOPLE



SHARE
Supporting Holistic & Actionable
Research in Education

Receptive and Expressive Language Model (RELM)

Development and Pilot Testing Across Four Countries

Fernanda Soares (UND), Caroline Freeman (UND), Cheng Liu (UND), Pooja Nakamura (AIR), Emily Kochetkova (IDP)



UNIVERSITY OF NAIROBI



UNIVERSIDADE
EDUARDO
MONDLANE



Philippine
Institute for
Development
Studies



UNIVERSITY OF
NOTRE DAME

Pulte Institute for Global Development
Global Center for the Development of the Whole Child



AIR
Advancing Evidence.
Improving Lives.



INCLUSIVE
DEVELOPMENT
PARTNERS



Agenda:

- I. RELM Rationale and Overview
- II. Preliminary Pilot Results
- III. Final thoughts



SHARE
Supporting Holistic & Actionable
Research in Education

I. RELM Rationale and Overview

Motivation for Developing RELM



SHARE
Supporting Holistic & Actionable
Research in Education

1. In the last two decades, widespread use of assessments such as EGRA and ASER in LMICs
2. Most literacy assessments assume learners' sufficient exposure to the language of instruction, which may not hold in multilingual contexts.
3. Prevalence of floor effects in reading comprehension tasks and lack of evidence identifying drivers of zero scores - i.e., if students' challenges with reading are related to “decoding” or “language comprehension”.

What are RELM's main objectives?



SHARE
Supporting Holistic & Actionable
Research in Education

- To assess pertinent aspects of learners' **language skills** in pre-primary and primary grades across low- and middle-income countries.
- RELM is proposed as an additional module for The Early Grade Reading Assessment (EGRA) but can also be implemented as a standalone assessment for language skills.
- RELM provides a tool to help policymakers to disentangle reading difficulties



RELM Sub-tasks



SHARE
Supporting Holistic & Actionable
Research in Education

Receptive Vocabulary

- 12 items total

Expressive Vocabulary

- 12 items total

Listening Comprehension

Short story listening comprehension

- 12 items total: two passages, six items each

Short story picture-matching

- 12 items total

Semantic Fluency

- Up to three items, in three different languages, depending how many the student uses

Core Subtasks

Supplemental Subtask

Receptive Vocabulary

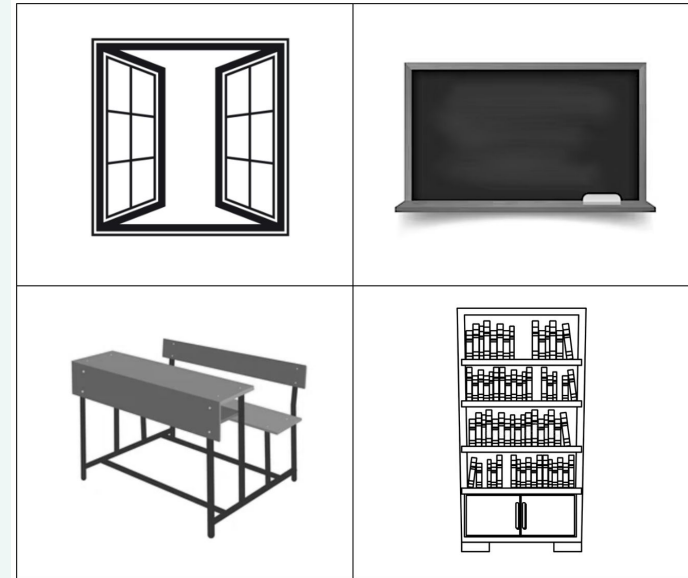


SHARE
Supporting Holistic & Actionable
Research in Education

Each item has a corresponding stimulus with **4 images** and **one target word**

- 1 image is the target (correct) image
- 3 images are the distractor (incorrect) images
- The child is asked to identify the image that matches the target word

Stimulus (visual)



Expressive Vocabulary



SHARE
Supporting Holistic & Actionable
Research in Education

Each item has a corresponding stimulus with **one image** and **one target word**

- The child is asked to say the word that matches the image that the assessor shows

Stimulus (visual)



Listening Story-Picture Matching



SHARE
Supporting Holistic & Actionable
Research in Education

Each item has a corresponding stimulus with **4 images** and **one target sentence or short story** increasing in complexity:

- Four items at the single sentence level.
- Four items at the 2-3 sentence level.
- Four items at the short story level
- 1 image is the target (correct) image
- 3 images are the distractor (incorrect) images
- The child is asked to identify the image that best represents the sentence or short story that is read out loud to them



Listening Comprehension Short Story



SHARE
Supporting Holistic & Actionable
Research in Education

2 passages, each passage has **6 associated multiple choice questions** of increasing complexity (parallel to picture-matching)

- First two questions are directly extractive from the passage (explicit in the sentence)
- 3rd-4th questions are cross-sentence
- 5th-6th questions are more inference making (e.g., How is she feeling?)
- 1 choice is the correct answer
- 3 choices are the distractor
- The child is asked to identify the best answer to the question about the short story

Example story: My name is Senga. I am 9 years old. I live on a farm with my mother, father, and sister Ana. Every year, the land gets very dry before the rains come. We watch the sky and wait. One afternoon as I sat outside, I saw dark clouds. Then something hit my head, lightly at first and then harder. I jumped up and ran towards the house. The rains had come at last.



SHARE
Supporting Holistic & Actionable
Research in Education

II. Preliminary Pilot Results

RELM Piloting Countries

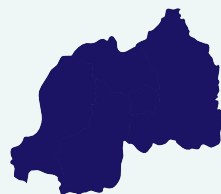


SHARE
Supporting Holistic & Actionable
Research in Education



MOZAMBIQUE

G3: Emakhua
G4: Portuguese



RWANDA

G3: Kinyarwanda
G4: English



KENYA

G3: Dhoulo
G4: English



THE PHILIPPINES

G3: Cebuano
G4: English

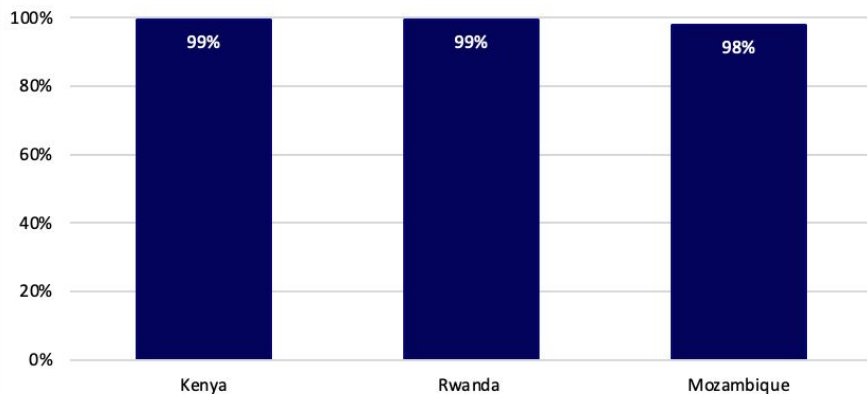
Round 1

Round 2

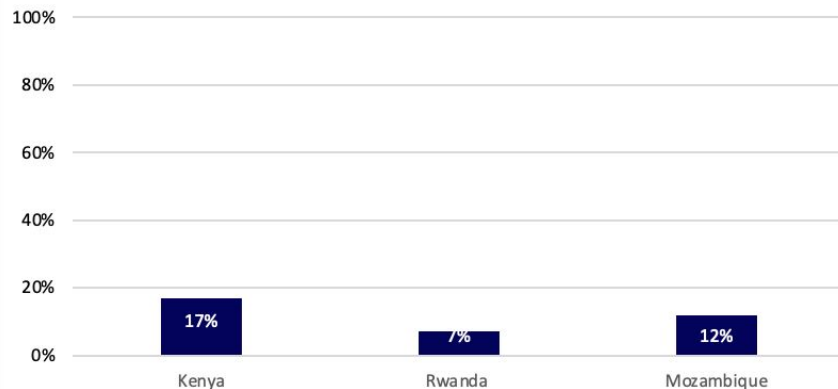
Student Sample: Language profile

- The majority of Grade 3 students speak the official LOI at home (an L1), but only a few Grade 4 students speak the official LOI at home (an L2+).

Percentage of Grade 3 Student who Speak LOI at Home



Percentage of Grade 4 Student who Speak LOI at Home



RELM Subtasks: Score Distributions



SHARE
Supporting Holistic & Actionable
Research in Education

- Grade 3 (L1) subtasks:
 - Ceiling effects, right-skewed distributions and limited discriminatory power
 - List. Comp. Short Story: subtask is able to capture more variability and approximate a normal distribution.
- Grade 4 (L2) subtasks:
 - Subtasks able to capture more variability. We observe mostly left-skewed distributions as well as distributions that approximate a normal distribution (exception of Kenya).

Disaggregated Analysis

- **By language:** Within each country and grade, **learners taught in a language spoken at home score higher on RELM subtasks** than learners who are taught in a language they do not speak at home
- **By SES:** Within each country and grade, **learners from the top wealth quintile score higher** than learners from the bottom wealth quintile
- **By gender:**
 - In Mozambique, **male learners score higher than female learners** in all RELM subtasks
 - No consistent pattern in Kenya and Rwanda

Grade 3 (L1) RELMs Internal Correlations



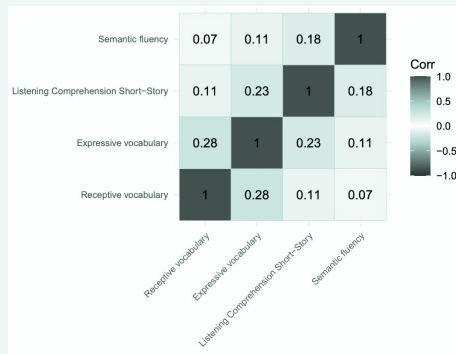
SHARE
Supporting Holistic & Actionable
Research in Education

- Low internal correlations between RELM **Grade 3 (L1)** subtasks in each country (below 0.30), likely driven by insufficient variation in responses (ceiling effects, right-skewed distributions)

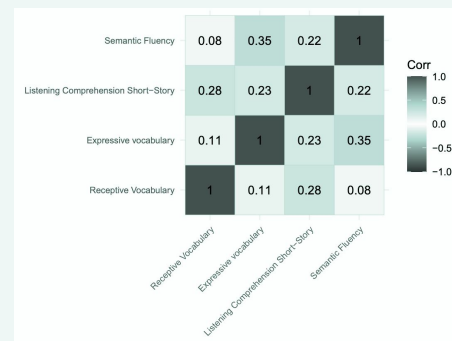
Kenya



Rwanda



Mozambique



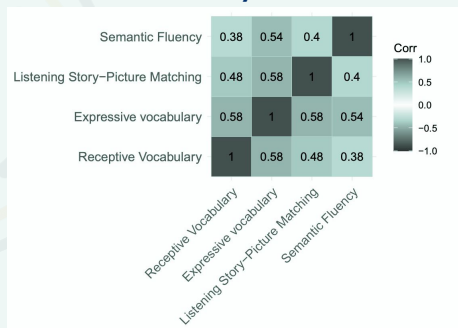
Grade 4 (L2+) RELMs Internal Correlations



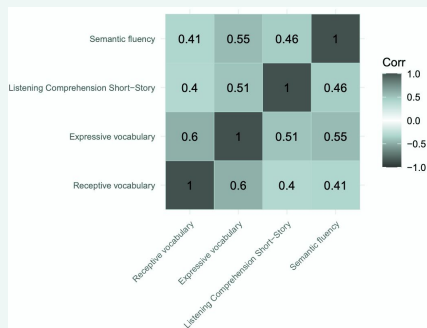
SHARE
Supporting Holistic & Actionable
Research in Education

- Moderate positive correlations between **RELM Grade 4(L2)** subtasks in each country (0.30 to 0.60)

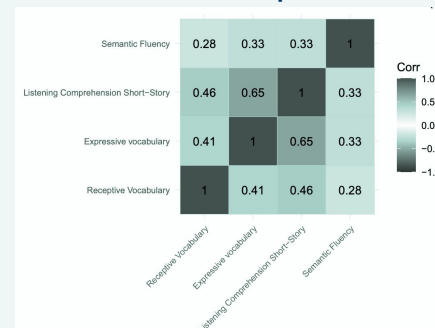
Kenya



Rwanda



Mozambique



Internal Consistency

- Reliable internal consistency observed between RELM **Grade 4 (L2)** subtasks.
- Low internal consistency observed between RELM **Grade 3 (L1)** subtasks.

	Cronbach's Alpha		
	Kenya	Rwanda	Mozambique
RELM Grade 3 (L1)	0.316	0.300	0.504
RELM Grade 4 (L2)	0.695	0.815	0.734

Evidence from Children with Reading Comprehension Difficulties



SHARE
Supporting Holistic & Actionable
Research in Education

- Grade 4 (L2+) weak comprehenders have less developed language skills compared to strong comprehenders, despite having similar decoding skills

	Kenya			Rwanda		
	Low reading comp.	High reading comp.	Diff. (H-L)	Low reading comp.	High reading comp.	Diff. (H-L)
	n = 57	n = 57		n = 63	n = 63	
Age	10.39 (0.94)	10.17 (0.91)	-0.22**	11.71 (1.45)	11.59 (1.39)	-0.13
SES	1.10 (0.39)	1.11 (0.43)	0	1.07 (0.51)	1.10 (0.49)	0.03
Invented Words	9.43 (6.26)	9.57 (6.23)	0.15	15.11 (8.11)	15.27 (8.18)	0.16
Reading Comprehension	5.52 (1.97)	9.43 (0.50)	3.91	2.24 (1.03)	5.68 (0.74)	3.44
Receptive Vocabulary	9.94 (1.52)	10.72 (1.23)	0.78**	6.35 (2.41)	8.03 (1.45)	1.68
Expressive Vocabulary	6.53 (2.50)	9.14 (1.96)	2.36**	1.82 (1.51)	4.14 (2.63)	1.45**
Listening Comprehension	9.50 (1.56)	10.35 (1.34)	0.88	3.26 (1.68)	4.78 (2.07)	1.14*



SHARE
Supporting Holistic & Actionable
Research in Education

III. Final thoughts

Key takeaways

- RELM sensitive to variation in an L2+, but not in the L1
 - Grade 3 (L1) RELMs: limited discriminatory ability, most likely because learners are being tested in a language they possess strong linguistic abilities
 - Grade 4 (L2+) RELMs: adequate reliability and validity evidence and enhanced discriminatory capacity
- Preliminary analysis of weak and strong reading comprehenders in Grade 4 (L2+) underscores the importance of language skills in reading development



SHARE
Supporting Holistic & Actionable
Research in Education

Thank you!

For more information about **RELM**, please contact Principal Investigator, **Fernanda Soares** at fassunca@nd.edu or RELM Research Associate, **Caroline Freeman** at cfreema5@nd.edu



SHARE
Supporting Holistic & Actionable
Research in Education

Annexes



Learnings from the process

- Challenges in identifying culturally and contextually relevant words and images applicable across whole countries and subpopulations
- Difficulty in determining vocabulary difficulty-level for each grade-level during assessment development
- Listening comprehension picture matching: some assessors reported students had difficulty comprehending instructions and recalling response options
- The adaptation process is longer and may require more time than adaptation of traditional literacy assessments.

Inter-Rater Reliability



SHARE
Supporting Holistic & Actionable
Research in Education

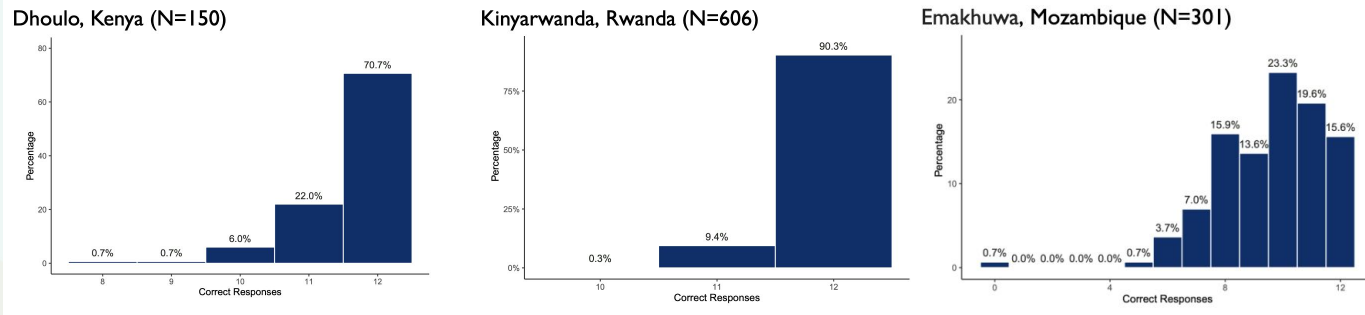
- High percent agreement between raters across all subtasks

Subtask	Kenya	Rwanda
Receptive vocabulary	99.12%	98.91%
Expressive vocabulary	97.5%	98.91%
Listening comp.	99.36%	98.72%

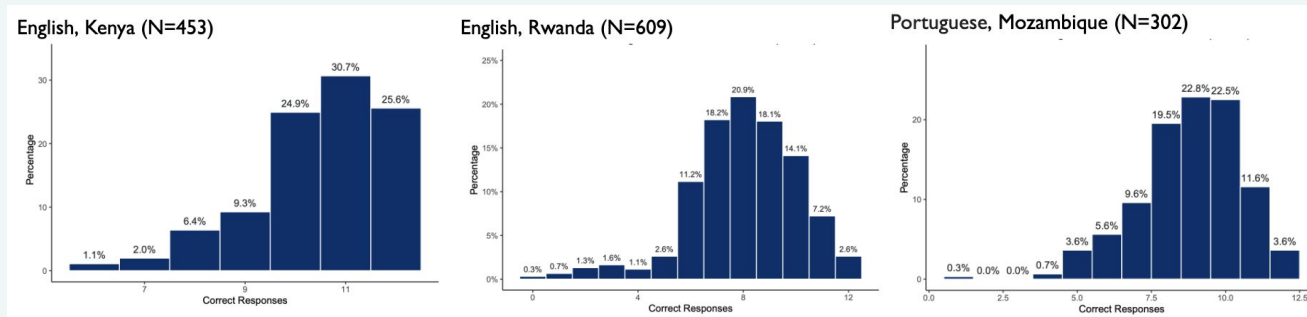
RECEPTIVE VOCABULARY: Grades 3 and 4 Score Distribution by Country



- Ceiling effect observed in **Grade 3** (L1) in Rwanda and Kenya, and right-skewed distribution in Mozambique



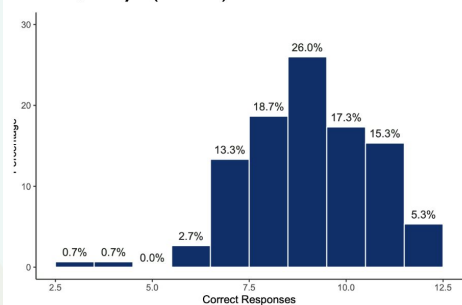
- Subtask capture more variability when assessing **Grade 4** (L2+), but with right-skewed distribution



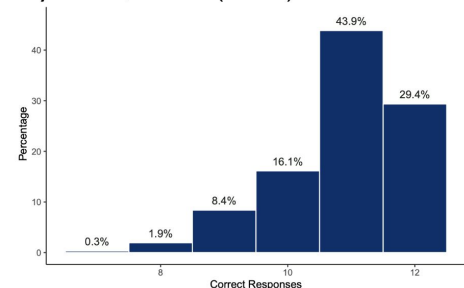
Expressive Vocabulary: Grades 3 and 4 Score Distribution by Country

- Ceiling effect observed in **Grade 3 (L1)** in Rwanda and right-skewed distribution in Kenya and Mozambique

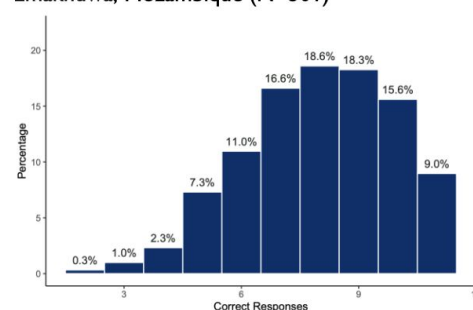
Dhoulou, Kenya (N=150)



Kinyarwanda, Rwanda (N=311)

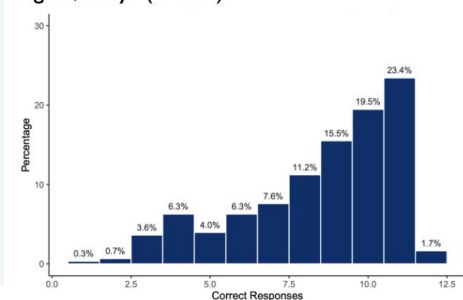


Emakhuwa, Mozambique (N=301)

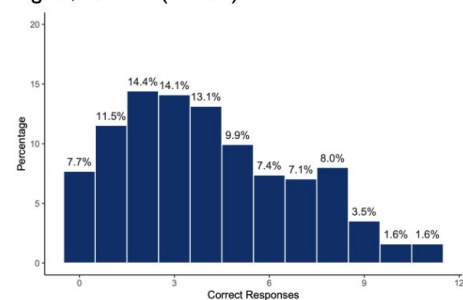


- Subtask capture more variability in **Grade 4 (L2+)**; left-skewed distribution in Rwanda and Moz.

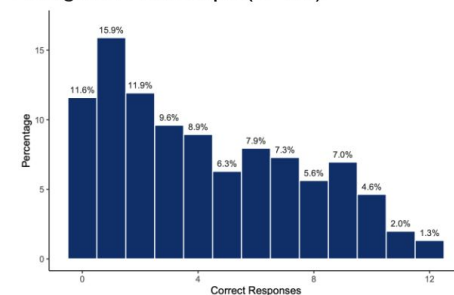
English, Kenya (N=303)



English, Rwanda (N=311)



Portuguese, Mozambique (N=301)



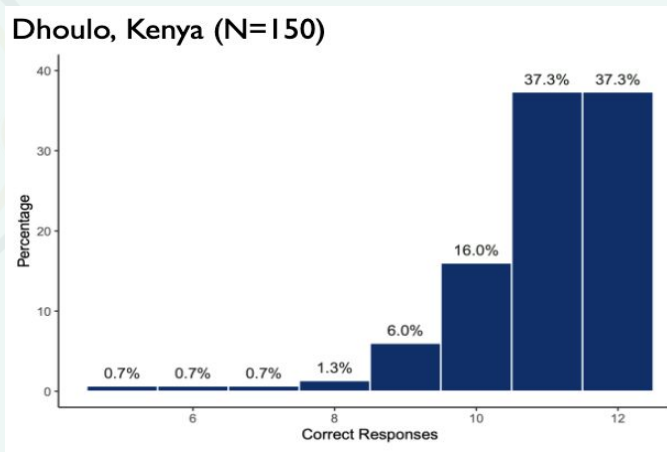
LISTENING COMPREHENSION PICTURE-MATCHING: Grades 3 and 4 Score Distribution in Kenya



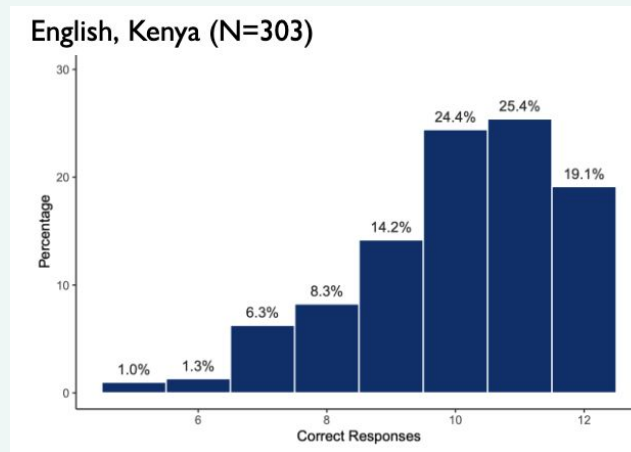
SHARE
Supporting Holistic & Actionable
Research in Education

- Ceiling effect observed in **Grade 3 (L1)**, right-skewed distribution in **Grade 4 (L2)**

Grade 3 (L1)

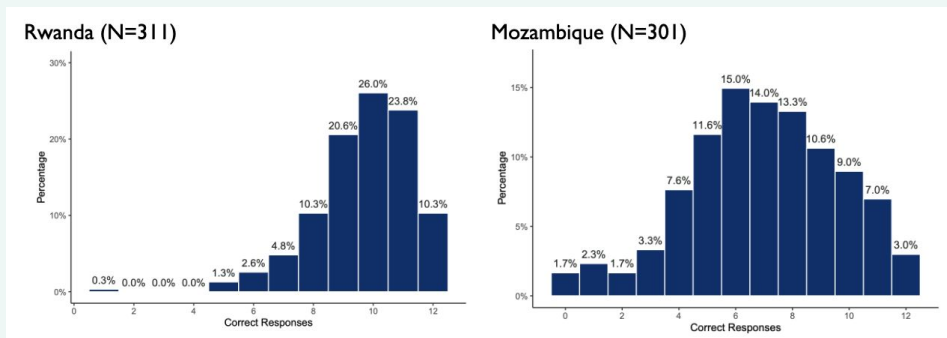


Grade 4 (L2)

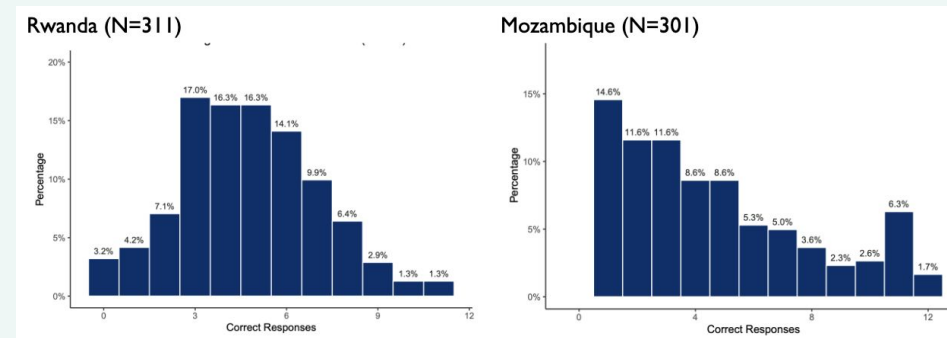


LISTENING COMPREHENSION SHORT-STORY: Grades 3 and 4 Score Distribution in Rwanda and Mozambique

Grade 3 (L1): right-skewed distribution in Rwanda, somewhat normally distributed in Moz.



Grade 4 (L2+): somewhat normally distributed in Rwanda, left-skewed distribution in Moz.



RECEPTIVE VOCABULARY:

Grade 3 Summary Statistics by Country



SHARE
Supporting Holistic & Actionable
Research in Education

- Ceiling effect observed in Grade 3 in Rwanda and Kenya when assessing students in an LI

	Mean	Median	Min	Max
Kenya (Dhoulou)	96.8	100	66.7	100
Rwanda (Kiswahili)	99.2	100	83.3	100
Mozambique (Emakhuwa)	80.5	83.3	41.7	100

Note: Scores are measured in percentage of correct items.

RECEPTIVE VOCABULARY:

Grade 4 Summary Statistics by Country



SHARE
Supporting Holistic & Actionable
Research in Education

- Ceiling effects no longer present when assessing Grade 4 students in an L2+ (second/later acquired language)
- Subtask capture more variability

	Mean	Median	Min	Max
Kenya (English)	87.5	91.7	50	100
Rwanda (English)	66.8	66.7	0	100
Mozambique (Portuguese)	73.3	75	8.3	100

EXPRESSIVE VOCABULARY:

Grade 4 Summary Statistics by Country



SHARE
Supporting Holistic & Actionable
Research in Education

	Mean	Median	Min	Max
Kenya (English)	70.8	75	8.3	100
Rwanda (English)	34	33.3	0	91.7
Mozambique (Portuguese)	35.5	33.3	0	100

Note: Scores are measured in percentage of correct items.

EXPRESSIVE VOCABULARY:

Grade 3 Summary Statistics by Country



SHARE
Supporting Holistic & Actionable
Research in Education

	Mean	Median	Min	Max
Kenya (Dhoulo)	75.3	75	25	100
Rwanda (Kiswahili)	90.8	91.7	58.3	100
Mozambique (Emakhuwa)	66.5	66.7	16.7	91.7

Note: Scores are measured in percentage of correct items.

Item Analysis



SHARE
Supporting Holistic & Actionable
Research in Education

- For Round 1 countries - recommendation to drop items that:
 - Were too easy or too difficult and did not discriminate well between higher- and lower-ability students
 - Had corresponding images that did not work well (captured through qualitative debriefs):
 - Images that did not capture local representation of noun or verb
 - Issues with image clarity
- For Round 2 (Philippines):
 - Recommendation to pilot more items than needed per subtask
 - Strengthened process to estimate item-difficulty before pilot

Relationship between RELM and Reading Comprehension

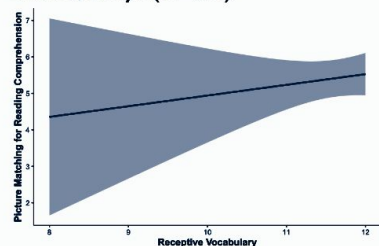


SHARE
Supporting Holistic & Actionable
Research in Education

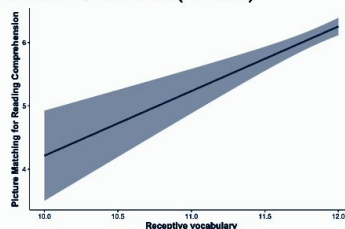
- Positive (linear) relationship between Receptive Vocabulary in Grades 3 and 4 and Reading Comprehension

Grade 3

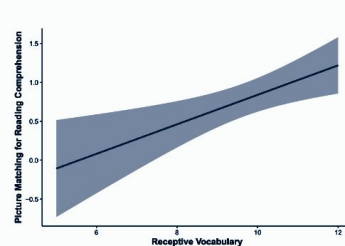
Dhoulou, Kenya (N=150)



Kiswahili, Rwanda (N=606)

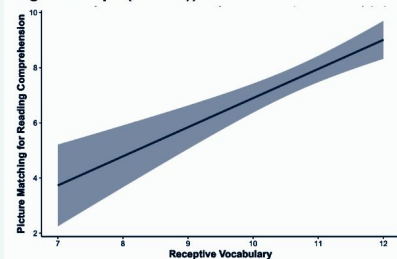


Emakhuwa, Mozambique (N=301)

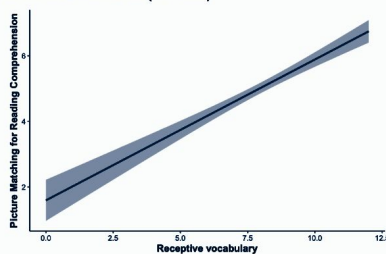


Grade 4

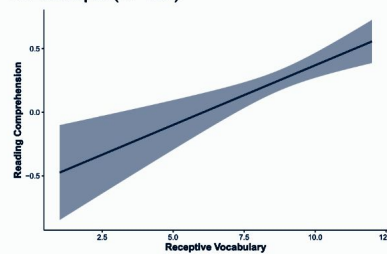
English, Kenya (N=454)



Kiswahili, Rwanda (N=606)



Mozambique (N=301)



Relationship between RELM and Reading Comprehension

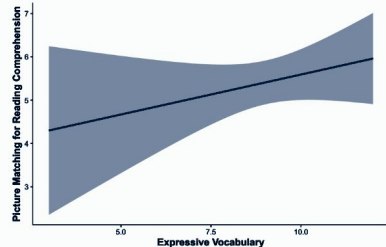


SHARE
Supporting Holistic & Actionable
Research in Education

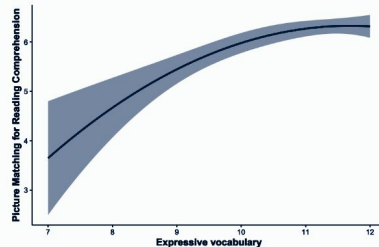
- Mostly positive relationship between Expressive Vocabulary in Grades 3 and 4 and Reading Comprehension

Grade 3

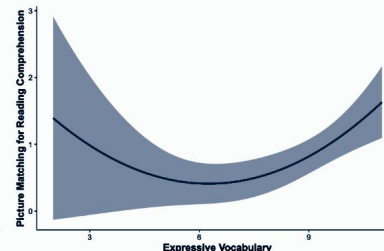
Dhoulou, Kenya (N=150)



Kiswahili, Rwanda (N=311)

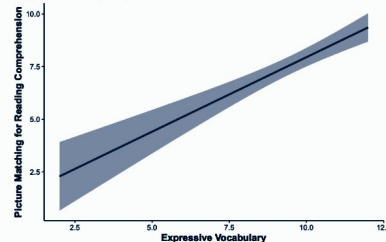


Mozambique (N=301)

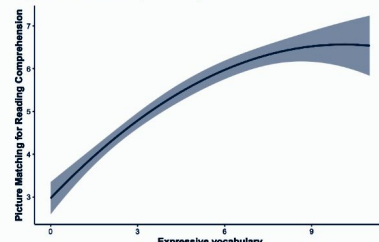


Grade 4

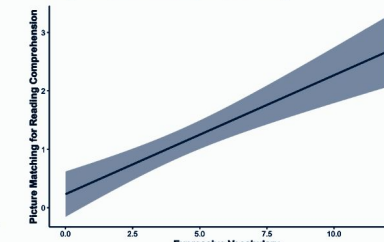
English, Kenya (N=150)



English, Rwanda (N=311)



Portuguese, Mozambique (N=301)

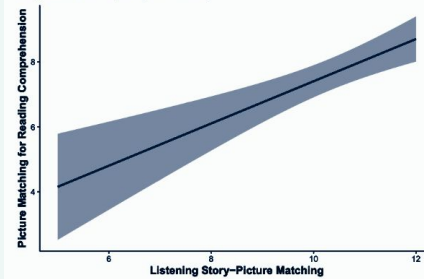


Relationship between RELM and Reading Comprehension

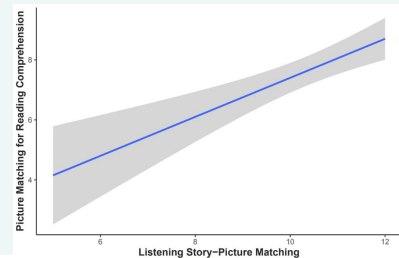
- Positive relationship between Listening-story Picture Matching in Grades 3 and 4 and Reading Comprehension

Grade 3

English, Kenya (N=303)

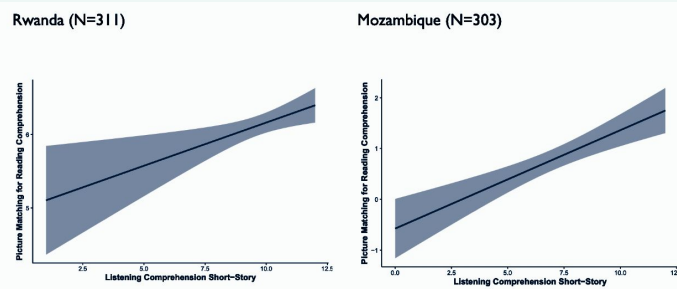


Grade 4

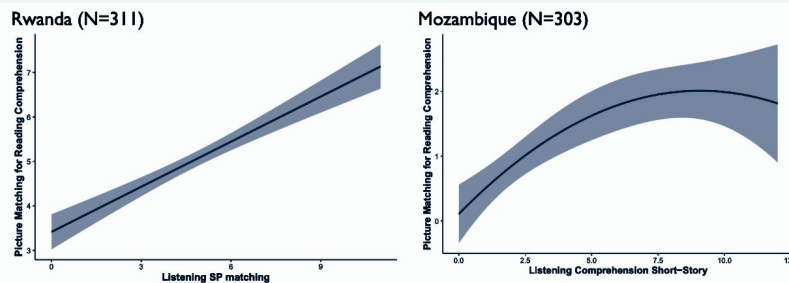


Relationship between RELM and Reading Comprehension

- Positive relationship between Listening Comp Short Story in Grades 3 and 4 and Reading Comprehension
- Grade 3



Grade 4

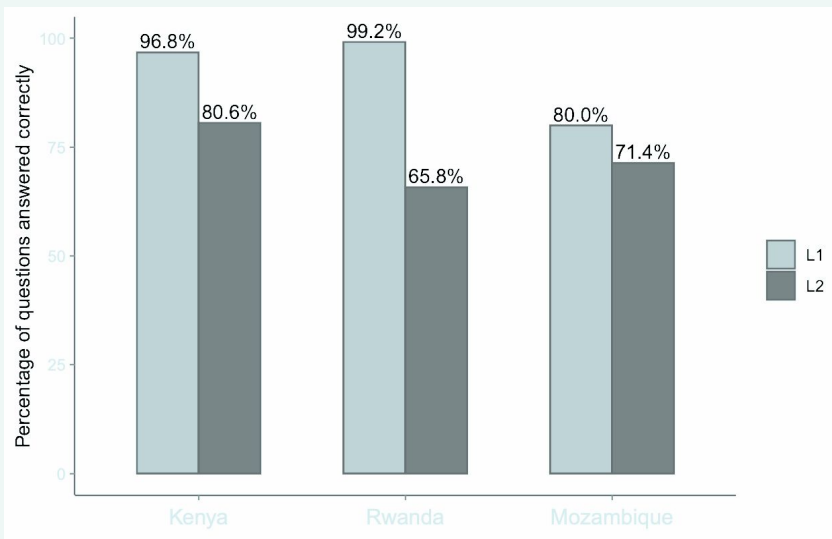


RECEPTIVE VOCABULARY:

Percentage of Items Answered Correctly by Grade 3 Students, by Language of Assessment



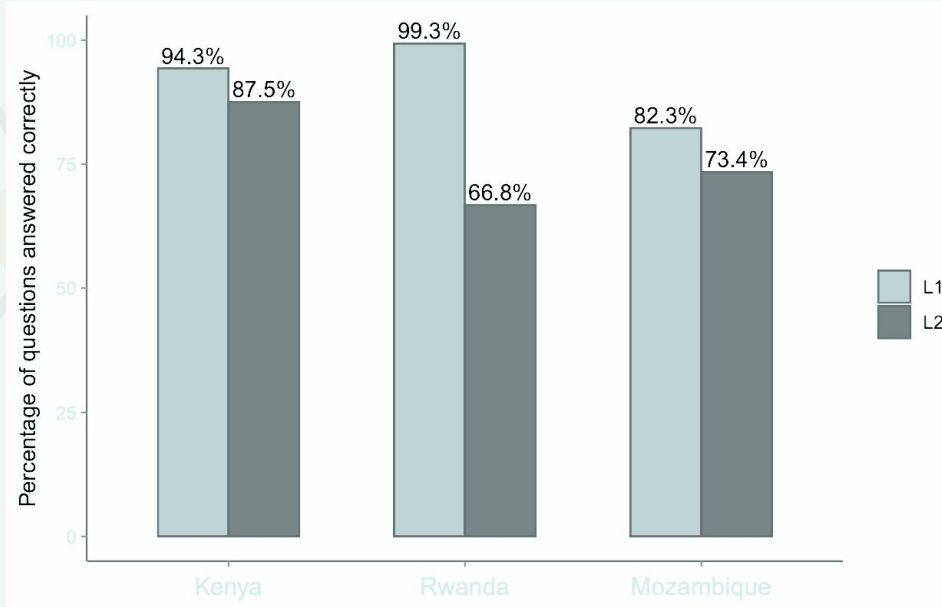
SHARE
Supporting Holistic & Actionable
Research in Education



- Students answered the largest proportion of question correctly in the L1, compared to L2
- Ceiling effect observed in Grade 3 in Rwanda and Kenya

RECEPTIVE VOCABULARY:

Percentage of Items Answered Correctly by Grade 4 Students, by Language of Assessment



- Students answered the largest proportion of question correctly in the L1, compared to L2
- Ceiling effect observed in Grade 4 in Rwanda and Kenya

Updated Adaptation Guidance: Expressive and Receptive Vocabulary



SHARE
Supporting Holistic & Actionable
Research in Education

Adaptation process for Receptive and Expressive Vocabulary Subtasks



Updated Adaptation Guidance: Expressive and Receptive Vocabulary



SHARE
Supporting Holistic & Actionable
Research in Education

- a) Item bank for possible nouns and verbs
 - i) Select 42 nouns and 15 verbs learned incidentally
 - ii) Decide on difficulty level based on subjective Age of Acquisition and word complexity
- b) Image bank to represent selected nouns and verbs
 - i) Ensure adequate representation of all sub-groups participating in the assessment
 - ii) inclusion of AI platforms such as Dall-E
- c) Updated expert review process
 - i) assessment of words evoked by images in relevant languages
 - ii) rating of pictures and words based on context, age and cultural appropriateness and difficulty
 - iii) analyze ratings and update item bank accordingly
- d) Finalize draft item bank and finalize subtasks

Example expert review assessment



SHARE
Supporting Holistic & Actionable
Research in Education

Image Assessment Sheet Template

[INSERT IMAGE 1]

- 1) How many words in [LANGUAGE] come to your mind when you see this image?

- 1) Please write the name of the word or words that come to your mind when you see this image

- 1) Do you consider this image to be appropriate to [NAME*] context and culture (is the object/action available for direct experience in your context)?
 - a) Yes
 - b) No
- 2) Do you consider this image to be appropriate to [AGE] age range (i.e., will children have had access/exposure to the object/action)?
 - a) Yes
 - b) No
- 3) When do children in [CONTEXT NAME] learn this word? Please estimate the age (in years) at which you think most children learn this word

- 1) Have you seen this word in [LANGUAGE] in any school textbooks or other teaching and learning materials? If yes, beginning at what grade level?

- 1) How difficult do you think it is for children [AGE RANGE] to understand the meaning of this word?
 - b) Easy
 - c) Medium
 - d) Difficult

Updated Adaptation Guidance: Listening Comprehension



SHARE
Supporting Holistic & Actionable
Research in Education

- I) Listening Comprehension Picture-Matching
 - a) Select 52 images based on scenes or scenarios familiar to the student sample
 - b) Develop simple stories reflective of local context and including relevant sub-populations
- 2) Listening Comprehension Multiple Choice
 - a) Select two short stories (~60 words) with a narrative structure and a plot reflective of culture and context
 - b) Develop comprehension questions with multiple-choice responses
 - i) 2 explicit
 - ii) 2 explicit inter-sentential
 - iii) 1 inferential

Additional Adaptation Guidance

- Adaptation team:
 - Should consist of local individuals such as teachers, Special Needs Education professionals, linguists, Ministry representatives
 - Adaptation team should participate in an adaptation workshop to build out subtasks (currently building)
- Expert Review
 - Recruit between 6-10 local former or current teachers/ teacher trainers to anonymously assess vocabulary and image selection
- Translation
 - Two cultural/ linguistic experts independently translate and then collaborate on final materials
- Subtask adaptation:
 - Image/story selection should ensure adequate representation of the context of the student population as well as all represented populations

Semantic Fluency



SHARE
Supporting Holistic & Actionable
Research in Education

Student **says as many words as possible** about the image, or anything else that comes to mind, in 60 seconds. The child will repeat the subtask in **up to three languages** they reported to use or understand.

The assessor will tally the total number of words the student says and score the student with that number.

We have 2 images that we need to recycle depending on the number of languages. For example, if the child says he/she speaks:

- 1 language we will use Image 1
- 2 languages we will use Image 1 for the first language and Image 2 for second language
- 3 languages we will use Image 1 for the first language, Image 2 for second language, and Image 1 again for the third language





USAID
FROM THE AMERICAN PEOPLE



SHARE
Supporting Holistic & Actionable
Research in Education

The End

The University of Notre Dame's Pulte Institute for Global Development is implementing the SHARE activity in partnership with local higher education and research institutions. Through SHARE and its partners, UND will advance USAID's Learning Priorities as a global good to improve education and learning outcomes as part of cooperative agreement No. 7200AA20CA00025. The contents of this presentation are the sole responsibility of the Pulte Institute and do not necessarily reflect the views of USAID or the United States Government.