Children's Storybooks and Early Literacy in Rural Kenya: Evidence from 1.5 Randomized Evaluations

October 2019

Lia Fernald, Pamela Jakiela, Heather Knauer, and Owen Ozier

Motivation

Investments in early childhood are critical to adult outcomes

Over 250 million children may not reach their developmental potential because of inadequate nutrition and stimulation in early childhood

- Inadequate stimulation during childhood impacts cognitive development, human capital, and income throughout adult life

Early childhood interventions can work, but they are often expensive

- Home visits by child development specialists (Gertler et al. 2014)

School-based interventions can be low-cost, but not always effective

- Many children do not attend preschool (Martinez et al. 2017)
- Difficult to improve school quality without engaging parents; they may have different objectives (Özler et al. 2018, Wolf et al. 2019)

Motivation

Many ECD interventions rely on household members as mediators

- Many parenting interventions run counter to cultural norms, e.g. in Sub-Saharan Africa (Jukes et al. 2018, Weber et al. 2017)

Storybooks are a nudge reminding parents to stimulate children

- Only 3 percent of African children have at least 3 storybooks (MICS)
- Mother tongue children's storybooks are almost never available
- Many parents do not realize importance of reading to children who have not started school and are not yet learning to read

Motivation

Many ECD interventions rely on household members as mediators

- Many parenting interventions run counter to cultural norms, e.g. in Sub-Saharan Africa (Jukes et al. 2018, Weber et al. 2017)

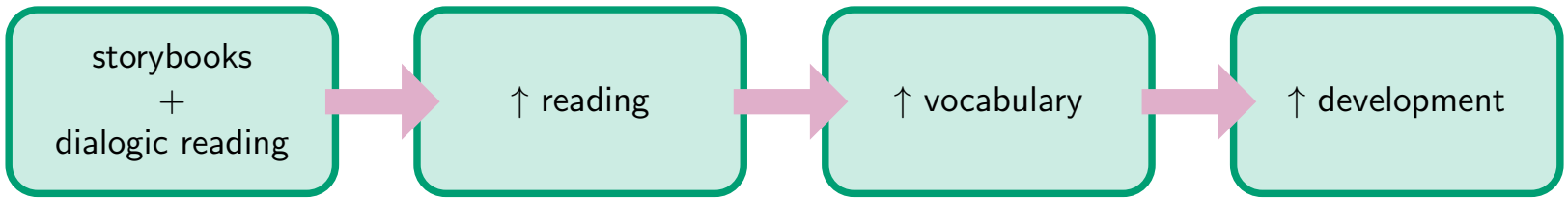
Storybooks are a nudge reminding parents to stimulate children

- Only 3 percent of African children have at least 3 storybooks (MICS)
- Mother tongue children's storybooks are almost never available
- Many parents do not realize importance of reading to children who have not started school and are not yet learning to read

How can we encourage parents to read with young children?

The EMERGE Project

The EMERGE Project



Storybooks are a fundamental technology for building pre-literacy skills

- Reading with preschool-aged children is uncommon in many LMICs
- Introducing parents to **dialogic reading** approach may improve both quantity and quality of shared reading, making it more interactive
- Reading — particularly dialogic reading — have been shown to improve children's vocabulary (mostly in high-income countries)
- Can a light-touch intervention change household (reading) behavior?

The EMERGE Project

2014 **Development of the EMERGE intervention**

Development of locally-appropriate storybooks for young children

Adaptation of modified dialogic reading training for caregivers

2015 Adaptation, validation of appropriate measures of child development

2016 **Within-village randomized trial to estimate short-term impacts**

350 households assigned to treatment arms varying in intensity

2017 **Cluster-randomized trial in 73 communities**

Census of all young children in 73 rural communities

Baseline survey of approximately 2,500 children aged 3–6

2018 Intervention delivered in 36 randomly-selected communities

Midline survey of random subset of children and households

Endline survey of caregivers, sample children, and siblings

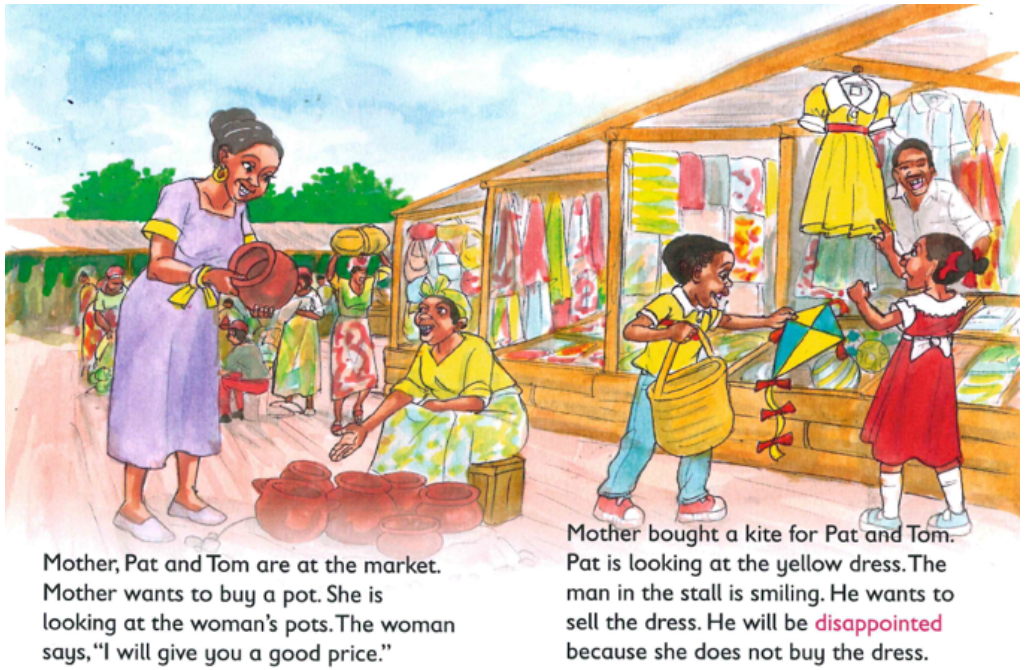
2019

Overview of the Talk

1. Development of the EMERGE intervention
2. Within-village pilot study documenting short-term impacts
3. Design of the cluster-randomized evaluation
4. Short-term impacts on shared reading
5. Measuring impacts at endline
6. Pre-analysis plan

Intervention Design

The EMERGE Project



Mother, Pat and Tom are at the market. Mother wants to buy a pot. She is looking at the woman's pots. The woman says, "I will give you a good price."

Mother bought a kite for Pat and Tom. Pat is looking at the yellow dress. The man in the stall is smiling. He wants to sell the dress. He will be disappointed because she does not buy the dress.

We create **local-language storybooks** appropriate for young children

- Adapt locally-appropriate English-language illustrated storybooks
- Translate into Luo language, the mother tongue of study children

The EMERGE Project

We develop a modified **dialogic reading** training for primary caregivers

- Adapted from dialogic reading interventions in Bangladesh (Aboud et al. 2009) and South Africa (Murray et al. 2016, Vally et al. 2015)
- Highlights critical importance of engaging young children in book-centered conversations (not just reading "to" them)
- Empowers illiterate caregivers to use books to stimulate a dialogue

Pilot Study

Evaluating the Short-Term Impacts of EMERGE

Existing evidence: dialogic reading improved child vocabulary in HICs, or when teachers/childcare providers in LMICs are trained (Mol et al. 2008)

- Study by Vally et al. (2015) finds positive effects of modified dialogic reading for mothers on young toddlers in South Africa
- Existing literacy interventions in SSA focus on school-age children

Evaluating the Short-Term Impacts of EMERGE

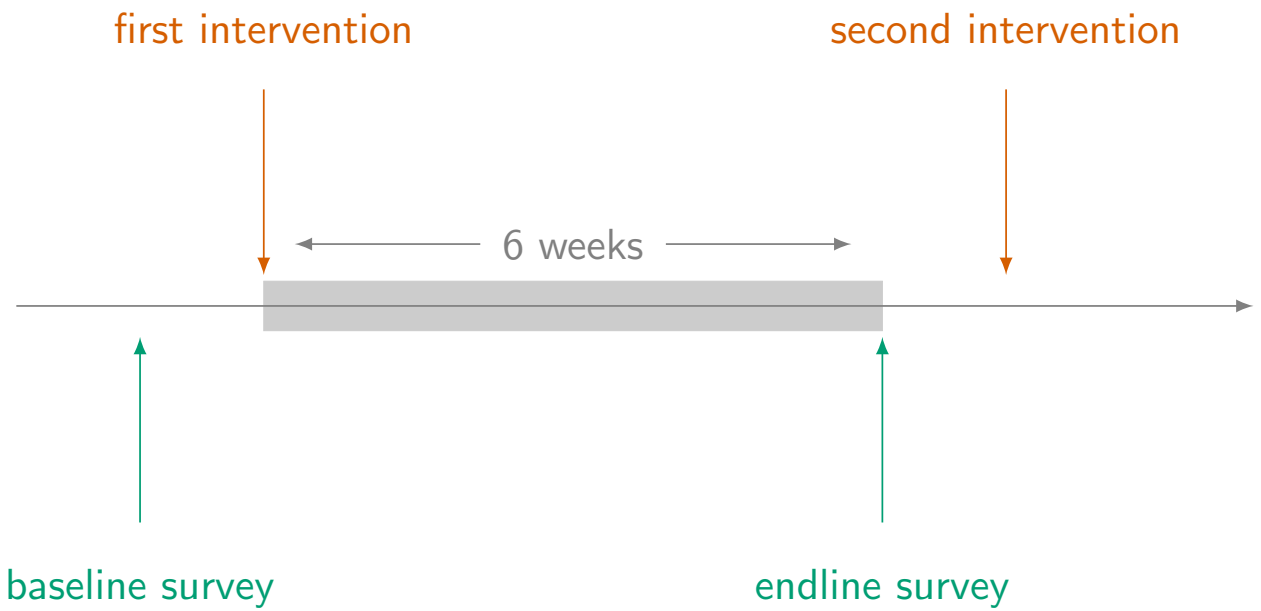
Existing evidence: dialogic reading improved child vocabulary in HICs, or when teachers/childcare providers in LMICs are trained (Mol et al. 2008)

- Study by Vally et al. (2015) finds positive effects of modified dialogic reading for mothers on young toddlers in South Africa
- Existing literacy interventions in SSA focus on school-age children

EMERGE intervention is (was) new, largely untested

- Do households value, use (local language) children's storybooks?
- Does our modified dialogic reading program change behavior?
- Do children learn new vocabulary words from the storybooks?
- What is the optimal training intensity? Is training necessary?

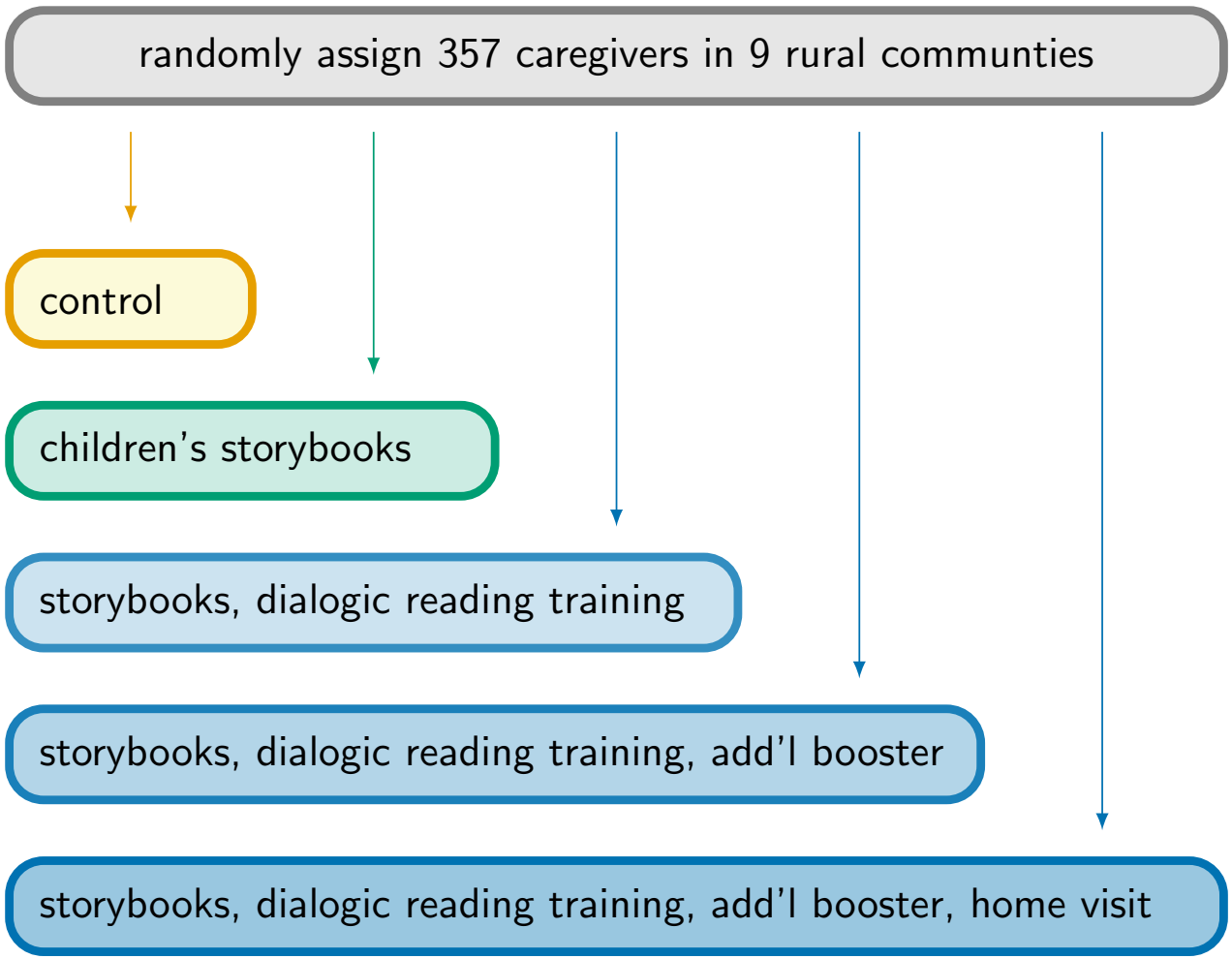
Evaluating the Short-Term Impacts of EMERGE



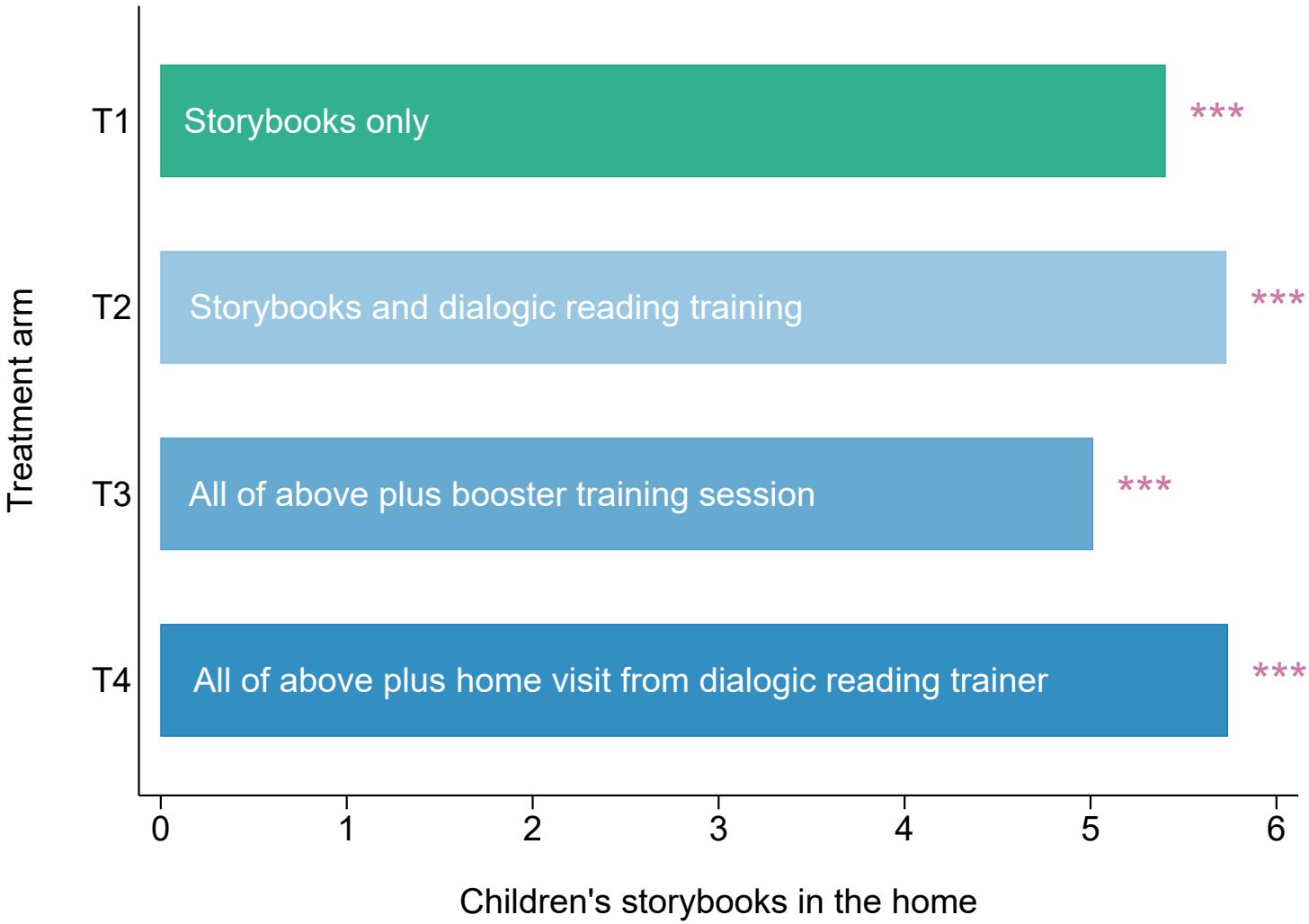
Within-community pilot study estimates short-term impacts of EMERGE

- Partially addresses concerns about (within-community) spillovers
- Parallels existing literature on dialogic reading in LMICs

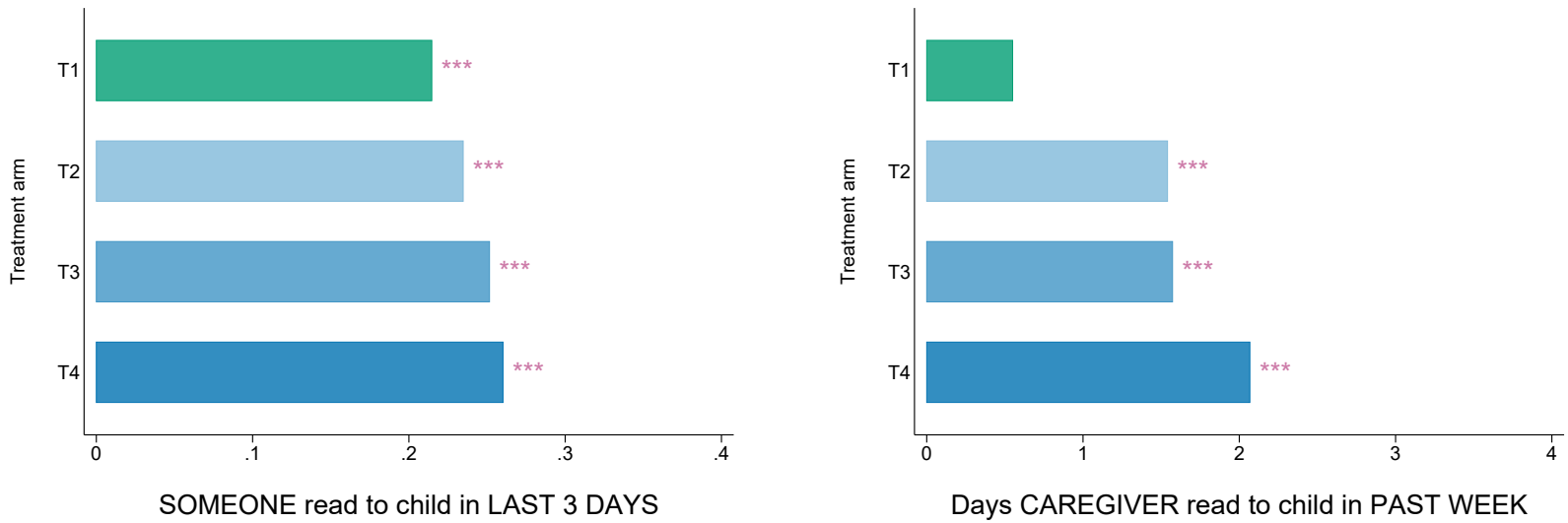
Evaluating the Short-Term Impacts of EMERGE



Short-Term Impacts on Home Literacy Environment



Short-Term Impacts on Reading Behaviors



All treatments increase (self-reported) likelihood of reading with child

- Dialogic reading training increases likelihood caregiver herself reports reading with child, and likelihood of shared reading almost every day
- In treatment arms, almost all households report some shared reading (control group mean in left panel of figure is 70 percent)

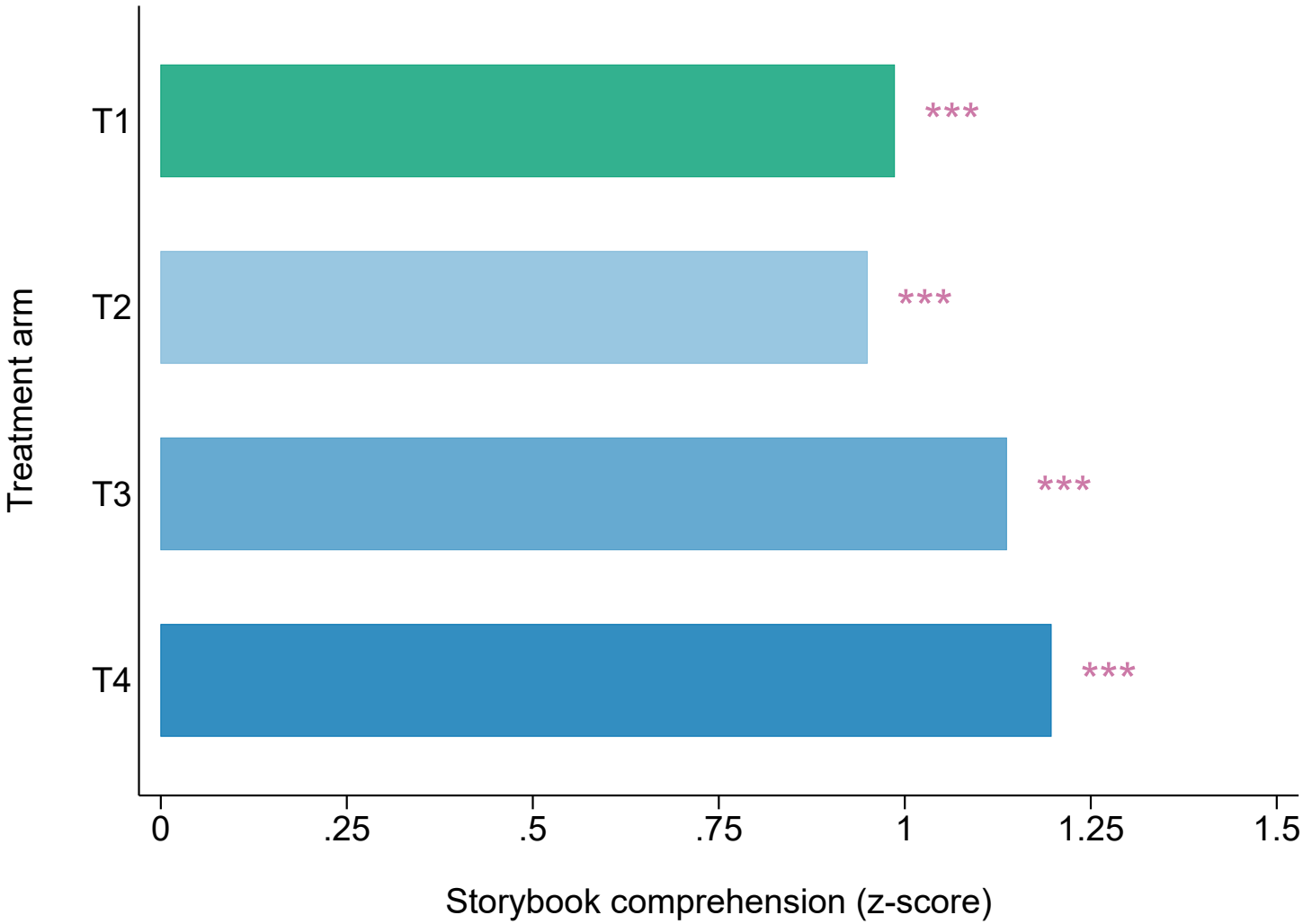
Short-Term Impacts on Familiarity with Stories

What 3 things did Caro forget at home today?

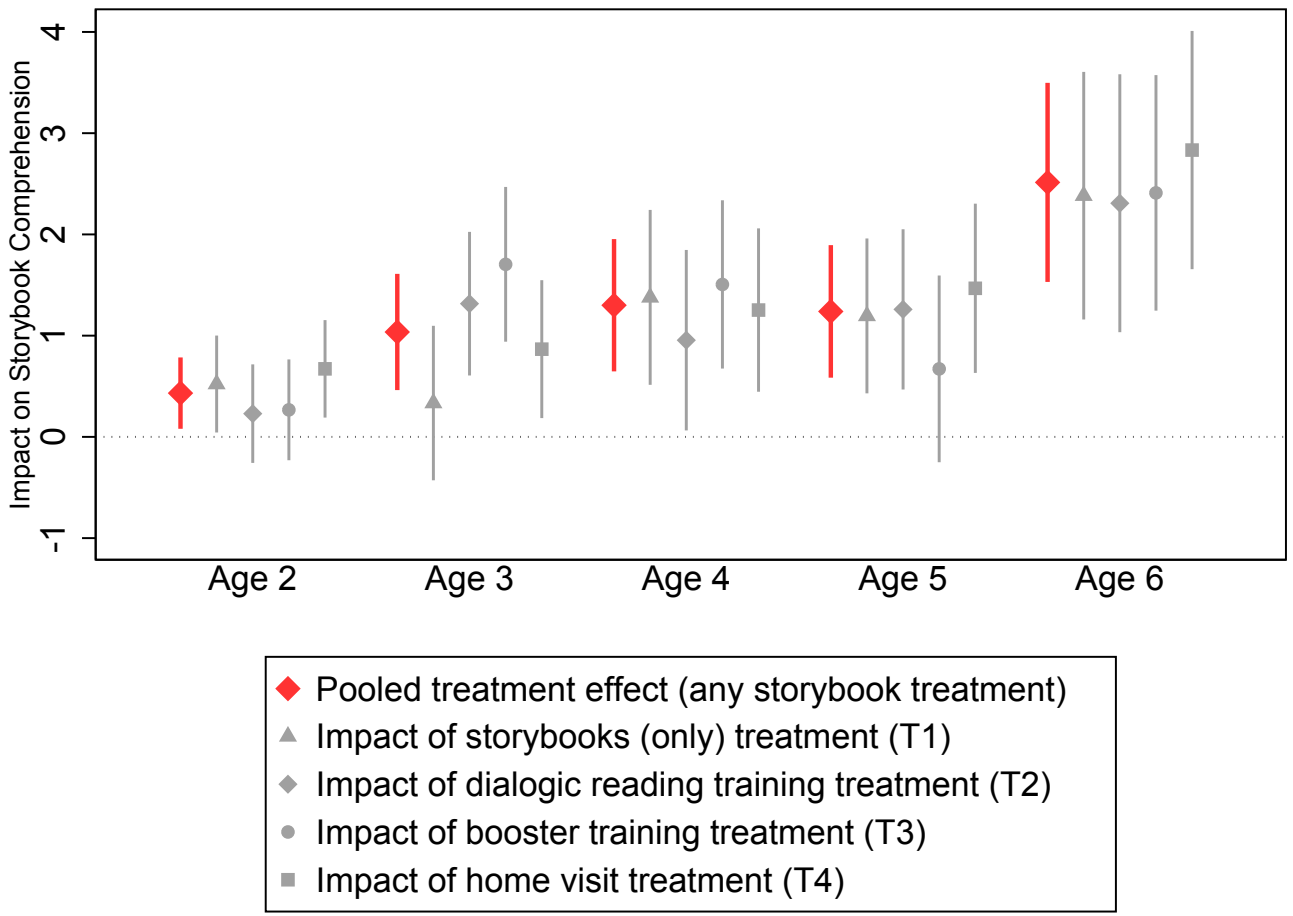
What do you think Meg and Ben are looking at?

We developed an objective measure of children's use of storybooks: 12 comprehension questions that children would be unlikely to guess

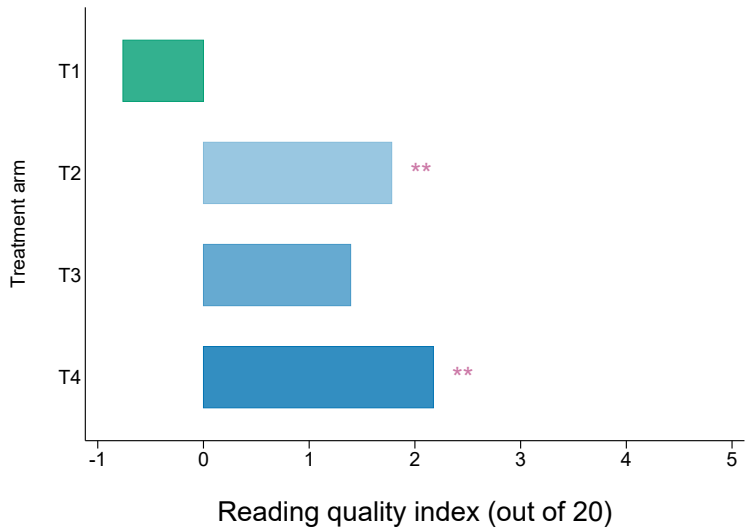
Short-Term Impacts on Familiarity with Stories

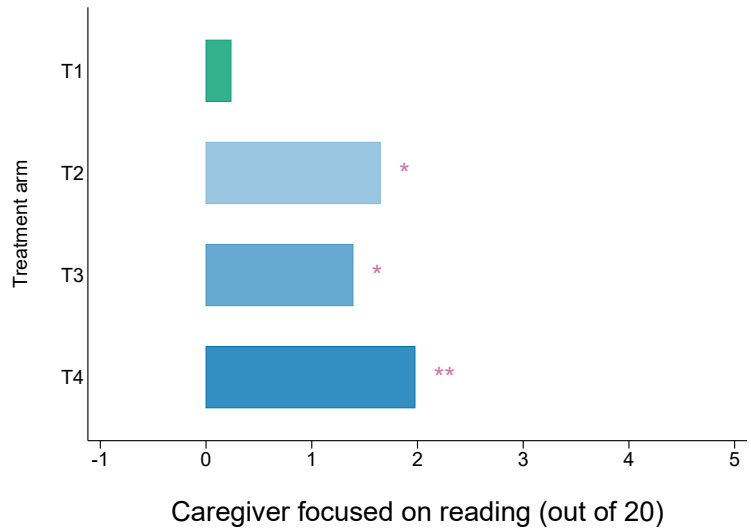


Short-Term Impacts on Familiarity with Stories



Short-Term Impacts on Reading Quality

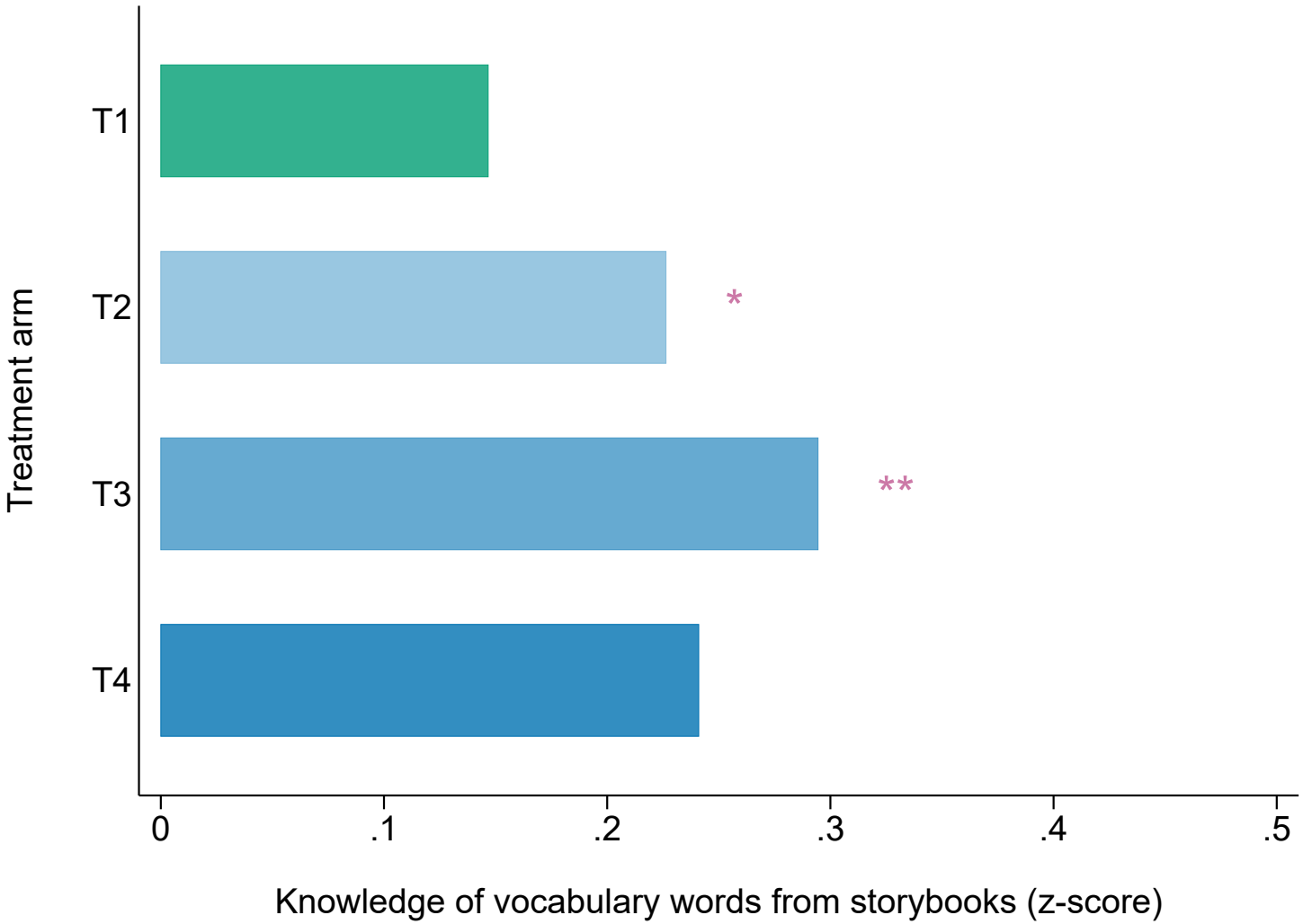




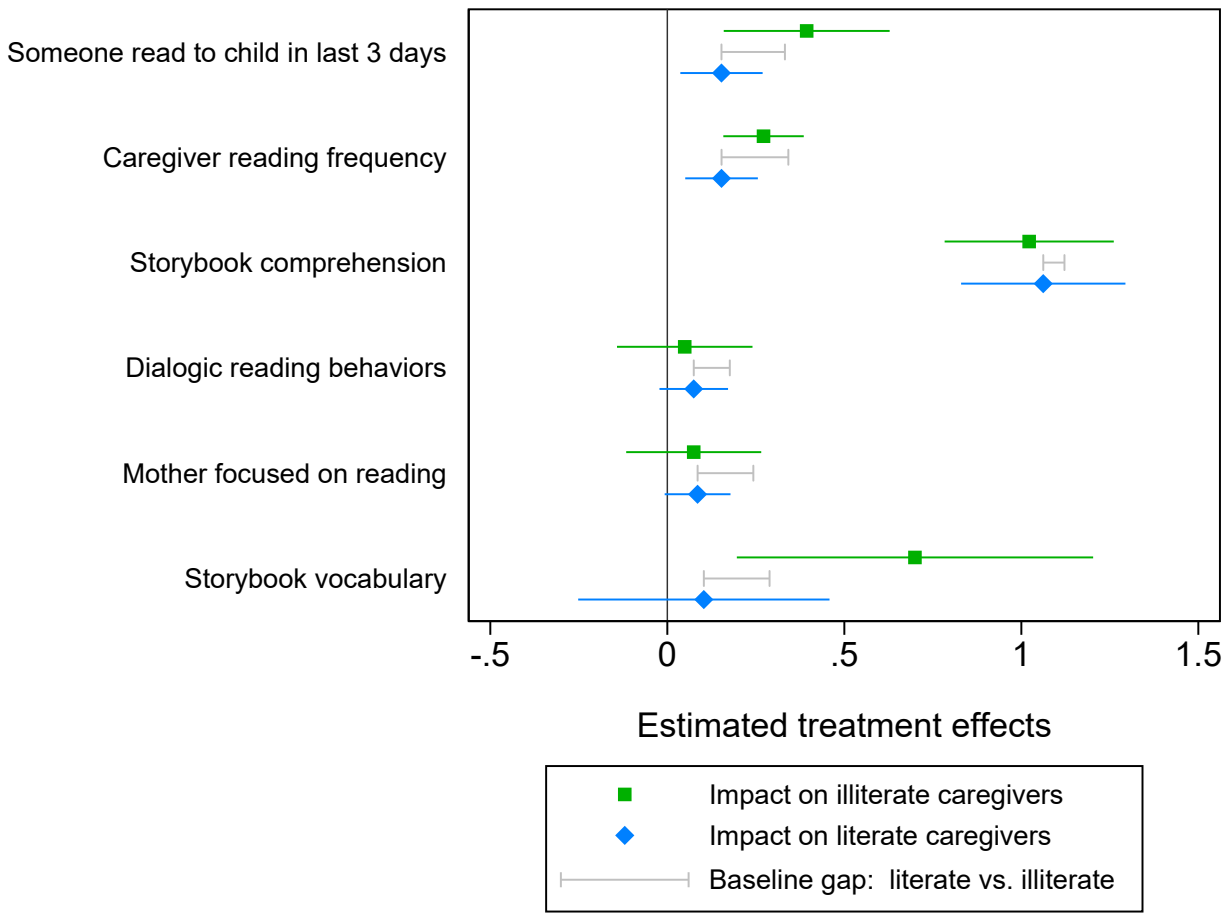
Measure reading quality using **Mother Child Picture Observation** tool

- Caregivers demonstrate reading (a new book) with a child
- Enumerators code types of reading activities every 20 seconds

Short-Term Impacts on Storybook Vocabulary



Heterogeneity by Caregiver Literacy



Short-Term Impacts of the EMERGE Intervention

Delivering books increased book-sharing in this context

- Large impacts of treatment on (self-reported) likelihood of reading with your child, children's comprehension of storybook content
- Training for parents does not have any additional impact (here)

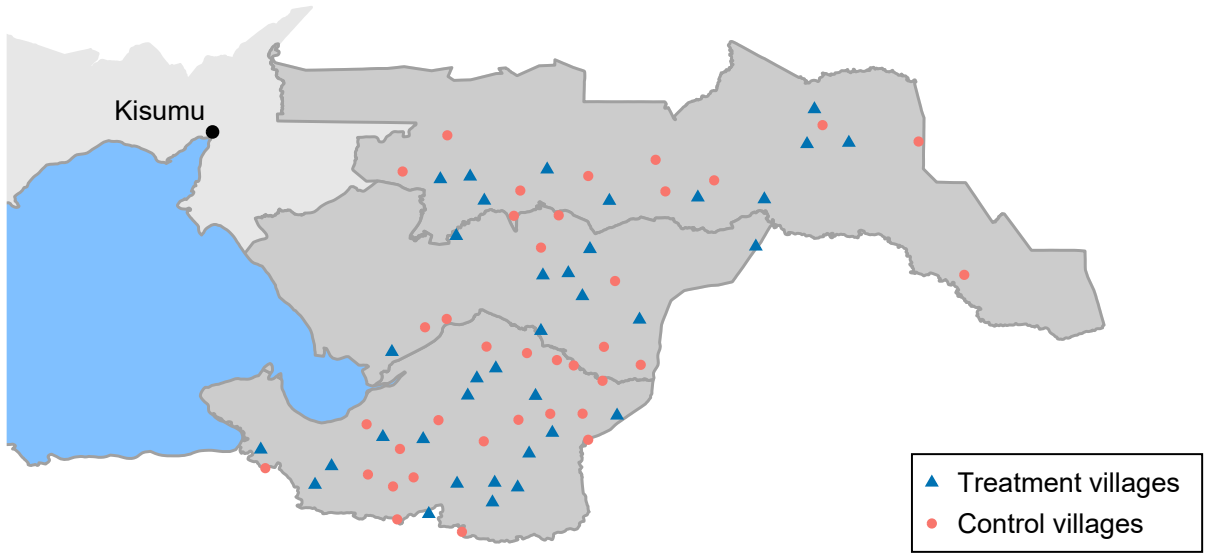
Dialogic reading training improves quantity, quality of reading

- Positive impacts on reading frequency (times per week)
- Positive impacts on quality of caregiver-child reading interactions
- Positive impacts on storybook vocabulary words

Booster training, home visits did not appear to increase impacts

Cluster-Randomized Evaluation

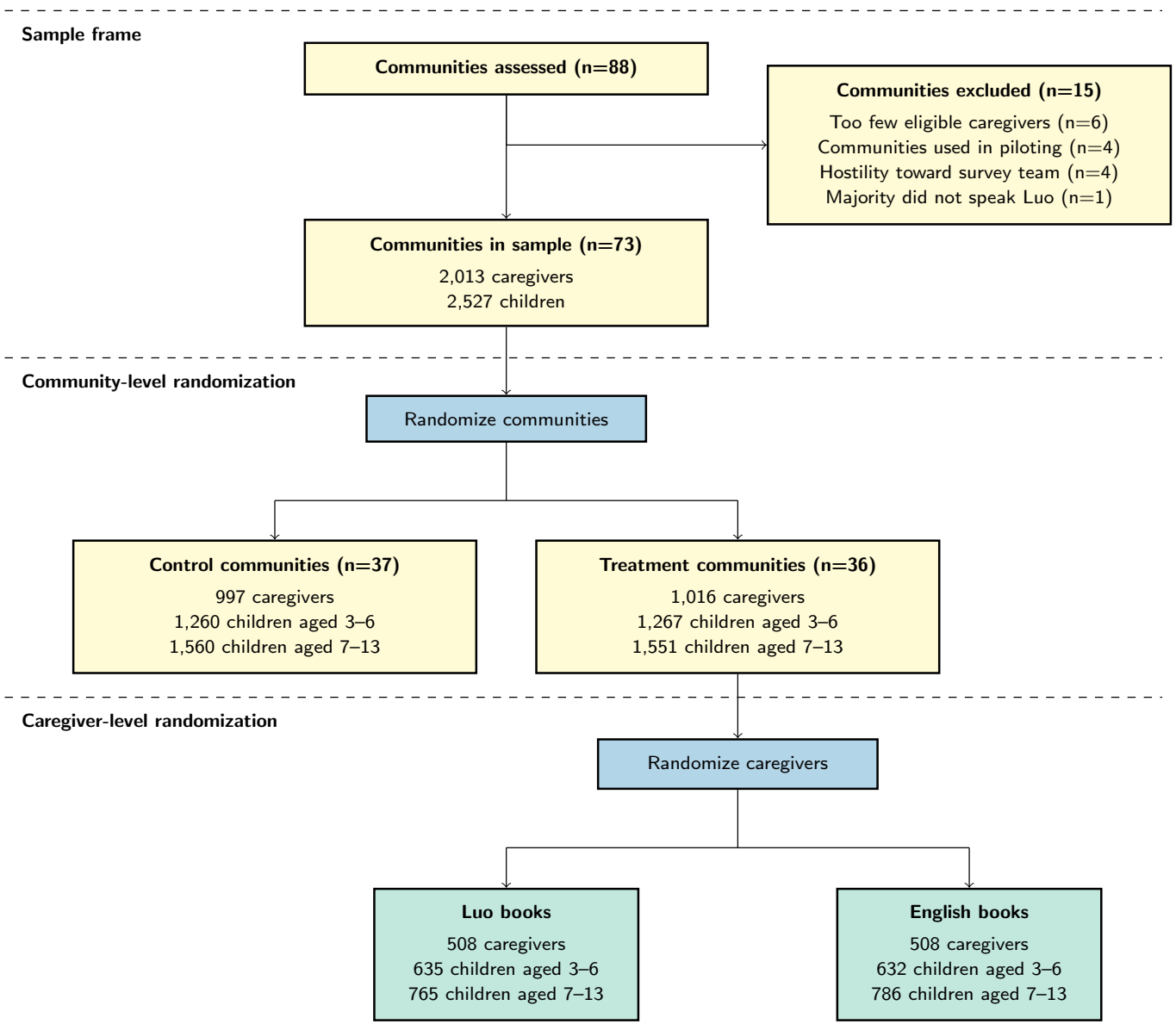
Research Design



Main study is a cluster-randomized trial in 73 rural communities

- Baseline sample frame includes all households with children aged 36 to 83 months living within 750 meters of a focal primary school

Research Design



Intervention Delivery

February and April of 2018: delivered treatment to 36 communities

- 88 percent of households sent at least one adult to the training
- We distributed storybooks to a further 78 households, so 95 percent of children in treatment communities received the EMERGE books

Short-Term Impacts at Midline

Estimating Short-Term Treatment Effects

We report intent-to-treat estimates of (pooled) EMERGE treatment:

$$Y_{ic} = \alpha + \beta EMERGE_c + \lambda_i + \delta_c + \varepsilon_{ic}$$

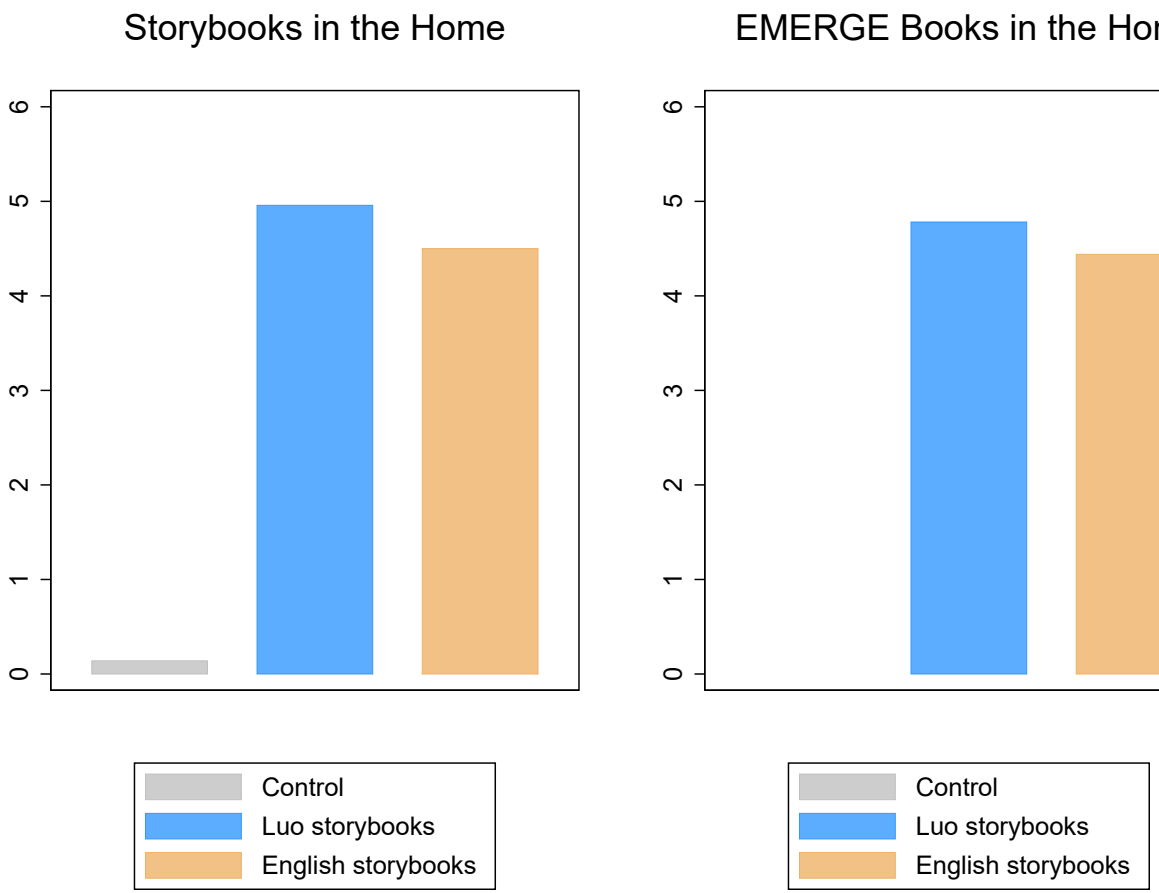
where:

- Y_{ic} is an outcome of interest for child i in community c
- $EMERGE_c$ is an indicator for (community-level) treatment
- λ_i is a vector of age fixed effects
- δ_c is a vector of stratum fixed effects
- ε_{ic} is a conditionally mean-zero error term

Midline survey: random sample of 300 children/households

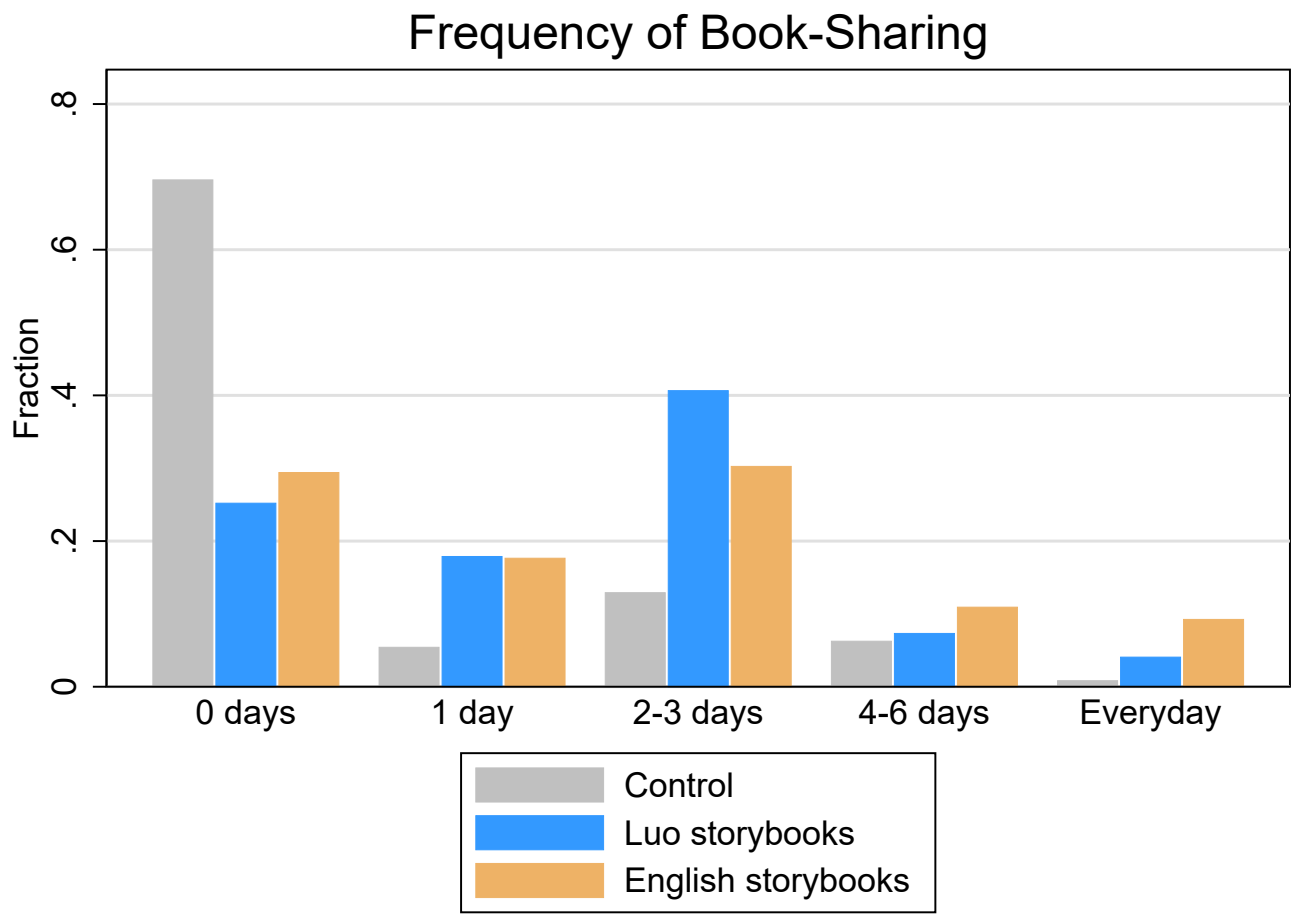
Short-Term Impacts

Result 1: the intervention improved the home literacy environment



Short-Term Impacts

Result 2: the intervention increased the quantity of book-sharing



Short-Term Impacts

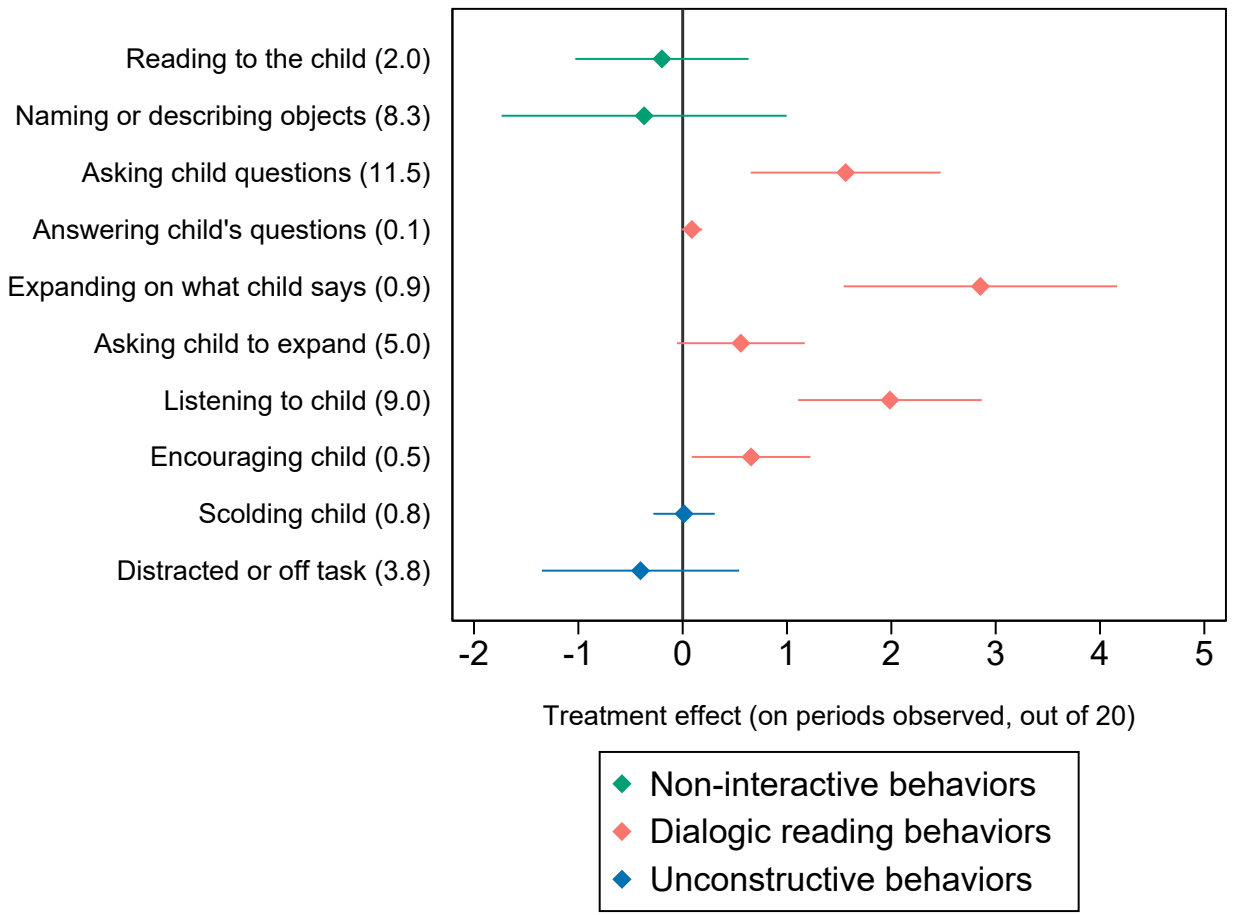
Result 3: the intervention increased the quality of book-sharing

We measure the quality of caregiver-child book-sharing activities through the **Mother Child Picture Observation** Assessment

- Caregivers provided with an unfamiliar storybook
- Enumerators observe caregiver reading with her child
- Behaviors coded in 20-second intervals:
  - Traditional (i.e. "non-interactive") reading
  - Dialogic reading
  - Other behaviors (e.g. scolding the child)

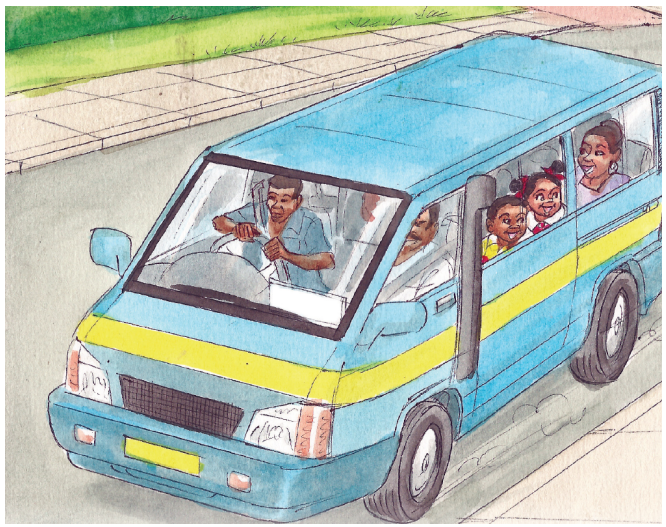
Short-Term Impacts

Result 3: the intervention increased the quality of book-sharing



Short-Term Impacts

Result 4: children use the EMERGE books and learn the stories

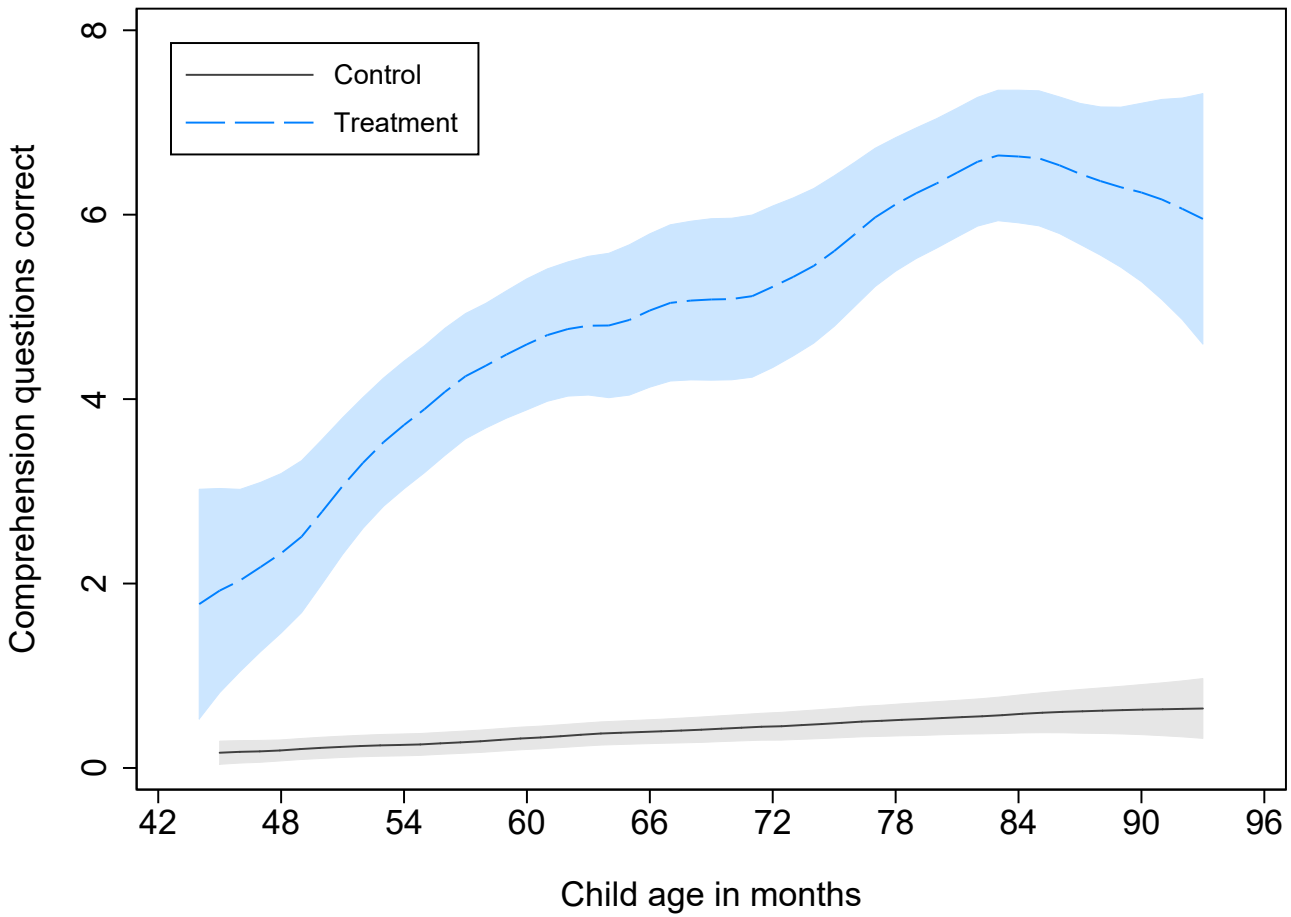


We developed a series of comprehension questions that children would be unlikely to guess (e.g. "Where is Ben going?" or "Why is Ben sad?")

$\Rightarrow$ Provides an objective measure of use of storybooks

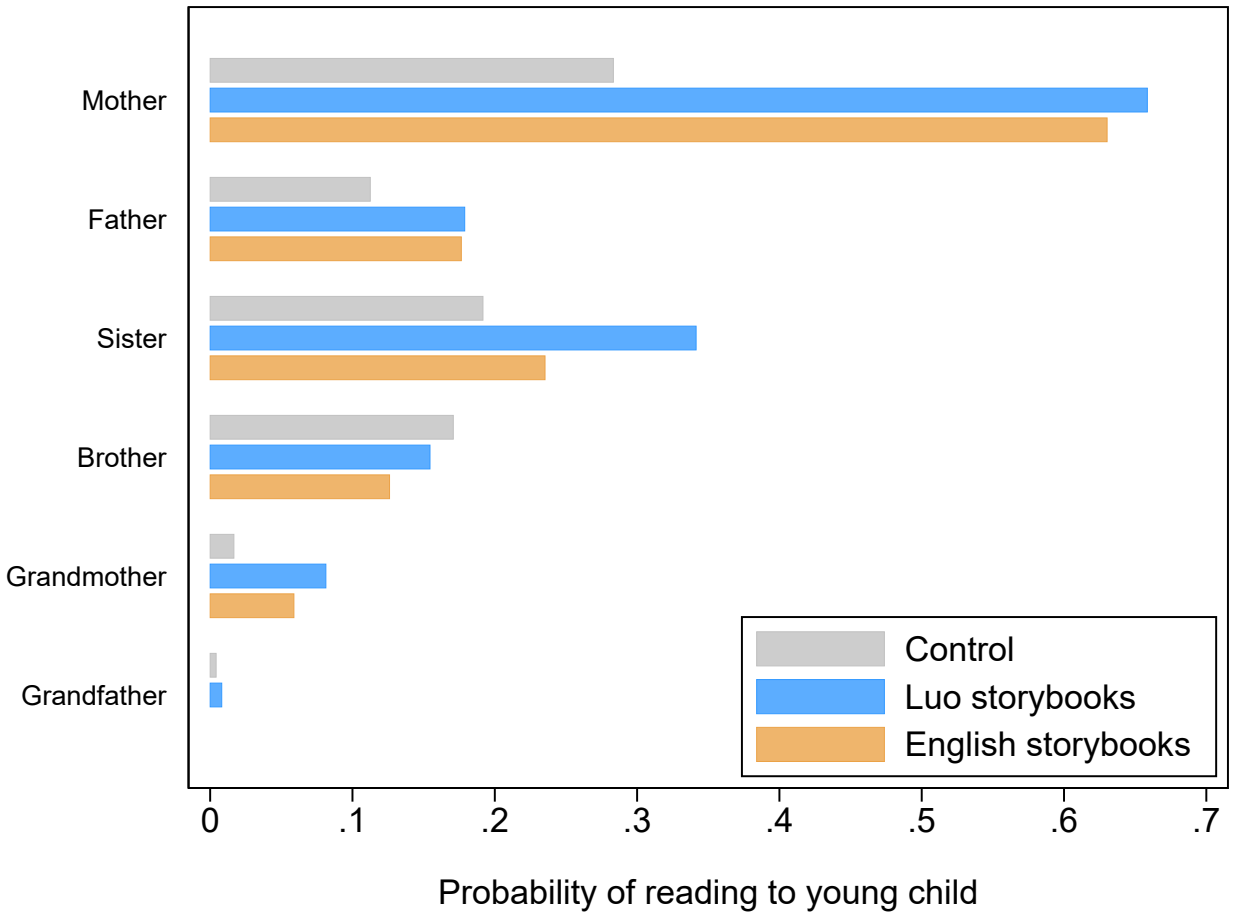
Short-Term Impacts

Result 4: children use the EMERGE books and learn the stories



Short-Term Impacts

Result 5: other women and girls also read the storybooks



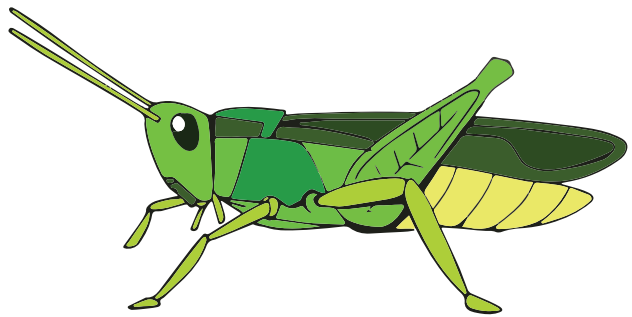
Measuring Impacts at Endline

Measuring Vocabulary

"Show me the DOG"

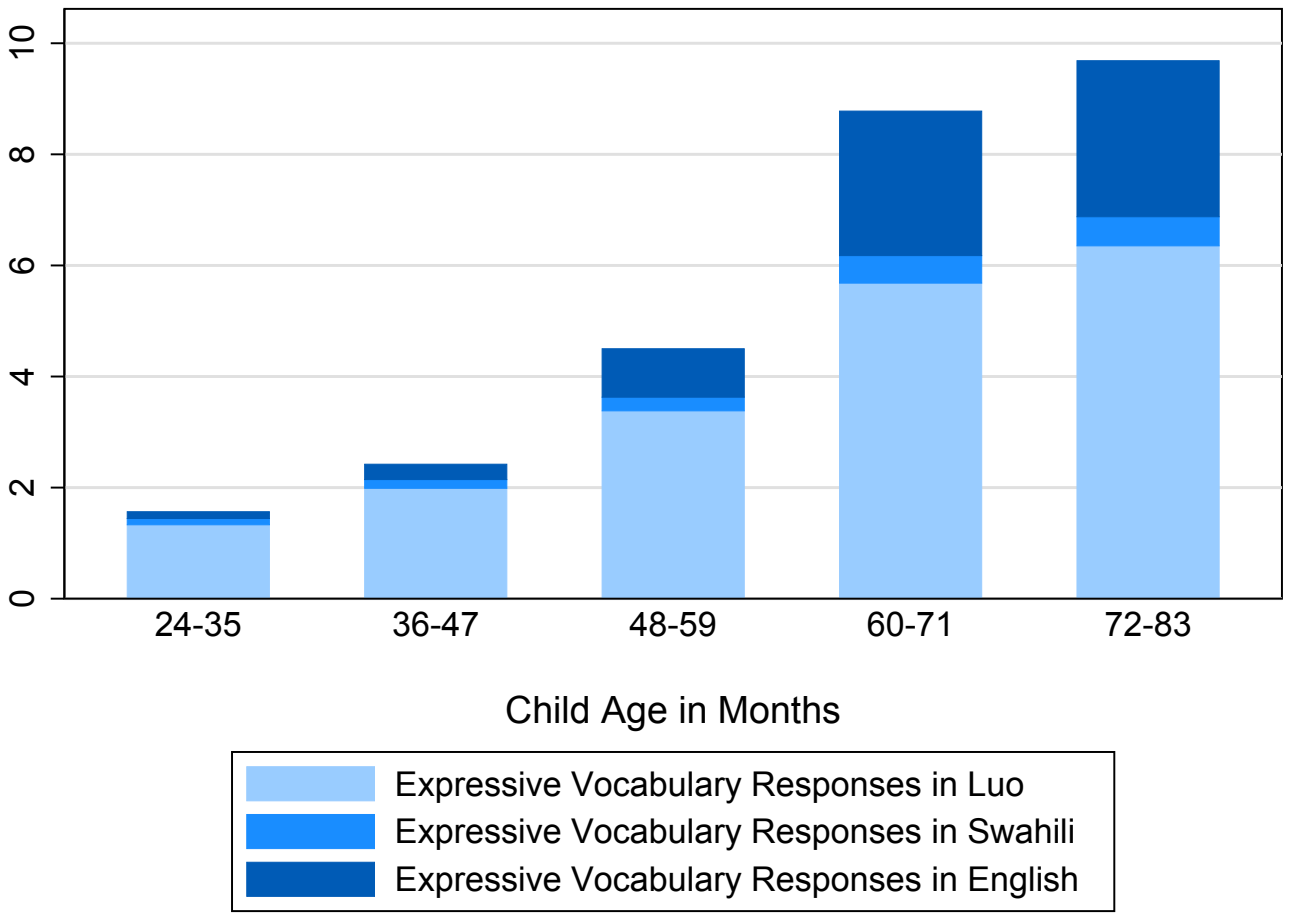
Receptive Vocabulary

"What is this?"



Expressive Vocabulary

Measuring Vocabulary



Source: Knauer et al. (2019)

Measuring Vocabulary

"banister" "tropical"

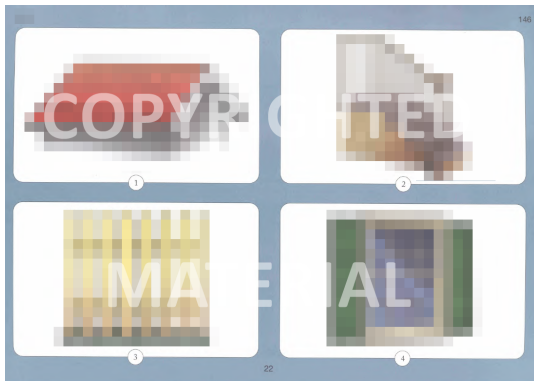
Source: British Picture Vocabulary Scale (Dunn, Dunn, and Styles 2009)

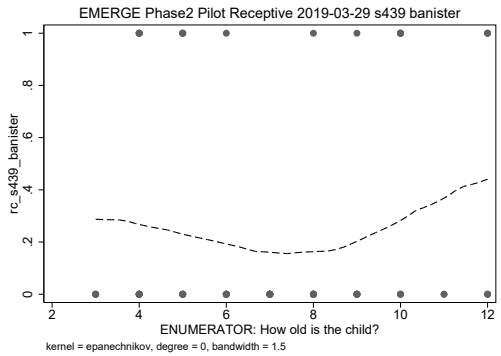
Inherent tension in measuring child development in LMIC settings

- Well-validated, widely used measures designed for HIC contexts
- May not capture child development in very different contexts

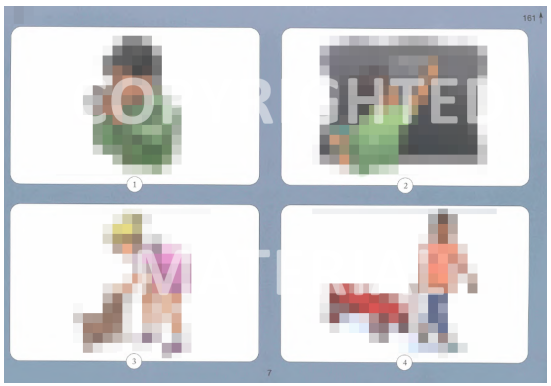
Measuring Vocabulary

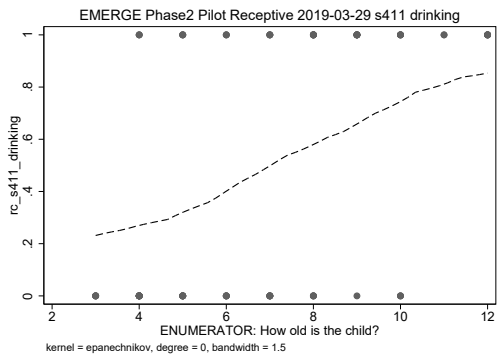
"banister"





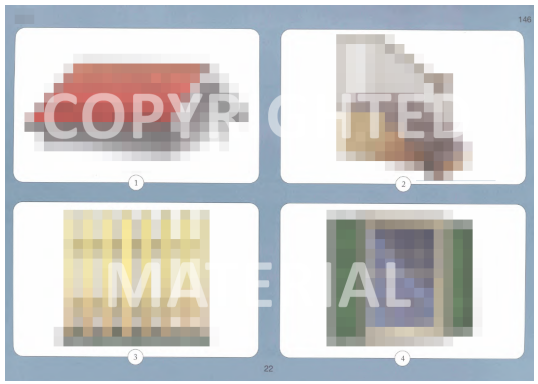
"drinking"

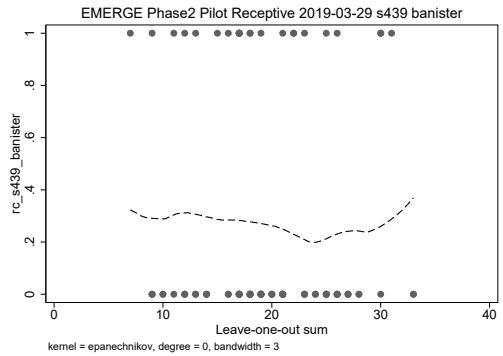




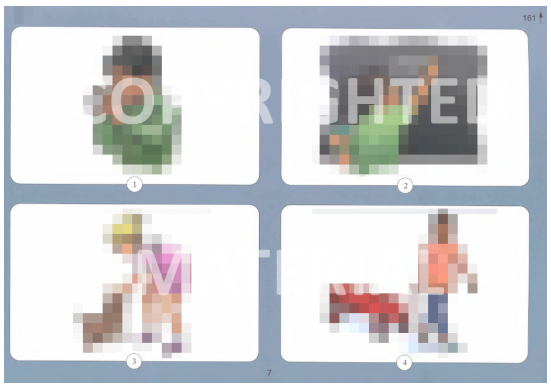
Measuring Vocabulary

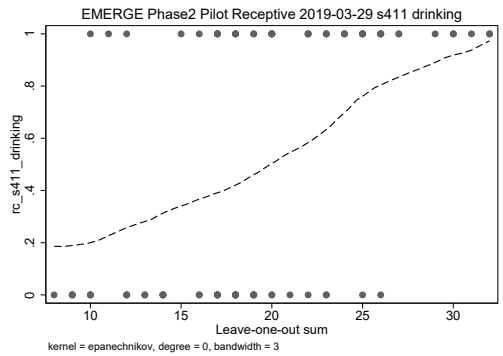
"banister"





"drinking"





Measuring Vocabulary

Objective measures of item performance:

- Correlation with child age
- Correlation with total score, leave-one-out sum
- IRT estimate of item discrimination
  - Discrimination may appear low for very easy/difficult items

Measuring Vocabulary

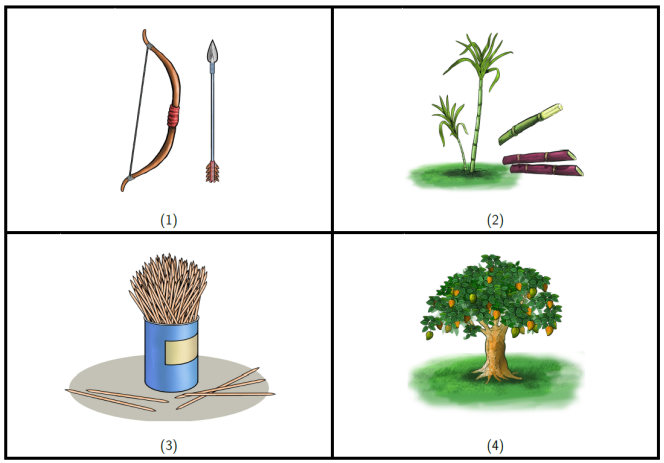
Objective measures of item performance:

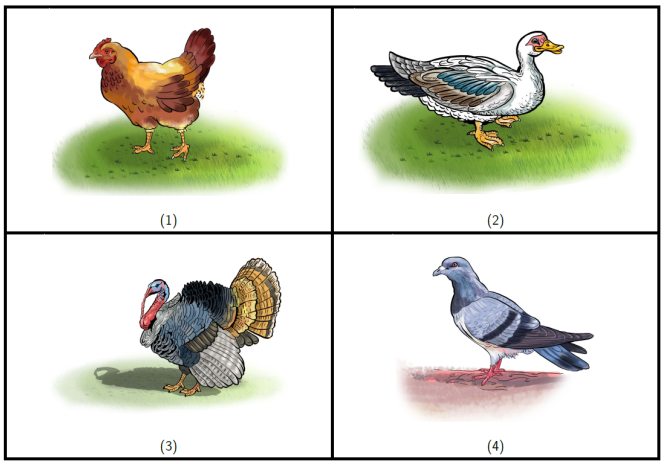
- Correlation with child age
- Correlation with total score, leave-one-out sum
- IRT estimate of item discrimination
  - Discrimination may appear low for very easy/difficult items

Why retain items that perform poorly?

- Not clear what it means to have comparable items when items do not have the same psychometric properties in different contexts
- We would not use English-language items in francophone countries
- Items that do not capture child development reduce statistical power

Measuring Vocabulary

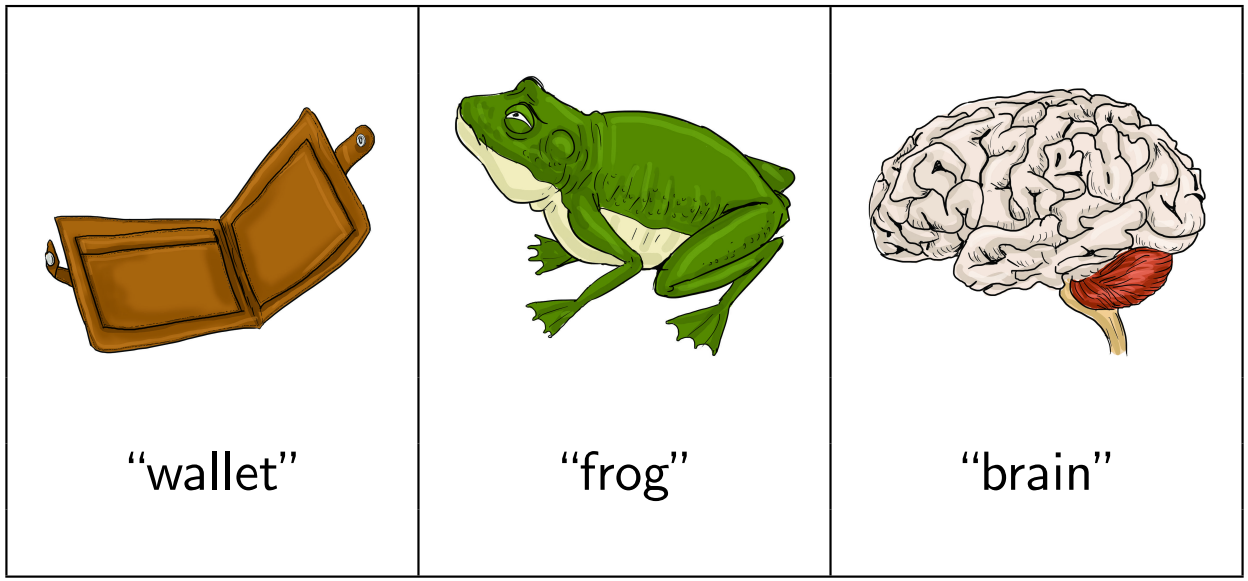




We adapt and extend the British Picture Vocabulary Scale (BVPS) to create English and Luo receptive vocabulary tests (Knauer et al. 2019)

- Adapt (seemingly) locally-appropriate items from BPVS
- Work with Kenyan artists to develop new, similarly-structured items
- Pilot all items with Kenyan schoolchildren, retaining those with reasonable psychometric properties while ensuring range of difficulty

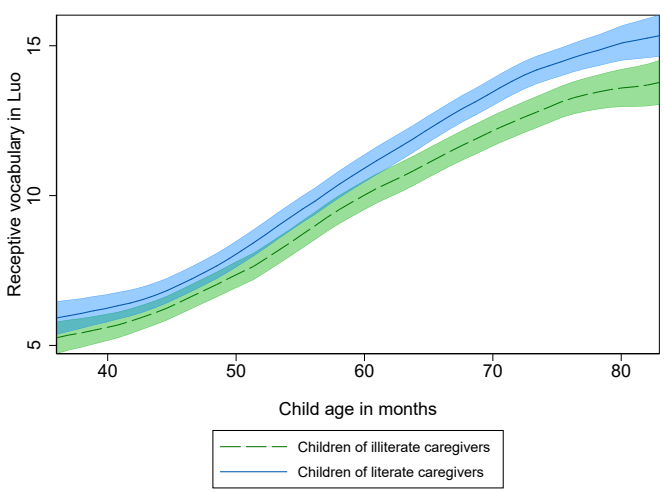
Measuring Vocabulary

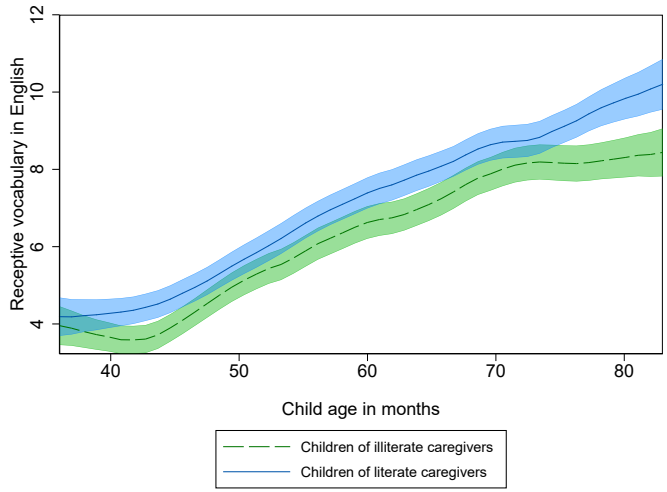


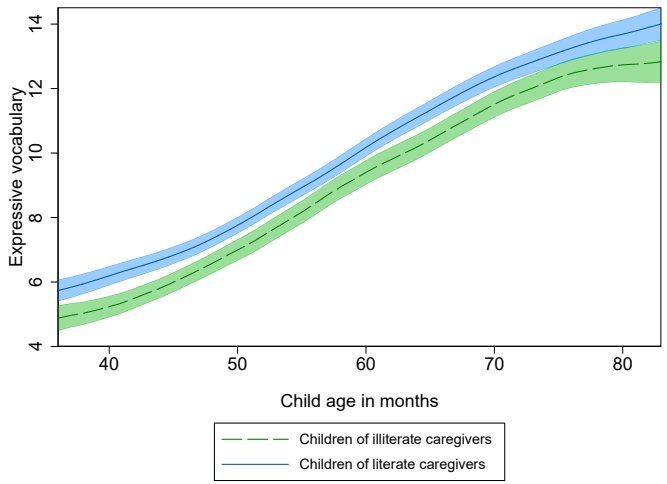
We develop a new, locally-appropriate expressive vocabulary assessment

- Existing tests relied on many inappropriate stimuli
- Use pilot data to characterize correct and common incorrect responses in all local languages to maximize inter-rater reliability

Measuring Vocabulary







Source: baseline data from EMERGE cluster-randomized evaluation

Measuring Other Outcomes

Child assessments:

- Early literacy assessment (EGRA) in both English and Luo
- Child activities/time use (through both parent and child reports)

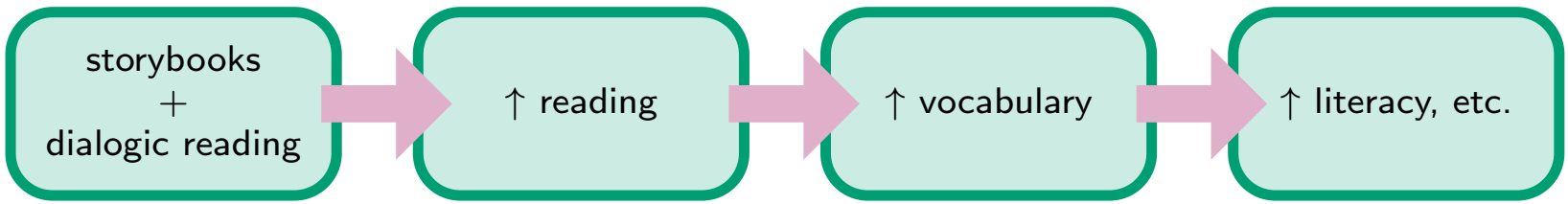
Primary caregiver survey:

- Shared reading and early childhood stimulation (expanded FCI)
- Women's time use (home production, childcare, leisure, etc.)
- Beliefs about children's vocabulary skills
- Demand for storybooks

Classroom observation, abbreviated teacher survey

Pre-Analysis Plan

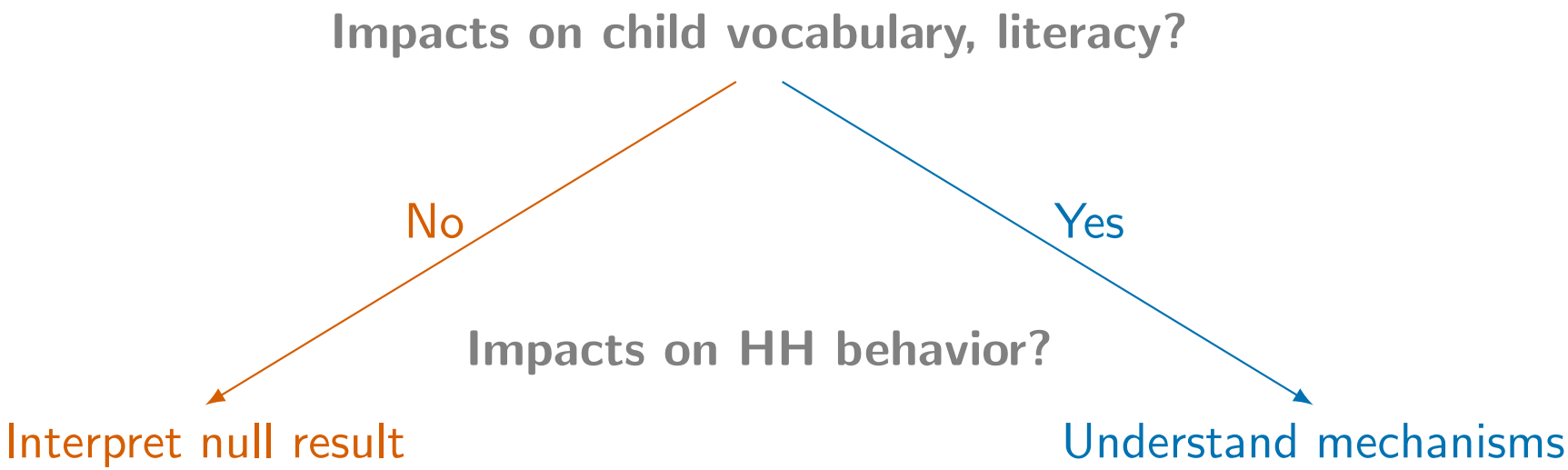
Primary and Secondary Hypotheses



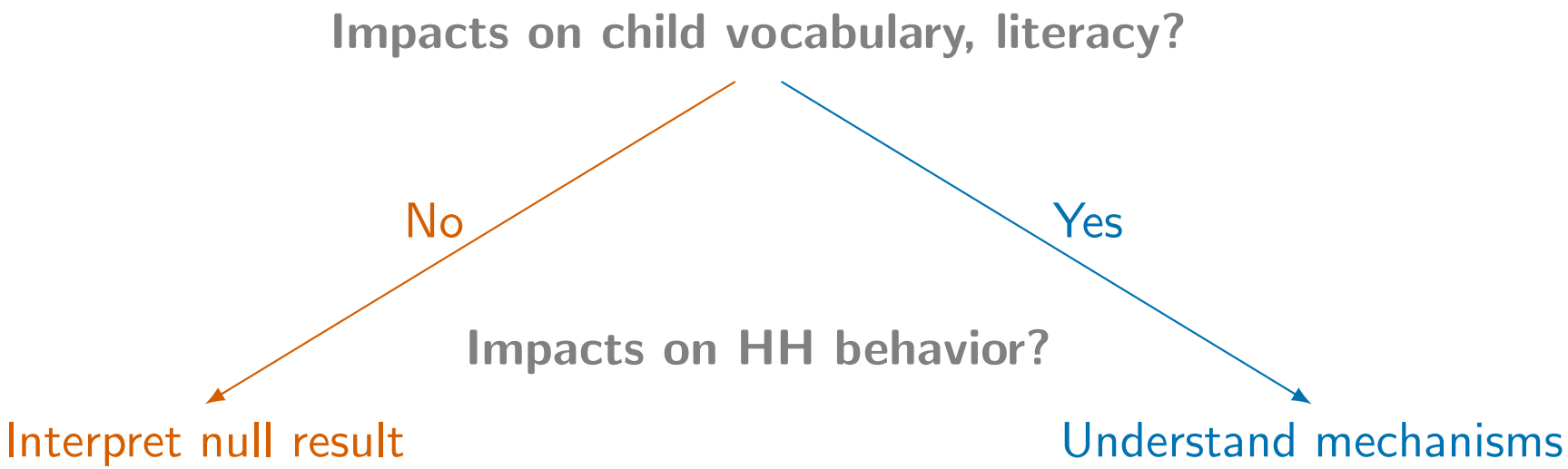
Primary hypothesis: EMERGE improved children's vocabulary, literacy

- Which types of vocabulary?
- Vocabulary, literacy in which language?
- Which children will be (measurably) impacted?
- Does the language of the storybook matter?

Primary and Secondary Hypotheses



Primary and Secondary Hypotheses



Secondary research questions:

- Are treated HHs reading more one year after treatment?
- Do children, siblings, parents know the stories?
- Do HHs engage in more child stimulation overall?
- Who does the stimulation, and what do they substitute away from?

Measuring Impacts on Vocabulary

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \sigma \cdot \sqrt{1 + (n_{groupsize} - 1)\rho}$$

Usually ≈ 2.8

Fixed given research design

Measuring Impacts on Vocabulary

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \sigma \cdot \sqrt{1 + (n_{groupsize} - 1)\rho}$$

Usually ≈ 2.8

Fixed given research design

Assumption (baseline data)

Measuring Impacts on Vocabulary

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \sigma \cdot \sqrt{1 + (n_{groupsize} - 1)\rho}$$

Usually ≈ 2.8 | Fixed given research design | Assumption (baseline data)

English receptive vocabulary:

Data from EMERGE pilot suggest $\rho \approx 0.026$

$MDE \approx 0.153\sigma$

Measuring Impacts on Vocabulary

Outcome	Coef.	S.E.	P-Val.	95% CI
Storybook Expressive	0.240	0.113	0.036	[0.016, 0.463]
Non-Storybook Expressive	0.041	0.107	0.703	[-0.171, 0.253]
Luo Receptive Vocabulary	0.022	0.139	0.869	[-0.242, 0.286]
English Receptive Vocabulary	0.147	0.135	0.276	[-0.119, 0.413]
Vocabulary Index	0.113	0.067	0.096	[-0.020, 0.246]

Measuring Impacts on Vocabulary

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \tilde{\sigma} \cdot \sqrt{1 + (n_{groupsize} - 1)\tilde{\rho}}$$

Use a one-sided test
(Anderson and Magruder 2017)

Measuring Impacts on Vocabulary

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \tilde{\sigma} \cdot \sqrt{1 + (n_{groupsize} - 1)\tilde{\rho}}$$

Use a one-sided test

Control for baseline Y, use residuals (calculated in pilot data)

Measuring Impacts on Vocabulary

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \tilde{\sigma} \cdot \sqrt{1 + (n_{groupsize} - 1)\tilde{\rho}}$$

Use a one-sided test

Control for baseline Y, use residuals (calculated in pilot data)

English receptive vocabulary:

Data from EMERGE pilot suggest $\tilde{\sigma} \approx 0.888, \tilde{\rho} \approx 0.005$

$MDE \approx 0.095\sigma$

Measuring Impacts on Vocabulary

	Unadjusted		Covariate-Adjusted		
Outcome	ρ	**MDE**	$\tilde{\sigma}$	$\tilde{\rho}$	**MDE**
Storybook Expressive	0.076	0.187	0.730	0.026	0.105
Non-Storybook Expressive	0.050	0.163	0.656	0.010	0.075
Luo Receptive	0.016	0.124	0.774	0	0.077
English Receptive	0.026	0.136	0.888	0.005	0.095
Vocabulary Index	0.060	0.172	0.606	0.015	0.074

Measuring Impacts on Vocabulary

We are well-powered to detect small effects, estimate precise nulls

- Baseline data available for all vocabulary outcomes
  - Child assessments are highly correlated over time
  - Measures were age-appropriate at baseline and endline
  - Intra-class correlation lower for residualized outcomes
- EMERGE pilot data, existing literature suggest positive impacts
  - Pre-analysis plan allows credible commitment to one-sided test
- Four (vocabulary) outcomes of interest
  - Pilot suggests largest impacts on English receptive vocabulary, existing literature suggests biggest impacts on expressive vocabulary
  - Focus on aggregate vocabulary index, adjust individual vocabulary outcomes for multiple hypothesis testing following Anderson (2008)

Measuring Impacts on Early Literacy

Baseline data on early literacy extremely limited

- EGRA not appropriate for preschool-aged children
- Predictive power of baseline covariates (for EGRA) unknown
  - Knowledge of letters, familiar word reading, vocabulary

Measuring Impacts on Early Literacy

Baseline data on early literacy extremely limited

- EGRA not appropriate for preschool-aged children
- Predictive power of baseline covariates (for EGRA) unknown
  - Knowledge of letters, familiar word reading, vocabulary

EGRA is appropriate for older children; we are measuring 2,043 of them

- Does increasing sample size by adding older siblings for whom baseline data is unavailable increase our statistical power?

Measuring Impacts on Early Literacy

Calculate MDE for baseline only sample, baseline+siblings sample

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \sigma \cdot \sqrt{1 + (n_{groupsize} - 1)\rho}$$

Measuring Impacts on Early Literacy

Calculate MDE for baseline only sample, baseline+siblings sample

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \sigma \cdot \sqrt{1 + (n_{groupsize} - 1)\rho}$$

Commit (through PAP) to using estimation strategy with smaller MDE

Sample	N	$\tilde{\sigma}$	$\tilde{\rho}$	**MDE**
Baseline sample	2,527	1	?	?
Baseline sample plus older siblings	4,570	?	?	?

Measuring Impacts on Early Literacy

Calculate MDE for baseline only sample, baseline+siblings sample

$$\text{MDE} = (t_{1-\kappa} + t_{\alpha/2}) \cdot \sqrt{\frac{1}{P(1-P)}} \cdot \sqrt{\frac{1}{N}} \cdot \sigma \cdot \sqrt{1 + (n_{groupsize} - 1)\rho}$$

Commit (through PAP) to using estimation strategy with smaller MDE

Sample	N	$\tilde{\sigma}$	$\tilde{\rho}$	**MDE**
Baseline sample	2,527	1	?	?
Baseline sample plus older siblings	4,570	?	?	?

Use LASSO to identify relevant baseline covariates from restricted subset

- Use different covariates to predict different outcomes w/o p-hacking

Secondary Hypotheses

- Differential impact by storybook language
  - Does language impact proficiency in English vs. Luo?
- Impacts on shared reading behaviors, stimulating activities
  - Number of children's storybooks in the home
  - Frequency of shared reading
  - Do young children, older siblings, adult caregivers know stories?
  - Overall level of early childhood stimulation
  - Demand for children's storybooks, beliefs about child ability
- Impacts on literacy, time use by mothers, older sibilings
- Heterogeneity by child age, caregiver literacy, other covariates (ML)

Next Steps

Next Steps

EMERGE team will deliver intervention to all communities in early 2020

- Building a model for scale-up, if intervention proves effective
- Evaluate messaging designed to encourage fathers to attend training
- Show appreciation to communities involved in our research

Thank you for listening!