DR. HEATHER A KNAUER (Orcid ID : 0000-0001-8617-8958)

Article type : Special Issue Article

Multilingual assessment of early child development: Analyses from repeated observations of children in Kenya

Heather A. Knauer[1*], PhD, Patricia Kariger[1], PhD, Pamela Jakiela[2], PhD, Owen Ozier[3], PhD, Lia C. H. Fernald[1], PhD

[1]University of California, Berkeley School of Public Health

[2]University of Maryland Department of Agriculture and Resource Economics

[3]The World Bank Development Research Group, Human Development Team

*Current institution: University of Michigan, School of Social Work

Corresponding author: Heather A. Knauer, email address: hknauer@umich.edu

Physical address: 1080 South University Ave, Rm 4864, Ann Arbor, MI 48109

Conflict of interest: The authors have no conflicts of interest to declare.

Acknowledgments: We thank Sheyda Esnaashari, Saahil Karpe, Rohit Chhabra, and Emily Cook-Lundgren for research assistance throughout this project. We thank the staff of Innovations for Poverty Action (Kenya), specifically, Patricia Gitonga, Michael Meda, Jessica Jomo, and the field team they led during data collection for this project. We are especially grateful to the KLPS team, who generously shared their previously adapted instruments with us. This work was supported by The World Bank, Washington, DC (via three facilities: the Strategic Impact

Evaluation Fund, the Early Learning Partnership, and the Research Support Budget). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily reflect the views of The World Bank, its board of executive directors, or the governments they represent.

Research highlights

- This study measured vocabulary among Kenyan children 2–6 years old, at two time points, across three languages: Luo (mother tongue), Swahili, and English (official languages).
- During testing, the youngest children strongly preferred to express themselves in Luo, whereas older children were more likely to respond in Luo and English.
- Luo receptive vocabulary among all children at baseline was significantly associated with English receptive vocabulary at follow-up, even accounting for baseline English and Swahili.
- Baseline caregiver literacy in Luo, rather than English, was robustly related to children's later receptive vocabulary in English.

Abstract

In many low- and middle-income countries, young children learn a mother tongue or indigenous language at home before entering the formal education system where they will need to understand and speak a country's official language(s). Thus, assessments of children before school age, conducted in a nation's official language, may not fully reflect a child's development, underscoring the importance of test translation and adaptation. To examine differences in vocabulary development by language of assessment, we adapted and validated instruments to measure developmental outcomes, including expressive and receptive vocabulary. We assessed 505 2-to-6-year-old children in rural communities in Western Kenya with comparable vocabulary tests in three languages: Luo (the local language or mother tongue), Swahili, and English (official languages) at two time points, 5–6 weeks apart, between September 2015 and October 2016. Younger children responded to the expressive vocabulary

measure exclusively in Luo (44–59% of 2-to-4-year-olds) much more frequently than did older children (20–21% of 5-to-6-year-olds). Baseline receptive vocabulary scores in Luo (β = 0.26, SE = 0.05, $p < 0.001$) and Swahili (β = 0.10, SE = 0.05, $p = 0.032$) were strongly associated with receptive vocabulary in English at follow-up, even after controlling for English vocabulary at baseline. Parental Luo literacy at baseline (β = 0.11, SE = 0.05, $p = 0.045$) was associated with child English vocabulary at follow-up, while parental English literacy at baseline was not. Our findings suggest that multilingual testing is essential to understanding the developmental environment and cognitive growth of multilingual children.

Keywords: school readiness, multilingual environments, language of instruction, BPVS, MDAT, PPVT

Up to 40% of children worldwide speak a mother tongue that is not used at school (Walter & Benson, 2012; Clegg & Simpson, 2016), which results in significant, negative consequences for hundreds of millions of children (Ball, 2011). For example, in a global study of fourth-grade students, children whose home language differed from the testing language were 10% less likely to achieve the most basic level of reading proficiency compared to students who spoke the testing language at home (UNESCO, 2018). Though mother tongue instruction is potentially complicated to implement in linguistically diverse environments, it may allow children to learn more and may better permit their parents to engage with teaching materials and monitor student performance (Konsonen, 2005; Benson, 2002; Lieberman, Posner, & Tsai, 2014). In the case of Kenya, 45% of mothers of school-aged children cannot read English at a second-grade reading level (Uwezo, 2015); in one study, 72% of parents reported not understanding how to interpret student-learning data (Lieberman et al., 2014). Thus, a country's policy regarding language of instruction (LOI) can have significant implications for children's development in ways that interact with poverty, parental literacy, ethnicity, and other risk factors faced by vulnerable children as they move through the formal education system.

Child Assessment in Multilingual Environments

Child development assessments allow teachers to understand how and what children are learning, to diagnose learning differences or language disorders, and to benchmark achievement

against national or international standards (Snilsveit et al., 2016; Armon-Lotem, de Jong, & Meir, 2015). Similarly, researchers and policymakers rely on child assessments to examine program effectiveness. In both academic and nonacademic settings, students are routinely tested in only one language, either the LOI or parents' preferred language. It is challenging to assess child development, language disorders, and school readiness in such populations, both because these children develop linguistic skills in multiple languages simultaneously and because most widely used measures of child development have not been validated in local languages and low- and middle-income country (LMIC) contexts. Many assessments created and validated in U.S. or European samples do not demonstrate the same strong psychometric characteristics when applied in different settings (Fernald, Prado, Kariger, & Raikes, 2017). To capture the linguistic development of children in LMIC contexts, it is crucial to adapt, or develop, and subsequently validate assessments in children's mother tongues (Prado et al., 2018).

Child development assessments conducted in a single language may not fully reflect a multilingual child's developmental outcomes and learning trajectory (Cummins, 1979; Cummins, 2001; Peña, Bedore, & Kester, 2015). Bilingual children's conceptual vocabularies are similar in size to those of monolingual children; however, their vocabulary size in each language is smaller than that for monolingual children (Hammer et al., 2014; Bialystok et al., 2010). The amount of overlap in children's vocabulary between the two languages may depend on how typologically related the two languages are (Hammer et al., 2014). Furthermore, bilingual children's performance on language assessments in their second language may have more to do with exposure to the second language than knowledge transfer based on first-language proficiency (Keller, Troesch, & Grob, 2015). For this reason, children may perform better on certain aspects of the tests, such as letter sounds, syllables, or reading fluency, when they are tested in the LOI as compared to their native language (Bialystok, Majumder, & Martin, 2003). Greater reading fluency or decoding skills in the LOI, however, do not necessarily indicate that children have greater reading comprehension in the LOI (Piper, Schroeder, & Trudell, 2016; Piper, Zuilkowski, & Ong'ele, 2016). For multilingual children, assessment of language and other domains of development should account for all of the child's languages (Pearson, Fernandez, & Oller, 1993). Furthermore, the assessment should ideally capture the complexity of the child's language environment or the extent to which a child's language is specific to a certain context (i.e., school, home, or community) (Toppelberg & Collins, 2010;

Pearson et al., 1993). To date, most studies of bilingual or multilingual child language development have been conducted in high-income countries (Barac, Bialystok, Castro, & Sanchez, 2014), although a few studies have been conducted in Sub-Saharan Africa (e.g., Cockcroft, 2016; Demuth, 2003; Alcock & Alibhai, 2013; Alcock, Holding, Mung'ala-Odera, & Newton, 2008; Alcock, 2017; Potgieter & Southwood, 2016). Thus, limited data are available to help us understand young children's verbal development in LMIC contexts.

Current Approaches to Child Assessment Across Contexts

There is an inherent tension between the desire to employ widely used, well-validated measures and the need to adapt items to local contexts. Assessments that are well validated in one context but not appropriately adapted for another may not maintain their properties (Peña, 2007) and may perform unreliably (Gibson, Jamulowicz, & Oller, 2017). This problem is particularly pronounced for tests designed and validated in high-income countries that, without thorough and careful adaptation, often generate items poorly suited to a LMIC context (Fernald et al., 2017; van de Vijver & Poortinga, 2005). Investigators generally have four approaches when using a measure in a new country or context: adoption (translation of an existing test without modification); adaptation (translation with careful modification of items, responses, and administration); expansion (adding items to an existing test to suit a particular cultural or linguistic context); or creation of new tests **(Figure 1**). These approaches have been used in the LMIC context (Weber, Fernald, Galasso, & Ratsifandrihamanana, 2015; He & van de Vijver, 2012) and in higher income contexts, where the parallel design of assessments is necessary to simultaneously test children's verbal development across multiple languages (Haman, Łuniewska, & Pomiechowska, 2015). When multiple tests are needed to comprehensively measure various capacities, a more diversified strategy may be to adopt some tests, adapt others, expand an existing test to include new test items, and create new tests that are internally valid for the context.

Language Policy in the African Context

Over 2,149 mother tongue languages are spoken in Africa (Lewis, Simons, & Fenning, 2016), and more than a quarter of the African population speaks a native language that is not in official use in the educational system or by the government (**Figure 2**; Lewis et al., 2016). In

spite of UNESCO's recent call for at least 6 years of mother tongue education (UNESCO, 2017), there are several reasons for resistance to mother tongue instruction. For example, parents and teachers sometimes believe that children who learn in the mother tongue language will fall behind those who learn in English (Jones, 2012; Trudell, 2007). In addition, linguistically appropriate teaching materials are not always available (Musau, 2003; Waithaka, 2017), and teachers may not be fluent in the local mother tongue (Trudell & Piper, 2014; Manyonyi et al., 2016). The misalignment between children's first languages and those used in schools has important implications for the assessment of school readiness and learning outcomes: namely, children from linguistically marginalized families risk being underserved by the educational system.

Current Study

Our study took place in the Luo-speaking region of western Kenya, a country with 68 spoken languages (Lewis et al., 2016). English and Swahili are the official languages (i.e., for all government proceedings and publications), but literacy rates in these languages, while perhaps relatively high within Sub-Saharan Africa, are still quite low. For example, only 55% of mothers of school-aged children, and about 51% of children aged 7–13 years can read English at a second-grade level (Uwezo, 2015). In our study area, only 31% of young primary school students are taught in Luo, while the rest are taught in either English or Swahili (Piper & Miksic, 2011).

The purpose of this study was to compile a set of child development assessments to evaluate the effects of a literacy promotion program on multilingual children's development. Our first aim was to validate language assessments for children aged 2–6 years. Our second aim was to understand children's performance on receptive vocabulary assessments in mother tongue (Luo) and official languages (English and Swahili), and the extent to which scores on each of these assessments were associated with children's receptive and expressive vocabulary at a 5- to 6-week follow-up. We hypothesized that baseline scores in all languages, but especially English, would be significantly associated with child English receptive vocabulary at follow-up, as they all measure aspects of language skill. Our final aim was to examine the relationship between caregiver literacy—in both the mother tongue (Luo) and the LOI (English)—and child receptive and expressive vocabulary and to test whether the strength of the association between mother

tongue and LOI literacy varied with caregiver literacy. We focused primarily on children's English vocabulary at follow-up as an indicator for school readiness, as this is the LOI at higher grade levels and the de facto language of instruction for many young children in our study area. We hypothesized that caregiver literacy in both languages at baseline would be significantly associated with child receptive and expressive language at follow-up, and that the association between the baseline measure of child receptive vocabulary in Luo and English vocabulary at follow-up would be strongest among children whose parents had lower English literacy.

Methods

Study design and sample description

The measures described in this paper were developed for an ongoing cluster-randomized trial in Kenya's Kisumu and Homa Bay Counties that is designed to evaluate the effectiveness of a book distribution and parenting training program on child development (see trial registry: https://doi.org/10.1186/ISRCTN68855267 and pilot results: Knauer, Jakiela, Ozier, Aboud, & Fernald, in press). Families with at least one child between the ages of 24 and 83 months were recruited from a set of nine primary school catchment areas in rural communities within two hours' drive from Kisumu. A total of 357 primary caregivers (one per household) and 510 children were assessed during household visits (average 1.43 children per household); five child assessments were incomplete, resulting in an analysis sample size of 505. A total of 442 children were assessed at follow-up (5–6 weeks later), with 68 children lost due to relocation or difficulty in making contact.

Measures

Overview of child assessments.

To develop our test battery, we used adoption, adaptation, expansion, and creation of new tests for different developmental domains. All assessments were translated to Swahili or Luo and then back-translated to English by a different team of translators (two for each language) who did not have access to the original measure. The first and second authors (HK and PK) then met with a group of translators and discussed each translation to ensure that words conveying the desired meaning were chosen over direct translation (In Swahili and Luo, several words were often possible depending on the intent of the item). The assessments were then pretested, and any additional study team concerns or discrepancies were addressed. Items for the vocabulary

assessments were ordered by difficulty, as measured in a small pilot sample (between 30 and 61 respondents).

The assessors hired to administer the tests in the current study had university degrees, were from the study area, spoke Luo as their mother tongue, and were trained on the full battery of tests by the first and second study authors. On a subset of 48 children, two assessors double coded the baseline assessment to assess interrater reliability (IRR) for each of the assessments (**Supplemental Table 1**).

Receptive vocabulary.

We created receptive vocabulary assessments based on the British Picture Vocabulary Scale III (BPVS III) (Dunn, Dunn, & Styles, 2009), which includes 168 items for individuals 3–16 years old (see details of translation and adaptation in **Appendix A**). Knowledge of words is measured by asking the respondent to point to one of four pictures that corresponds to a word (object, person, or action) spoken by the assessor. The BPVS has been adapted for use in South Africa (Cockcroft, 2016) and Indonesia (Prado, Alcock, Muadz, Ullman, & Shankar, 2012) and is the British adaptation of the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), which has also been used in neighboring areas of Kenya (Ozier, 2018). As we wanted to capture young children's knowledge of Luo, Swahili, and English words, we created three sets of nonoverlapping words of varying difficulty, with 27 Luo items, 32 Swahili items, and 34 English items. Administration ended when a child failed six out of a set of eight items.

Expressive vocabulary.

We developed our own measure of expressive vocabulary after reviewing various expressive vocabulary tests and concluding that the stimulus words and/or pictures were not appropriate to the context (see details in **Appendix B)**. The assessment was a picture-naming task, in which children were presented with flash cards bearing a single illustrated stimulus item or object (noun) per card and were asked in the child's preferred language, "What is this?" for each item. Children were not instructed as to which language to respond in, but responses in any language were accepted. We did not provide further instruction because code-switching during conversation is common in this area, and very young children may not be aware which language they are actually speaking for a given word. Thus, a child could respond to each item in the 20-item test in English, Luo, or Swahili to score a pass for expressing the word verbally. Administration ended with three consecutive fails.

Other child-level assessments.

The Malawi Developmental Assessment Tool (MDAT) was created and validated for use in rural Malawi with children 1–84 months of age (Gladstone et al., 2010). It includes four 34-item subscales (fine motor/perception, language/hearing, gross motor, social-personal), with many items adapted from existing Western tests (see details of our adaptation in **Appendix C**). The MDAT is currently being used in various countries, including Mali, Sierra Leone, Rwanda, Burkina Faso, and Zimbabwe (M. Gladstone, personal communication, June 24, 2016). The western Kenya adaptation was initiated by the first, second, and fifth authors (HK, PK, and LCHF) for the Kenya Life Panel Survey (e.g., Baird, Hicks, Kremer, & Miguel, 2016), a longitudinal study that examines the intergenerational effects of health investments. We used the translations and piloting data from that study to further adapt and expand the language and fine motor/perception subtests of the MDAT for current study. The final adapted language test had 26 items. To further reduce the overall length of the test, we created start and stop rules for three different age groups (24–35 months; 36–59 months; 60–71 months) based on pass rates during piloting.

Caregiver survey.

Data were gathered on household assets, housing quality, household size and composition, and the age and education level of primary caregivers. In addition, we assessed caregiver literacy by asking caregivers to read a simple, five-word (second-grade level) sentence in each language adapted from the Early Grade Reading Assessment (EGRA; Gove & Wetterberg, 2011). Caregivers who read more than one word incorrectly in all three languages were categorized as illiterate. Working memory in caregivers was assessed using a summary score of the forward and backward digit span test (Ozier, 2018; out of 20 possible), and mental health was measured using an adapted version of the Centers for Epidemiological Studies-Depression scale (CES-D; Radloff, 1977; scores range 0-60). Household support for learning was measured with a set of items drawn from the HOME Inventory, Family Care Indicators, and UNICEF MICS4 (Bradley, Corwyn, McAdoo, & Coll, 2001; Hamadani et al., 2010; Kariger et al., 2012).

Statistical Analysis

To address the first aim of validating our assessments by examining their psychometric properties, we measured: 1) the internal consistency of the measures using Cronbach's alpha; 2) IRR using Cohen's kappa, Krippendorff's alpha, and percent agreement; 3) construct validity by examining the correlations between the measures; and 4) convergent validity by examining associations with known covariates in bivariate regressions. For our second aim, to better understand the relationships between baseline measures of mother tongue and LOI receptive vocabulary and scores on subsequent vocabulary assessments, we estimated a series of ordinary least squares (OLS) regression models to examine the associations between baseline age-standardized receptive vocabulary scores in all languages (English, Swahili, and Luo) and English receptive vocabulary at follow-up. We repeated this analysis for follow-up measures of child expressive vocabulary as well as Swahili and Luo receptive vocabulary; we present these results as supplemental analyses. Our final aim was to examine the association between caregiver literacy and child vocabulary at two time points. We used OLS regression to examine the association between baseline caregiver literacy in Luo and English and child English and Luo receptive and expressive vocabulary scores at follow-up. To test whether the relationship between baseline mother tongue and LOI receptive vocabulary scores and follow-up LOI receptive vocabulary varied with caregiver literacy, we estimated OLS regression models that included both caregiver literacy and baseline child receptive vocabulary (in English, Luo, and Swahili). All regressions used age-adjusted *z*-scores for child vocabulary, and standard errors were adjusted for household clustering. All statistical analyses were conducted using Stata 14.2.

Results

Descriptive Statistics

The average age of children in the study was 54.42 months (range 24–83 months) (**Table 1**). About one quarter (27%) of caregivers were illiterate. Maternal and household characteristics were similar to those observed in the representative 2014 Kenya Demographic and Health Survey sample for the study area.

Psychometric Properties of the Instruments

The internal consistency of the vocabulary measures ranged from $\alpha = 0.57$ to 0.90: Cronbach's alphas were lowest for the expressive and English receptive tests and highest for the

MDAT language test (**Supplemental Table 1**). The internal consistency of the receptive vocabulary assessments was higher for Luo ($\alpha = 0.78$) and Swahili ($\alpha = 0.76$) than for English ($\alpha = 0.57$). The IRR of the receptive vocabulary tests was $\kappa = 1$ for Luo, $\kappa = 0.89$ for Swahili, and $\kappa = 0.95$ for English. The internal consistency of the expressive vocabulary test was $\alpha = 0.67$, while the IRR was $\kappa = 0.95$. The internal consistency of the MDAT fine motor and language tests was $\alpha = 0.94$ and $\alpha = 0.90$, respectively. IRR of the total score for each measure was $\kappa = 0.93$ for fine motor and $\kappa = 0.86$ for language.

Correlations among the baseline child development assessments ranged from $r = 0.32$ to 0.56, and all correlations were statistically significant at the $p < 0.001$ level (**Supplemental Table 2**). The three age-normalized receptive vocabulary scores were all moderately correlated with each other, the expressive vocabulary score, and the MDAT scores, while the expressive vocabulary score was also moderately correlated with both MDAT scores. The MDAT tests had the strongest correlation ($r = 0.56$) with each other; among the vocabulary assessments, they had the strongest correlations with Luo vocabulary ($r = 0.48$–0.49).

The associations between baseline child, caregiver, and household characteristics with child age-adjusted child development scores are presented in **Supplemental Table 3**. In bivariate regression analyses adjusted for household clustering, child height-for-age *z*-score was significantly associated with all child development assessments ($\beta = 0.25$–0.33, SD = 0.03–0.04, $p < 0.001$ for all). Caregiver characteristics (education, literacy, and cognition) were most strongly associated with child expressive vocabulary and MDAT scores, while caregiver depressive symptoms were not associated with any child assessments. Finally, household characteristics were not consistently associated with child assessments.

The Role of Language in Child Development Assessment

At baseline, 2-year-old children knew, on average, 3.45 of 27 (SD = 3.40) Luo receptive vocabulary words, 3.40 of 34 (SD = 3.41) English receptive vocabulary words, 5.08 of 31 (SD = 4.20) Swahili receptive vocabulary words, and 1.71 of 20 expressive vocabulary words in any language (SD = 2.02) (**Table 2**). Children's vocabulary progressed with age (**Figure 2**), such that 6-year-olds knew, on average, 14.48 (SD = 5.39) Luo, 9.17 (SD = 3.70) English, and 11.94 (SD = 5.43) Swahili receptive words, and 9.78 expressive words in any language (SD = 5.83).

Overall, 189 children answered the expressive vocabulary test entirely in Luo, while 13 children answered entirely in English, and 6 children answered entirely in Swahili (**Table 2**). The

other 297 children (58%) answered in more than one language; the number of children answering in only one language decreased with age. Across all ages, children answered more expressive vocabulary words in Luo, followed by English and then Swahili. The fraction of expressive responses given in Luo decreased from about 89% among 2-year-olds to about 67% among 6-year-olds, while the fraction of responses given in English increased from about 5% among 2-year-olds to about 29% among 6-year-olds (**Table 2**). The percentage of responses given in Swahili was small (7% among 2-year-olds) and decreased slightly with age. The youngest children showed a clear preference for expressing themselves in their mother tongue, as was evident in the patterns of response in our expressive vocabulary test (**Figure 3**).

Children's vocabulary at baseline and follow-up.

In bivariate analyses, the baseline measure for each language was most strongly associated with the corresponding follow-up measure (**Tables 3, S4, & S5**; Models 1–3). When baseline measures of the other languages were included in the analyses (Models 4–6), however, our follow-up receptive English measure performed differently from other languages in at least two ways: first, English benefited most from the inclusion of baseline measures of the other two languages; second, English was much more strongly associated with the next-best baseline assessment than were other follow-up languages (**Figure 4**). The baseline Luo assessment was also notable for being the next-strongest correlate of both follow-up English and follow-up Swahili.

Child baseline Luo receptive vocabulary was significantly associated with follow-up English receptive vocabulary ($\beta = 0.18$, SD = 0.05, $p < 0.001$), even after accounting for baseline English ($\beta = 0.26$, SD = 0.05, $p < 0.001$) and Swahili ($\beta = 0.10$, SD = 0.05, $p = 0.032$; Model 6) (**Table 3**). Adding Luo to the English test (moving from Model 1 to Model 4 or 6) increased the *R*-squared substantially (**Figure 5**). There was a greater gain in *R*-squared by testing in both English and Luo (Model 4: *R*-squared = 0.1737) than English alone, or in English and Swahili (Model 5: *R*-squared = 0.1560). Testing in all three languages yielded an *R*-squared of 0.1821 (Model 6). In contrast, baseline receptive vocabulary for all languages was not associated with children's Swahili or Luo vocabulary at follow-up (**Supplemental Tables 4 and 5**). Though the other languages were sometimes statistically significant in the full models (Model 6), they did little to increase the overall explanatory power above the bivariate model of only baseline receptive vocabulary scores in the same language (Model 1).

Children's expressive vocabulary at follow-up was significantly associated with baseline receptive vocabulary measures in all three languages; Luo receptive vocabulary had the strongest association ($\beta = 0.26$, SE = 0.06, $p < 0.001$; Model 8) (**Supplemental Table 6**). When baseline expressive vocabulary was included in the full model, however, the receptive vocabulary measures were no longer statistically significant (Model 9).

Associations Between Caregiver Literacy and Child Vocabulary

In our final analyses, we examined caregiver baseline and child follow-up measures in English (the primary LOI at older grade levels) and Luo (the mother tongue for 95% of our sample). After adjusting for caregiver education and household wealth, caregiver literacy in Luo was significantly associated with children's receptive vocabulary in English ($\beta = 0.11$, SD = 0.05, $p = 0.045$), while caregiver literacy in English was not (**Table 4**). Caregiver literacy in either language was not significantly associated with children's receptive vocabulary in Luo or their expressive vocabulary. Moreover, controlling for caregiver literacy (in English and Luo) did not alter the pattern of associations between children's baseline receptive vocabulary (in English, Luo, and Swahili) and follow-up English receptive vocabulary. Similar patterns were observed among the children of literate and illiterate caregivers, though baseline Swahili receptive vocabulary was more strongly associated with endline English receptive vocabulary among the children of literate caregivers (**Supplemental Table 7**).

Discussion

In this study, we assessed the language development of 2-to-6-year-old multilingual children at two time points in a rural, ethnically homogenous region of Kenya. Notably, we found that English receptive vocabulary was less strongly associated with other measures of children's language development than expected, especially among the youngest children. Instead, baseline Luo receptive vocabulary seemed best able to capture general language skill. Specifically, children's baseline Luo receptive vocabulary was significantly associated with English receptive vocabulary at follow-up, even after taking baseline English receptive vocabulary into account. Luo was also the form of receptive vocabulary most strongly associated with subsequent expressive vocabulary.

Children's follow-up English receptive scores were significantly associated with their baseline receptive vocabulary scores in all three languages. However, English receptive

vocabulary at baseline was not a significant correlate of children's later receptive vocabulary in Swahili or Luo. These findings suggest that, when measuring children's vocabulary at a very young age, an assessment of mother tongue receptive vocabulary provides a strong indication of overall language ability, while LOI receptive vocabulary does not provide a full assessment of vocabulary development. Our findings are consistent with existing work in the study area, in which children demonstrated greater reading fluency in English than in Luo, but significantly lower reading comprehension scores in English than in Luo (Piper, Schroeder, & Trudell, 2016). While children's familiarity with English through their classroom exposure is high, their actual understanding of English is often quite low (Trudell & Piper, 2014). This situation is likely to be common to many African contexts since many children learn to read in a language other than their mother tongue (Lewis et al., 2016).

In the process of vocabulary development, children typically first acquire receptive knowledge of a word (i.e., they recognize and understand the word when it is spoken or read), only later developing the ability to produce the word (expressive vocabulary) either by speaking or writing (Burger & Chong, 2011). By age six, children's receptive vocabulary is usually larger than their expressive vocabulary, although they may also learn to say words before they fully understand them (Burger & Chong, 2011). In our examination of the relationship between children's receptive and expressive vocabulary, we found that the strongest measures of language development at follow-up were baseline expressive vocabulary (in any language) followed by receptive vocabulary in Luo. However, expressive vocabulary is often not measured in research studies in LMICs—for example, because very young children can be too shy to respond, or may respond correctly in any one of several languages, which makes it more complicated to code responses. Our finding that 59% of children used multiple languages in their expressive responses confirmed our assumption that code-switching was common.

Caregivers' literacy in mother tongue at baseline provided an indicator of children's school readiness (as measured by English vocabulary), while caregivers' English literacy skills at baseline did not. Additionally, caregiver literacy at baseline in either language was not associated with children's vocabulary in Luo at follow-up. There may be several reasons for these findings. First, young children are often cared for by multiple caregivers, including other children. As a result, their Luo vocabulary may depend less on their primary caregiver's literacy because they hear conversation among other family and caregivers in Luo. Conversely, most families and

neighbors do not converse in English, so children would have less regular exposure to the language. Greater caregiver engagement in stimulating activities with their children was associated with higher English, but not Luo receptive vocabulary, suggesting that caregivers may deliberately teach their children English words. Finally, while our measurement of caregiver literacy was designed to be at a second-grade reading level, it proved more difficult for caregivers in English than in Luo; only three caregivers could read all five words in the English sentence.

A central limitation to this study is that it took place among a rural and ethnically homogenous group of children, so the findings may not generalize to an urban or ethnically mixed setting (Hungi, Njangi, Wekulo, & Ngware, 2017). In mixed ethnicity households or very diverse communities, the associations between mother tongue vocabulary and subsequent LOI vocabulary may not be as strong. However, even within this homogenous group, we had to navigate a multilingual environment to implement language assessments, which presented several inherent challenges. First, items (e.g., "playground") that perform well in high-income contexts may be unknown to children in other settings. Additionally, concepts that are represented by a single more difficult word in the original test language may translate to a phrase built from much simpler words: for example, "nest" translates in Luo to *od* ("covering" or "housing") *winyo* ("bird") (Capen, 1998), making it an easier word in Luo than the same word in English; thus, the ordering of item difficulty may no longer be appropriate. Finally, even linguistically accurate translations may not retain what some have called "psychological similarity" (van de Vijver & Poortinga, 2005). This is when an item taken from one setting may not have the same psychological meaning in a different context, such as "What do we do before crossing the road?" Therefore, a core strength of our study is the rigorous adaptation, translation, and validation process that we performed for our assessments and our testing of children across a broad age range in multiple languages. This process allowed us to document more fully how children's vocabulary in different languages evolves with age and how receptive vocabulary measures in mother tongue and the LOI were associated with vocabulary development 5–6 weeks later.

In a multilingual context, as is common in LMICs, there is a question of how best to support young children's language and cognitive development. Should pre-primary educational materials be in the local mother tongue—i.e., children's first language—or in English, the

language in which children will eventually be instructed and tested in primary school? Our findings raise the possibility that to best support the language development of children before school age, early childhood interventions—especially those targeting parents—might do well to include instruction and materials in mother tongue, as a child's first language lays the foundation for learning in other languages and for general readiness for school (see also Altan & Hoff, 2018; Hoff & Ribot, 2017).

A recent review of language of instruction policies in Eastern and Southern Africa found that 14 out of 21 countries introduce English as the LOI before fifth grade (Trudell, 2016a). However, it may be particularly challenging in Africa to implement UNESCO's guidelines of at least six years of mother tongue education because of the continent's high degree of linguistic heterogeneity. As a concrete example, Kenya's formal educational policy mandates that early primary instruction be conducted in the mother tongue in rural areas and in Swahili in urban areas—with a transition to English at Grade 4 in either case— however, this policy is only loosely followed in practice, illustrating the practical challenges inherent in such complex environments (Trudell, 2016b; Manyonyi, Mbori, & Okwako, 2016).

Vocabulary assessment of young children in only one language, particularly if not in their mother tongue, risks inadequately capturing children's development. Foundational work in the study of bilingual education has pointed out the interdependence of language skills across languages for bilingual children, but has focused exclusively on high-income country examples (e.g., Cummins, 1979). Monolinguals and bilinguals may learn school-centric words in the LOI equally quickly, but bilingual children may differentially know home-centric words in their first language rather than the LOI, thereby complicating the interpretation of assessments conducted in a single language (Bialystok, Luk, Peets, & Yang, 2010). As a specific example of the interplay of languages in Africa, Shin et al. (2015) found that in Malawi, Chichewa literacy in Grade 2 was a predictor of subsequent English skills in Grade 3.

A recent study in Kenya found no additional benefit from mother tongue instruction in primary school on children's language development, but only assessed children's linguistic development in English and Swahili (Piper, Zuilkowski, Kwayumba, & Oyanga, 2018). However, a separate study found that the PRIMR program (which provides teacher training and instructional supports to improve language and math skills in early primary grades) improved oral reading fluency and reading comprehension in mother tongue (Piper, Zuilkowski, &

Ong'ele, 2016). Our findings suggest that receptive vocabulary in a child's mother tongue may be a particularly important measure of linguistic development, even when the outcome of interest is the language of instruction. Children's vocabulary in their mother tongue may better reflect the level of stimulation and conversation they receive at home, while children's vocabulary in the LOI indicates their exposure to that language. Multilingual testing of parents and children is essential in order to understand the developmental status of multilingual children as well as factors that affect their development in LMICs.

Data availability:

The data that support the findings of this study are available upon reasonable request to oozier@worldbank.org.

References

Alcock, K. J. (2017). Production is only half the story—First words in two East African languages. *Frontiers in Psychology*, *8*, 1898.

Alcock, K. J., & Alibhai, N. (2013). Language development in sub-Saharan Africa. In M. J. Boivin & B. Giordani (Eds.), *Neuropsychology of Children in Africa* (pp. 155–180). Chapter 8] New York: Springer New York LLC.

Alcock, K. J., Holding, P. A., Mung'ala-Odera, V., & Newton, C. (2008). Constructing tests of cognitive abilities for schooled and unschooled children. *Journal of Cross-Cultural Psychology*, *39*(5), 529–551.

Altan, A., & Hoff, E. (2018). Complex structures in the child-directed speech of native and nonnative speakers. In D. Buğa & M. C. Ögeyik (Eds.), *Psycholinguistics and Cognition in Language Processing* (pp. 127–139). Hershey, PA: IGI Global.

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children. disentangling bilingualism from language impairment*. Bristol, England: Multilingual Matters.

Baird, S., Hicks, J. H., Kremer, M., & Miguel, E. (2016). Worms at work: Long-run impacts of a child health investment. *Quarterly Journal of Economics*, *131*, 1637–1680.

Ball, J. (2011). *Enhancing learning of children from diverse language backgrounds: Mother tongue based bilingual or multilingual education in the early years*. Paris, France: United Nations Educational, Scientific and Cultural Organization [UNESCO].

Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*(4), 699–714.

Benson, C. J. (2002). Real and potential benefits of bilingual programmes in developing countries. *International Journal of Bilingual Education and Bilingualism*, *5*(6), 303–317.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525–531.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, 24, 27–44.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Coll, G. C. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*(6), 1844–1867.

Burger, A., & Chong, I. (2011). Receptive vocabulary. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and development*. Boston, MA: Springer.

Capen, C. A. (1998). *Bilingual Dholuo–English dictionary.* Tuscon, AZ: Carole A. Capen.

Clegg, J., & Simpson, J. (2016). Improving the effectiveness of English as a medium of instruction in sub-Saharan Africa. *Comparative Education, 52*(3), 359–371.

Cockcroft, K. (2016). A comparison between verbal working memory and vocabulary in bilingual and monolingual South African school beginners: Implications for bilingual language assessment. *International Journal of Bilingual Education and Bilingualism*, *19*(1), 74–88.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222–251.

Cummins, J. (2001). Bilingual children's mother tongue: Why is it important for education? *Sprogforum*, *7*(19), 15–20.

Demuth, K. (2003). The acquisition of Bantu languages. In D. Nurse & G. Philippson (Eds.), *The Bantu languages* (pp. 209–222). London, England: Routledge.

Dunn, L. M., & Dunn, D. M. (1997). *PPVT-III: Peabody Picture Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *British Picture Vocabulary Scale*, 3rd ed. London, England: GL Assessment.

Fernald, L. C., Prado, E., Kariger, P., & Raikes, A. (2017). *A toolkit for measuring early childhood development in low and middle-income countries*. Washington, DC: The World Bank.

Gibson, T. A., Jarmulowicz, L., & Oller, D. K. (2017). Difficulties using standardized tests to identify the receptive expressive gap in bilingual children's vocabularies. *Bilingualism: Language and Cognition*, *21*, 1–12.

Gladstone, M., Lancaster, G. A., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N. R., & Smyth, R. L. (2010). The Malawi Developmental Assessment Tool (MDAT): The creation, validation, and reliability of a tool to assess child development in rural African settings. *PLoS Med, 7*(5).

Gove, A., & Wetterberg, A. (2011). *The early grade reading assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI International.

Hamadani, J. D., Tofail, F., Hilaly, A., Huda, S. N., Engle, P., & Grantham-McGregor, S. M. (2010). Use of family care indicators and their relationship with child development in Bangladesh. *Journal of Health, Population and Nutrition, 28*(1), 23–33.

Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–240). Bristol, England: Multilingual Matters.

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*(4), 715–733.

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, *2*(2), 1–19.

Hoff, E., & Ribot, K. M. (2017). Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 years. *Journal of Pediatrics*, *190*, 241–245.

Hungi, N., Njangi, J., Wekulo, P., & Ngware, M. (2017). Effects of language of instruction on learning of literacy skills among pre-primary school children from low-income urban communities in Kenya. *Early Childhood Education Journal, 46*(2), 187–199.

Jones, J. M. (2012). The effect of language attitudes on Kenyan stakeholder involvement in mother tongue policy implementation. *Journal of Multilingual and Multicultural Development*, *33*(3), 237–250.

Kariger, P., Frongillo, E. A., Engle, P., Britto, P. M., Sywulka, S. M., & Menon, P. (2012). Indicators of family care for development for use in multicountry surveys. *Journal of Health, Population and Nutrition, 30*(4), 472–486.

Keller, K., Troesch, L. M., & Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Frontiers in Psychology, 6*.

Knauer, H. A., Jakiela, P., Ozier, O., Aboud, F. E., & Fernald, L. C. (in press). Enhancing young children's language acquisition through parent–child book-sharing: A randomized trial in Rural Kenya. *Early Childhood Research Quarterly*. https://doi.org/10.1016/j.ecresq.2019.01.002

Konsonen, K. (2005). Education in local languages: Policy and practice in Southeast Asia. In *First languages first: Community-based literacy programmes for minority language contexts in Asia* (96–134). Bangkok, Thailand: UNESCO Asia and Pacific Regional Bureau for Education.

Lewis, M. P., Simons, G. F., & Fenning, C. D. (Eds.). (2016). *Ethnologue: Languages of the world,* 19th ed. Dallas, TX: SIL International.

Lieberman, E. S., Posner, D. N., & Tsai, L. L. (2014). Does information lead to more active citizenship? Evidence from an education intervention in rural Kenya. *World Development*, *60*, 69–83.

Manyonyi, J., Mbori, B., & Okwako, E. (2016). Attitude of teachers towards use of mother tongue as medium of instruction in lower primary schools in Bungoma South Sub-County, Kenya. *International Journal of Education and Research, 4*(8), 315–334.

Musau, P. M. (2003). Linguistic human rights in Africa: Challenges and prospects for indigenous languages in Kenya. *Language, Culture and Curriculum, 16*(2), 155–164.

Ozier, O. (2018). Exploiting externalities to estimate the long-term effects of early childhood deworming. *American Economic Journal: Applied Economics*, *10*(3), 235–262.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*(1), 93–120.

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*(4), 1255–1264.

Peña, E. D., Bedore, L. M., & Kester, E. S. (2015). Assessment of language impairment in bilingual children using semantic tasks: two languages classify better than one. *International Journal of Language and Communication Disorders*, *51*, 192–202.

Piper, B., & Miksic, E. (2011). Mother tongue and reading: Using early grade reading assessments to investigate language-of-instruction policy in East Africa. In A. Gove & A. Wetterberg (Eds.), *The early grade reading assessment: Applications and intervention to improve basic literacy* (pp.139–182). Research Triangle Park, NC: RTI Press. Retrieved from http://www.rti.org/pubs/bk–0007–1109-wetterberg.pdf.

Piper, B., Schroeder, L., & Trudell, B. (2016). Oral reading fluency and comprehension in Kenya: Reading acquisition in a multilingual environment. *Journal of Research in Reading*, *39*(2), 133–152.

Piper, B., Zuilkowski, S. S., Kwayumba, D., & Oyanga, A. (2018). Examining the secondary effects of mother-tongue literacy instruction in Kenya: Impacts on student learning in English, Kiswahili, and mathematics. *International Journal of Educational Development*, *59*, 110–127.

Piper, B., Zuilkowski, S. S., & Ong'ele, S. (2016). Implementing mother tongue instruction in the real world: Results form a medium-scale randomized controlled trial in Kenya. *Comparative Education Review, 60*(4), 776–807.

Potgieter, A. P., & Southwood, F. (2016). A comparison of proficiency levels in 4-year-old monolingual and trilingual speakers of Afrikaans, isiXhosa and South African English across SES boundaries, using LITMUS-CLT. *Clinical Linguistics and Phonetics*, *30*(2), 87–100.

Prado, E. L., Alcock, K. J., Muadz, H., Ullman, M. T., & Shankar, A. H. (2012). Maternal multiple micronutrient supplements and child cognition: A randomized trial in Indonesia. *Pediatrics, 131*, 536–546.

Prado, E. L., Phuka, J., Ocansey, E., Maleta, K., Ashorn, P., Ashorn, U., Adu-Afarwuah, S., Oaks, B. M., Lartey, A., & Dewey, K. G. (2018). A method to develop vocabulary checklists

in new languages and their validity to assess early language development. *Journal of Health, Population and Nutrition, 37*, 13.

Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Shin, J., Sailors, M., McClung, N., Pearson, P. D., Hoffman, J. V., & Chilimanjira, M. (2015). The case of Chichewa and English in Malawi: The impact of first language reading and writing on learning English as a second language. *Bilingual Research Journal, 38*(3), 255–274.

Snilsveit, B., Stevenson, J., Menon, R., Philips, D., Fallagher, E., Geleen, M., Jobse, H., Schmidt, T., & Jimenez, E. (2016). *The impact of education programmes on learning and school participation in low- and middle-income countries: A systematic review summary report*. London, England: International Initiative for Impact Evaluation.

Toppelberg, C. O., & Collins, B. A. (2010). Language, culture, and adaptation in immigrant children. *Child and Adolescent Psychiatric Clinics of North America*, *19(4)*, 697–717.

Trudell, B. (2007). Local community perspectives and language of education in sub-Saharan African communities. *International Journal of Educational Development*, *27*(5), 552–563.

Trudell, B. (2016a). *The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa*. Nairobi, Kenya: UNICEF Eastern and Southern Africa Regional Office, Basic Education and Gender Equality Section.

Trudell, B. (2016b). Language choice and education quality in Eastern and Southern Africa: A review. *Comparative Education, 52(3)*, 281–293.

Trudell, B., & Piper, B. (2014). Whatever the law says: Language policy implementation and early-grade literacy achievement in Kenya. *Current Issues in Language Planning, 15*, 4–21.

UNESCO. (2017). *Global education monitoring report 2017/8, accountability in education: Meeting our commitments*. Paris, France: UNESCO Publishing.

UNESCO. (2018). *World inequality database on education*. Retrieved from: https://www.education-inequalities.org.

Uwezo. (2015). *Are our children learning? The state of education in Kenya in 2015 and beyond*. Kampala, Uganda: Twaweza East Africa.

Van De Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting*

educational and psychological tests for cross-cultural assessment (pp. 39–63). Mahwah, NJ: Lawrence Erlbaum Associates.

Waithaka, E. N. (2017). Choice of the medium of instruction in Kenyan preschools: Averting xenocentrism. *Journal of Education and Practice, 8*(9), 210–216.

Walter, S., & Benson, C. (2012). Language policy and medium of instruction in formal education. In B. Spolsky (Ed.), *The Cambridge handbook of language policy* (pp. 278–300). Cambridge, UK: Cambridge University Press.

Weber, A. M., Fernald, L. C., Galasso, E., & Ratsifandrihamanana, L. (2015). Performance of a receptive language test among young children in Madagascar. *PLoS One, 10*(4), e0121767.

Table 1. Descriptive characteristics of children, caregivers, and households at baseline

	Obs.	Mean	SD	Min.	Max.
Child characteristics:					
Child age in months	505	54.42	17.51	24	83
Child is male	505	0.53	0.50	0	1
Height-for-age *z*-score	489	–0.23	1.40	–4.58	4.39
Child is stunted (HAZ <–2SD)	489	0.08	0.27	0	1
Primary caregiver characteristics:					
Caregiver is child's mother	353	0.85	0.36	0	1
Caregiver is child's father	353	0.02	0.14	0	1
Caregiver is child's grandmother	353	0.11	0.31	0	1
Caregiver mother tongue is Luo	353	0.95	0.23	0	1
Caregiver education: no formal schooling	353	0.02	0.15	0	1
Caregiver education: incomplete primary school	353	0.48	0.50	0	1
Caregiver education: completed primary, not secondary	353	0.41	0.49	0	1
Caregiver education: completed secondary school	353	0.08	0.28	0	1
Caregiver illiterate	353	0.27	0.45	0	1

Caregiver working memory (out of 20)	353	2.43	1.74	0	9
Caregiver depressive symptoms (out of 26)	346	10.33	5.36	0	26
Household characteristics:					
Household size	353	3.03	0.85	2	6
Any children's books in the home	353	0.13	0.34	0	1
Number of children's books in the home	353	0.22	0.71	0	5
Someone has read to the child in past 3 days	353	0.48	0.50	0	1
Family care indicators score (out of 17)	353	7.95	3.58	0	17
Rooms per person	353	0.89	0.43	0.20	3.00
Household has cement floor	353	0.24	0.43	0	1
Household has iron roof	353	0.96	0.20	0	1
Household has electricity	353	0.30	0.46	0	1
Household has latrine	353	0.69	0.46	0	1
Household wealth index	353	0.04	2.22	–3.51	9.80

Notes. Summary statistics on 505 children and 353 caregivers for whom baseline data is available. Baseline height data is missing for 16 children. Seven caregivers declined to answer the questions on depressive symptoms.

Table 2. Descriptive statistics of child measures at baseline

	2 years		3 years		4 years		5 years		6 years		Overall
	N = 106		N = 76		N = 110		N = 113		N = 105		N = 505
	mean	sd / n	mean	sd / n	mean	sd / n	mean	sd / n	mean	sd / n	%
Receptive vocabulary											
English (out of 34)	3.40	3.41	4.26	3.44	6.51	3.76	8.58	3.92	9.17	3.70	
Swahili (out of 31)	3.45	3.40	5.19	3.96	8.44	4.81	11.04	5.00	14.48	5.39	
Luo (out of 27)	5.08	4.20	5.99	3.87	8.77	3.87	11.53	4.90	11.94	5.43	
Expressive vocabulary (out of 30)	1.71	2.02	2.60	2.37	4.82	3.86	8.73	5.33	9.78	5.83	
% of responses in English	4.44%	15.29%	6.42%	16.30%	15.57%	24.08%	28.34%	26.04%	29.02%	27.10%	19.21%
% of responses in Swahili	7.02%	25.54%	4.90%	18.24%	5.03%	16.13%	4.83%	12.97%	3.80%	7.28%	4.96%
% of responses in Luo	88.54%	28.78%	88.68%	24.97%	79.39%	30.25%	66.83%	29.00%	67.18%	29.13%	75.83%
% answered only in English	0.94%	1	0.00%	0	2.73%	3	3.54%	4	4.76%	5	2.55%
% answered only in Swahili	3.77%	4	1.32%	1	0.91%	1	0.00%	0	0.00%	0	1.18%
% answered only in Luo	44.34%	47	59.21%	45	47.27%	52	21.24%	24	20.00%	21	37.06%
Multiple language response	50.94%	54	39.47%	30	49.09%	54	75.22%	85	75.24%	79	59.22%
	N = 106		N = 76		N = 110		N = 113		N = 105		N = 505
Adapted MDAT fine motor (out of 43)	5.52	3.62	12.99	5.34	18.34	4.85	22.65	4.85	25.98	5.05	
Adapted MDAT language (out of 36)	10.22	1.93	13.81	3.28	16.93	3.27	20.78	2.92	22.22	2.64	

Notes. English, Swahili, and Luo vocabulary are raw total receptive vocabulary scores, measured using three assessments based on the British Picture Vocabulary Scale (BPVS). Expressive vocabulary raw scores were measured using a tool developed from the PPVT. Percent (%) English, Swahili, and Luo expressive are the mean (and sd) percent of responses given in each language for the expressive vocabulary test. Only in English, Swahili, and Luo are the percent (and n) of children who answered the expressive vocabulary exclusively in each language. Multiple language response are the percent (and n) of children who answered in more than one language. The adapted MDAT are the raw scores from the Kenya adaptation of the Malawi Developmental Assessment Tool (MDAT).

Table 3. The association between baseline receptive vocabulary and follow-up English receptive vocabulary

Outcome: Follow-up English receptive vocabulary	e1	e2	e3	e4	e5	e6
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
Receptive vocabulary in English (*z*-score)	0.3554****			0.2811****	0.3035****	0.2593****
	0.0478			0.0477	0.0487	0.0485
	<0.001			<0.001	<0.001	<0.001
Receptive vocabulary in Swahili (*z*-score)		0.2669****			0.1618****	0.1034**
		0.0472			0.0469	0.048
		<0.001			0.0006	0.0318
Receptive vocabulary in Luo (*z*-score)			0.3157****	0.2139****		0.1812****
			0.0465	0.0451		0.0463
			<0.001	<0.001		0.0001
Constant	–0.0119	–0.0025	–0.003	–0.0104	–0.0107	–0.0099
	0.0469	0.0483	0.0474	0.0454	0.0458	0.045
	0.7998	0.9585	0.95	0.8187	0.8158	0.8267
R-squared	0.1332	0.0696	0.1009	0.1737	0.1560	0.1821
N. of cases	442	442	442	442	442	442

Notes. English, Swahili, and Luo receptive vocabulary are measured using three assessments based on the British Picture Vocabulary Scale (BPVS). Vocabulary scores are age-adjusted *z*-scores for children ages 2 to 6 years. Baseline and follow-up were conducted

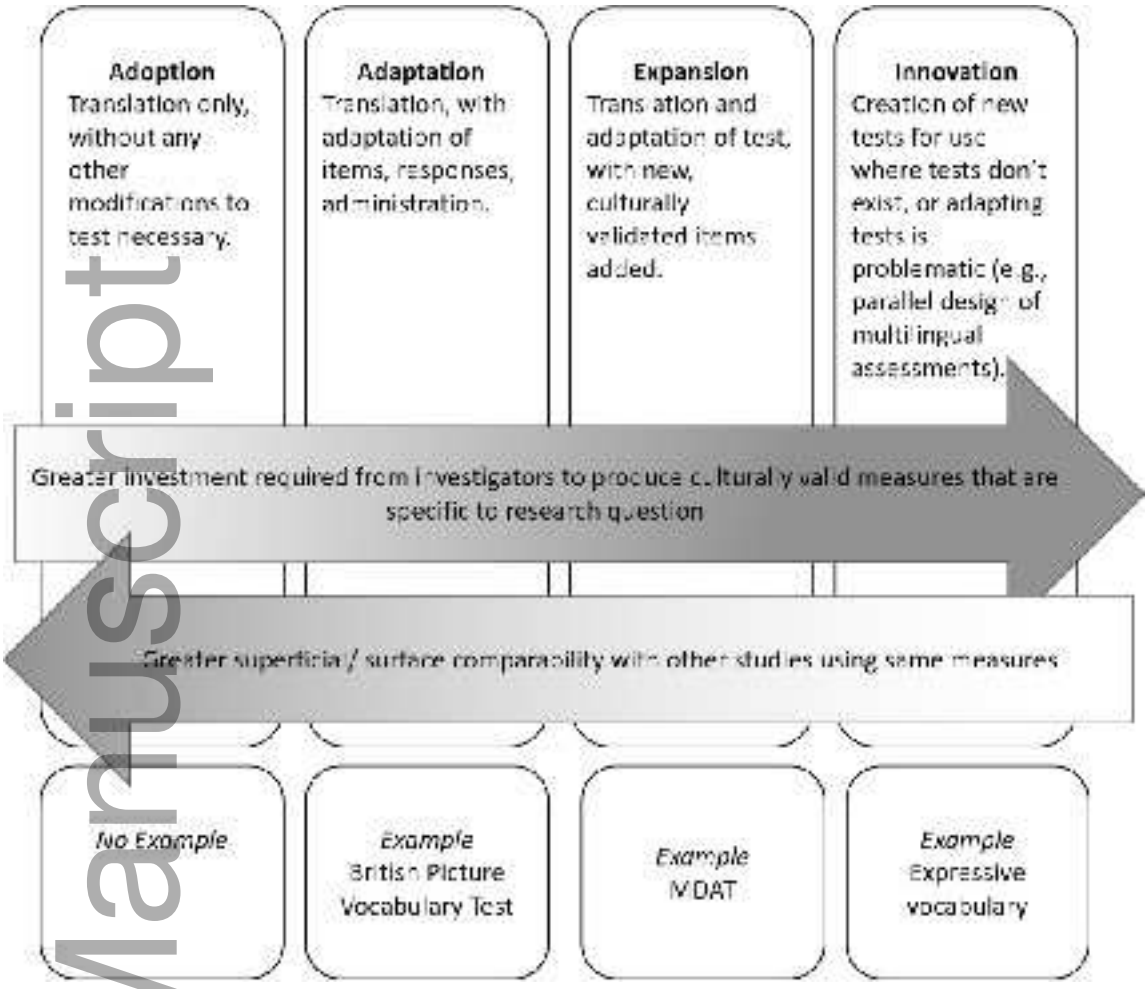
approximately 5 weeks apart. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.001$.

Table 4. The association of baseline caregiver literacy in English and Luo and child follow-up vocabulary scores

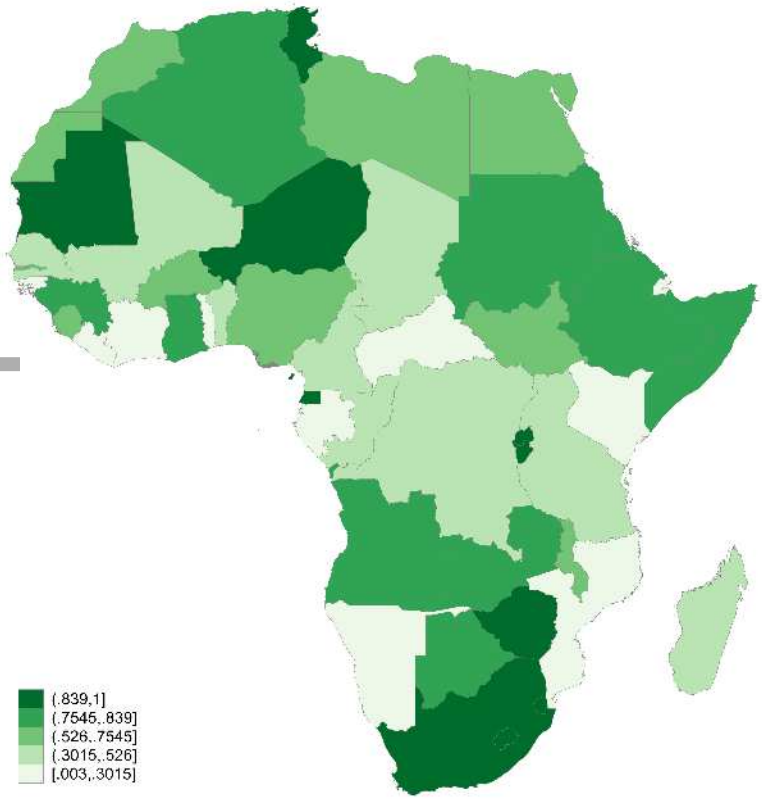
	English receptive vocabulary				Luo receptive vocabulary				Expressive vocabulary			
	cg1	cg2	cg3	cg4	cg5	cg6	cg7	cg8	cg9	cg10	cg11	cg12
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
Caregiver literacy												
English	0.070**		–0.057	–0.098	0.0003		–0.040	–0.077	0.049*		0.027	0.021
	0.028		0.073	0.073	0.029		0.088	0.092	0.029		0.081	0.084
	0.013		0.431	0.181	0.993		0.651	0.405	0.088		0.737	0.800
Luo		0.067***	0.109*	0.110**		0.005	0.034	0.028		0.039*	0.019	0.023
		0.022	0.056	0.055		0.023	0.068	0.068		0.022	0.062	0.063
		0.002	0.056	0.045		0.815	0.619	0.680		0.082	0.760	0.711
Constant	–0.190**	–0.234***	–0.224**	–0.237***	–0.003	–0.021	–0.014	–0.039	–0.142	–0.143	–0.148	–0.143
	0.089	0.088	0.090	0.087	0.0936	0.092	0.094	0.093	0.090	0.089	0.090	0.091
	0.034	0.008	0.013	0.007	0.9735	0.822	0.883	0.677	0.117	0.109	0.103	0.118
R-squared	0.015	0.023	0.024	0.045	0	0	0.001	0.016	0.008	0.008	0.008	0.015
N. of cases	442	442	442	442	442	442	442	442	442	442	442	442

Notes. Receptive English and Luo vocabulary scores are age-adjusted *z*-scores for children ages 2 to 6 years, measured using separate assessments based on the British Picture Vocabulary Scale (BPVS). Expressive vocabulary *z*-scores were measured using a tool developed from the PPVT. Caregiver literacy is number of words (out of 5) a caregiver could read from a simple sentence at a second-grade reading level, adapted from the Early Grade Reading Assessment (EGRA). Baseline and follow-up were conducted approximately 5 weeks apart. The first two models for each vocabulary assessment (cg1 & cg2, cg5 & cg6, cg9, & cg10) are bivariate regressions. The third model for each vocabulary assessment (cg3, cg7, and cg11) include caregiver literacy in both languages. The fourth model for each vocabulary assessment (cg4, cg8, and

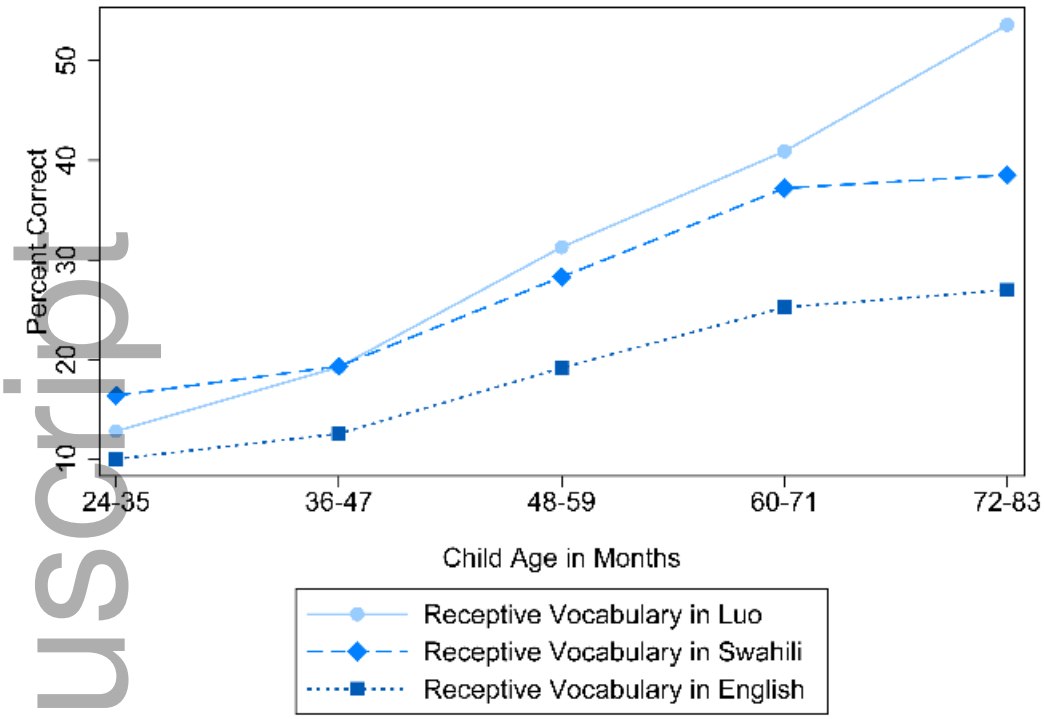
cg12) add controls for caregiver education and household wealth. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.001$.



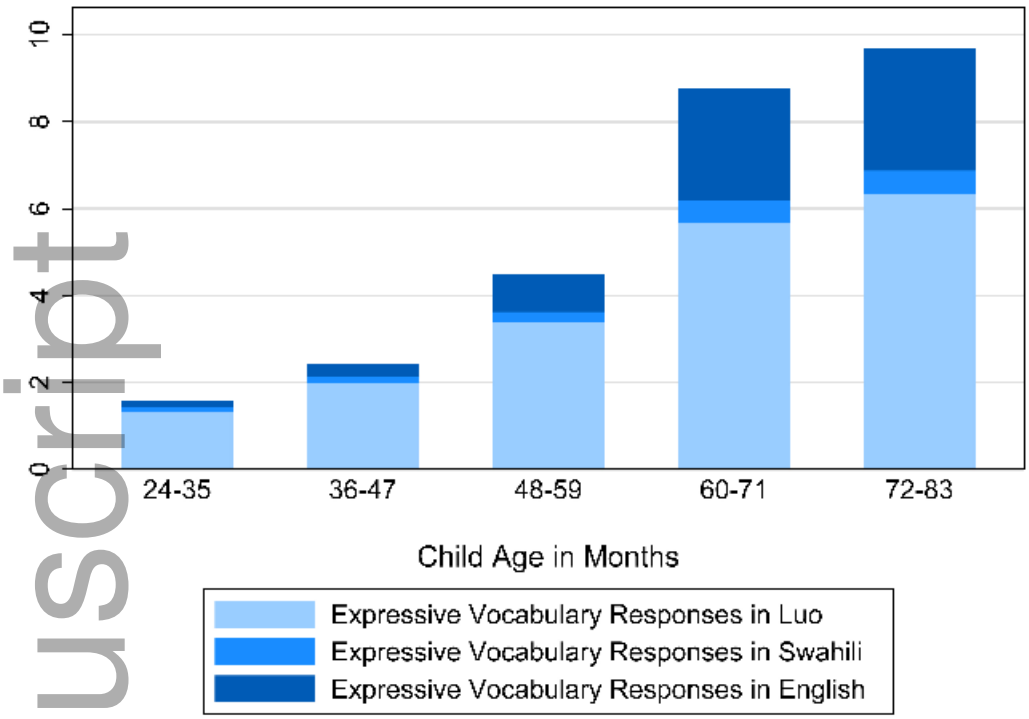
desc_12875_f1.jpg



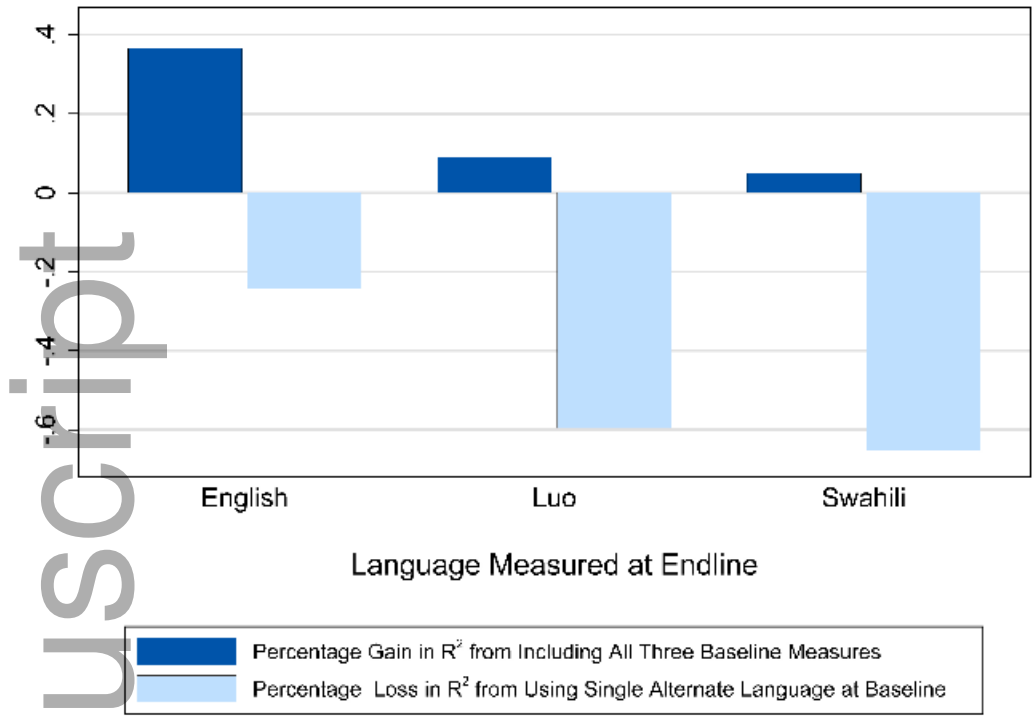
desc_12875_f2.jpg



desc_12875_f3.jpg



desc_12875_f4.jpg



desc_12875_f5.jpg