Examining the lexical development of multilingual children in Ghana using an adapted vocabulary checklist

Joseph R. COFFEY[1], Elizabeth SPELKE[2], & Jesse SNEDEKER[2]

[1] Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University

[2] Department of Psychology, Harvard University

Author Note

This study is registered in The American Economic Association's registry for randomized controlled trials under RCT ID AEARCTR-0000015. The study protocol was approved by the Ghana Health Service Ethics Review Committee, and the IRBs of MIT, Stanford, and IPA. The funding for this study was provided by the British Academy, the JPAL Post-Primary Education Initiative, and USAID-DIV. We have no conflicts of interest to declare. This manuscript was produced using RMd (Xie et al., 2018) and Papaja (Aust & Barth, 2024) on R (R Core Team, 2024) for increased reproducibility. We would like to thank the members of the Spelke Lab who contributed to the development of the vocabulary checklist. We would like to thank Esther Duflo, Pascaline Dupas, Mark Walsh, Erin Grela, Gabriella Fleischman, and Flora Bergounhon at the Abdul Latif Jameel Poverty Action Lab for their work on the Ghana Children and Caregivers Panel Survey, from which our subjects, checklist, and survey measures are drawn. We would also like to thank the staff of Innovations for Poverty Action in Ghana: Stephanie Adjovu, Kwame Akakpo, Renaud Comba, and Isaac Arthur. We would like to thank all our surveyors for their years-long work on data collection: Belinda Bakah, Vivian Ahiatornyo, Mercy Chidi Atanley, Augustina Quarshie, Mabel Akorfa Quame-Amaglo, Gifty Amaney, Anette Fafa Amegayi-Senaya, Stellina Pana, Reindorf Appiah, Kawamu-Din Ibrahim, Jonathan Tetteh, and Amankwah Maxwell.

Correspondence concerning this article should be addressed to Joseph R. COFFEY, LSCP, 29 rue d'Ulm, Paris, France 75005. E-mail: jrcoffey@g.harvard.edu

Abstract

We examine vocabulary growth in children aged 14- to 25-months in Ghana (n=937) using a novel 88-word checklist adapted from the CDI Words and Gestures short form with assistance from local consultants. The measure was designed to assess the downstream impacts of an earlier high school education intervention children's parents had enrolled in as adolescents. Most children were spoken to in a dialect of Akan; half had exposure to at least one other language, most commonly English. Children's primary caregivers were read a list of 88 vocabulary items in their native language and asked to indicate which items their children say or understand. Caregiver responses were found to have high internal consistency, and most items were effective at distinguishing children by ability. Vocabulary size was significantly and positively related to age, female gender, exposure to multiple languages, and caregiver engagement in stimulating activities, but no significant associations with birth order, caregiver education, or living in an urban setting were found. Children's vocabulary composition underwent predictable changes over time: children with lower vocabularies tended to know words for people and social routines, while children with higher vocabularies knew more predicates and closed class items. Nouns were over-represented in children's vocabularies, suggesting the presence of a noun bias. Our results suggest that children in Ghana undergo many similar patterns of development in vocabulary growth as have been found in other languages, reaffirming the usefulness of caregiver-reported language assessments as a tool for studying language acquisition across cultural contexts.

Keywords: vocabulary; Ghana; multilingualism; parent report; item-response theory; Akan

Examining the lexical development of multilingual children in Ghana using an adapted vocabulary checklist

Introduction

The lack of diversity in language acquisition research has received increased attention in recent years. Kidd and Garcia (2022) found that 87% of articles published in major language acquisition journals from 1975 to 2020 were based on North American or European samples, and only 1.5% of the world's languages were represented in those journals. Documenting the course of language acquisition across cultural and linguistic groups is critical for both theory and practice. To build a robust and accurate theory of the human capacity to learn and create languages, we need to consider the full range of contexts in which children are raised. One example of a cross-cultural success story is the noun bias. Early work established that, in a variety of contexts, children initially learn more nouns than verbs (Gentner, 1982; Gentner & Boroditsky, 2001). Subsequent studies identified two cases in which verbs were acquired earlier, diluting or eliminating the noun-bias: Mandarin (Tardiff et al., 1999; Frank et al., 2021) and Tseltal (Casillas et al., 2024). These contrasts provide critical insight on the mechanisms underlying this bias (e.g., the concreteness of nouns and verbs and their frequency in one-word utterances). Thus, our understanding of language acquisition in general is strengthened with each language we study.

Cross-cultural and cross-linguistic data may also shed light on the patterns of individual differences observed within a culture. Some of these patterns appear to be cross-culturally robust (in the contexts studied so far), while others are more variable. Female children and firstborn children show early vocabulary advantage over male and later-born children, though the magnitude of these differences depends on how vocabulary is measured (Huttenlocher et al., 1991; 2010; Hoff-Ginsberg, 1998; Bornstein et al., 2004; Frank et al., 2021). Household education is also a consistent predictor of children's language development (Hoff, 2003; Rowe, 2008; Frank et al., 2021). This may be because more educated caregivers more often engage in activities that have been found to be associated with early language learning, such as play and reading (Jeong et al., 2017; Cuartas et al., 2020). Some studies also find that children living in urban settings have larger vocabularies than children from rural communities, perhaps due to differences in household income and education (Bornstein & Cote, 2005; Vogt et al., 2015; Southwood et al., 2021). This pattern, however, seems to depend on the wider context of urbanization: a recent study comparing rural families in China to recently relocated peri-urban families finds that rural children have larger vocabularies on average (Ma et al., 2024). There is also variation in studies comparing monolingual and bilingual children. While most studies find that multilingualism predicts lower vocabularies in each individual language, the findings on total vocabulary (the sum of both languages) or total conceptual vocabulary (the number of concepts for which the child has a word) are more variable (Bialystok et al., 2010; Cote & Bornstein, 2014; Singh et al., 2023; Byers-Heinlein et al., 2024).

Under-representation also creates real-world consequences that can be felt in the lives of children and caregivers. The growth of children's vocabulary is a key component of early language acquisition that is linked to language and literacy skills in school (Lee, 2011; Can et al., 2013; Bornstein et al., 2014). Supporting language learning is a target for intervention in supporting early childhood education (Suskind et al., 2016; Weber et al., 2017; Wong et al., 2020; Dupas et al., 2023), particularly in low- and middle-income countries where children are at heightened risk of delayed development (Black et al., 2017; Sania et al., 2019). By excluding under-represented groups from this research, we risk failing to learn about the range of different paths leading to healthy development (Rogoff et al. 2017). Using research based on a minority of the world's population to inform early childcare and education policies internationally compromises the effectiveness of these policies and threatens the autonomy of vulnerable communities. Thus, there is an urgent need for more studies focusing on the acquisition of understudied languages in diverse social contexts.

In this paper, we will examine the early vocabularies of 937 children whose parents had participated in a longitudinal randomized control trial as adolescents in 2009. Parents were from low-income households in rural areas of southern and central Ghana and had qualified for admission to high school but failed to enroll due to financial constraints. Roughly one-third of participants were given scholarships that covered high school fees. Since then, these participants have been followed over time to study the effects of high school access on health, income, education, and family status (Duflo, Dupas, & Kremer, 2021).

Prior studies of vocabulary growth in African languages

Caregiver reported measures offer several advantages over direct assessment of young children, particularly in field studies. First, caregiver reports are easier to implement. Direct assessments of children require highly trained surveyors for consistent administration, while caregiver reports can be administered with little training. Direct assessments also require quiet, private spaces to avoid distraction, which is often impossible in settings where testing is conducted outdoors. They also involve testing children, who are less compliant and more distractible than adults, limiting the number of questions or trials that can be given. In contrast, caregiver reports allow for more thorough and broad assessments of the child's behavior. A popular caregiver reported language measure is the MacArthur-Bates CDI, a vocabulary checklist that has been adapted into almost 100 languages (Marchman & Dale, 2023). Adaptations of the MB-CDI have been previously used to measure the effects of child educational interventions in low- and middle-income countries, particularly short form CDIs that provide a brief snapshot of children's language development (Fenson et al., 2000). In other cases, the MB-CDI Short Form has been used as a starting point for creating vocabulary checklists specialized for use in particular contexts, such as highly multilingual settings (Vogt et al., 2015; Prado et al., 2018).

We located six papers that have adapted some version of the CDI (Words and Gestures or Words and Sentences, both short and long forms) for use in six African countries (Table 1). These adaptations typically focused on rural communities outside of major cities, but some included communities living in or around urban centers. Many of the communities were multilingual; in these cases, researchers either focused on a single linguistic group within a country (Childers et al., 2007; Weber et al., 2018), created a CDI for each language spoken (Alcock et al., 2015; Southwood et al., 2021), or created a conceptual CDI that tracked vocabulary across all languages spoken in a household (Vogt et al., 2015; Prado et al., 2018). Half of the studies sought formal approval from the CDI advisory board; primarily those that created dedicated checklists for an individual language. Most of the studies began by translating an existing CDI and then refining the measure through local consultation and piloting; however, Prado et al. (2018) created an initial word list by eliciting common child words from caregivers.

All of these studies reported statistics that supported the validity of the measure. For example, four studies found vocabulary to be correlated with the age of the child, and four found correlations between the parent report and the child's production in a speech sample. Two studies reported convergent validity with other assessments of child development adapted for the setting, such as the MDAT (Gladstone et al., 2010), or with assessments that were being normed in parallel, such as the Wolof Milestone Checklist (Weber et al., 2018). In addition, Prado et al. (2018) reported a later follow-up with their Ghanaian participants in preschool where their previously measured vocabulary was positively related to performance on language and cognitive testing, demonstrating the predictive validity of the checklist. Reliability statistics were less commonly reported; Cronbach's alpha was reported as a measure of internal consistency in two studies. The size of these samples varied greatly, from 8 children to almost 900 (Prado et al., 2018), all between the ages of 12 and 30 months, the age range in which the CDI is typically administered in other settings.

These studies also frequently reported correlations between vocabulary scores and the demographic characteristics of their sample. For instance, girls were found to have larger vocabularies on average (Vogt et al., 2015; Weber et al., 2018; Southwood et al., 2021), as were children living in more urban areas (Vogt et al., 2015; Southwood et al., 2021). The relationship between household education and vocabulary varied: Vogt et al. (2015) reported a positive correlation (with productive vocabulary, but not receptive vocabulary) and Southwood et al. (2021) found no correlation with their productive vocabulary checklist.

In some cases, researchers adapted the CDI specifically to test the efficacy of randomized controlled trials. Prado et al. (2018) created the Malawi and Ghana checklists to test the effects of a lipids-based nutritional supplement on 18-month-olds' language development (Prado et al., 2017). Weber et al. (2018) adapted the CDI-WG Short Form into Wolof to test the effects of an educational intervention in which Senegalese mothers from rural, low-income communities were enrolled in a parenting program that encouraged verbal engagement with their infants (Weber et al., 2017).

Other researchers have used their adapted CDIs to explore theoretical questions in understudied populations. Childers et al. (2007) used the checklist to investigate the presence of a noun bias in Ngas-speaking children; they found little evidence for an early bias, which they argued reflect properties of Ngas that made verbs easier to acquire. They also found that time spent in joint attention with caregivers predicted overall vocabulary size. Mastin and Vogt (2016) used the Mozambican CDI adaptation to examine how different social experiences shaped the vocabularies of children living in urban and rural communities. They found that time spent in joint attention with caregivers predicted vocabulary in urban children vocabulary, while time spent in one-on-one interactions with caregivers predicted vocabulary in rural children.

Table 1

Summary of previous adaptations of vocabulary checklists in Africa

Paper	Country	Languages	Adaptation	Participants	Validity	Reliability
Childers (2007)	Nigeria	Ngas	WG-Short	16 (1;0-2;0)	Age	None
Alcock (2015)*	Kenya	Kiswahili, Kigiriama	WG/WS-Short/Long	300 (1;0-2;6)	Age, speech, tests, other	Cronbach's, test-retest, parallel form
Vogt (2015)	Mozambique	Many	WG-Short (conceptual)	637 (1;1-2;1)	Age, speech	None
Prado (2018)	Malawi and Ghana	Many	WG-Short (conceptual)	30 and 869 (1;5-2;1)	Speech, tests	None
Weber (2018)*	Senegal	Wolof	WG-Short	500 (1;6-2;6)	Speech, tests; Rasch modeling	Cronbach's, person-sep
Southwood (2021)*	South Africa	Afrikaans, SAE, Xhosa, Tsonga	WS-Long	98-115 each (1;4-2;8)	Age	None

Note. WG: Words and Gestures; WS: Words and Sentences; Short form: ~100 words; Long form: ~350 words for WG & ~700 words for WS; Conceptual: total vocabulary across all languages known; Validity gives correlations between checklist score and other variables (child age, naturalistic speech samples, language testing, etc.); * indicates official CDI adaptations of each language listed with advisory board approval

Current Study

Beginning in 2017, surveyors conducted biennial check-ups on the children of the original intervention participants to assess the possible downstream effects of high school scholarships on child health and cognitive development (Duflo et al., 2024). At each visit, children were administered an age-adjusted testing battery that assessed their language, executive function, and numerical, spatial, and mental-state reasoning. For 14- to 25-months-old children, their primary caregiver was asked to report their vocabulary using a checklist adapted from the American English MacArthur Bates CDI (Fenson et al., 1994). Children of scholarship recipients did not differ in vocabulary size from controls (Duflo et al., 2024). However, children of female scholarship recipients were found to have a lower mortality rate and their caregivers engaged in more preventative health behaviors (e.g., prenatal care), as well as more play and conversation with their child, as measured by both reported activity and naturalistic daylong recordings.

Here, we explore children's performance on the vocabulary checklist, collapsing across the treatment and control conditions. Ghana is a highly multilingual country, with most people speaking at least two local languages (Omane et al., 2023). Most of the children in the sample were spoken to in a dialect of Akan, a Niger-Congo language that is the most prevalently used language in the country (Osam, 2003). Akan syntax is primarily SVO. Akan phonology permits V, C, and CV syllables and exhibits vowel harmony within words (Dolphyne, 1988). Unlike languages in other Niger-Congo families like Bantu, Akan lacks a noun class system but relies on a complex system of prefixing and suffixing to mark both singular and plural (Osam, 1993). Verbal morphology is marked for tense, aspect, and mood through a combination of affixation and grammatical tone (Osam, 2003). Due to contact with English, loan words are also common, particularly in education, politics, health, and sports (Apenteng & Amfo, 2014). Akan is composed of two major mutually-intelligible dialect groups, Twi and Fante. However, English is the official language of Ghana and is used in government, education and business. Given this context, we used the checklist to assess children's productive and receptive conceptual vocabularies across all languages they were exposed to at home.

It is important to highlight that this checklist is not a formal adaptation of the CDI Words and Gestures, which would require far more controlled and language-specific vetting than was able to be performed here. Instead, it is a caregiver reported tool created for the purpose of evaluating a randomized controlled trial on a diverse, multilingual sample of families in Ghana, using the CDI Words and Gestures as a starting point (similar to e.g., Vogt et al., 2015; Prado et al., 2018). According to a review of 237 randomized controlled trials registered through the American Economic Association, less than one-third of interventions targeting children's school-related skills reported any measure of contextual validity or reliability for their assessment (Macours, Williams, & Wolf, 2023). Because of this, the first part of our analysis focuses on establishing its effectiveness as a measure of individual differences. We have also tried to control for sources of variance, such as differences across surveyors and differences in form translation, that may have been introduced during administration. It is our hope that our evaluation procedure might help other researchers improve the ecological validity of measures designed for use in settings where data collection is challenging.

First, we investigate the reliability and validity of the checklist. A previous study of the assessment battery on the children in the control condition found that reported receptive vocabulary was associated with later performance on a receptive picture vocabulary test constructed to measure the intervention effects at 3 years ($r=0.19$), providing evidence for its predictive validity (Coffey & Spelke, 2024). However, no item-level analysis of checklist performance was reported. To do so, we constructed a Rasch model, a psychometric model that uses item responses to produce estimates of item difficulty and respondent ability. This model allows us to determine whether responses to individual items deviate significantly from their predicted values, suggesting their inadequacy as test items. Second, we explored the predictors of children's vocabulary size, including gender, multilingualism, birth order, child-centered stimulating activities, and whether children resided in or close to a city. Finally, we examined differences in the syntactic composition of children's vocabularies. In keeping with previous cross-linguistic studies of caregiver-reported vocabulary, we expected to find that nouns and social words made up a larger proportion of words known than predicates and closed-class items for children with lower total vocabularies, compared to children with larger vocabularies (e.g., Braginsky et al., 2019).

Methods

Participants

In total, 986 caregivers were administered the vocabulary checklist. After excluding children who were tested outside of the original age range (n=3), children who were missing information about caregiver education (n=5) and birth order (n=32), and duplicate tests (n=9), we arrived at a final sample size of 937 children (Table 2). The vocabulary assessment was initially administered to children from 14- to 22-months of age, and up to 25-months beginning after 2018 (Duflo et al., 2024). All data was reported by the child's primary caregiver, who in almost all cases was the child's biological mother (95%). Most caregivers reported speaking to their children mostly in some dialect of Akan (Twi or Fante) (Table 3). However, over half of caregivers also reported their children were regularly exposed to more than one language, typically Akan and some combination of Ghanaian English, Nzema, and/or Ewe, corroborating other surveys of language exposure in Ghanaian homes (Omane et al., 2023). Based on caregiver-reported language use, these homes were considered multilingual. A little over half of parents lived far outside of urban areas (n=NA, 71 missing). The sample reflected the proportion of participants assigned to the treatment condition, with two-thirds of children coming from control households.

Primary caregiver education ranged from no primary school to post-secondary education. Caregivers were also surveyed on whether anyone in the home had engaged the child in storytelling, playing, singing, reading, or drawing, counting, and object-naming in the last month. These activities were based on those used in the Multiple Indicators Cluster Survey (UNICEF, 2005) and selected based on previous research that has found them to be associated with the pace of children's cognitive development (Jeong et al., 2017; Cuartas et al., 2020). On average, caregivers reported that children had engaged in at least 3 of these activities in the last month, most commonly playing and singing.

Materials

The MacArthur-Bates Communicative Development Inventories Words & Gestures Short Form (hereafter CDI-WG for brevity) was used as a starting point for the checklist (Fenson et al., 2000). The CDI-WG short form was designed for use with infants 8 – 18 months of age. We chose it because our plan was to focus data collection efforts on children 14- to 22-months of age, thus our group fell between the age range for the CDI-WG and the age range for the Words & Sentences short form (CDI-WS) which targets 16- to 30-month-olds. The CDI-WG includes words that are more appropriate for younger children than the CDI-WS and asks parents to report not only on vocabulary production, but also on comprehension, which is typically difficult to report accurately for older children with larger vocabularies. We decided to base our checklist on the instrument for the younger age group for two reasons. First, we reasoned that there was a greater risk of getting floor effects if we used the CDI-WS than there were of getting ceiling effects if we used the CDI-WG given the pattern observed at the youngest and oldest ages at which each measure is used (Frank et al., 2017). Second, the success of the Oxford English CDI, which was normed on children between the ages of 12-24 months suggested that it was possible to measure vocabulary comprehension and production with a single instrument across this age range (Hamilton et al., 2000).

Table 2

Summary of child information surveyed from caregivers

	n	(%)
Total children	937	
Condition		
Treatment	321	(34%)
Control	616	(66%)
Gender		
Female	478	(51%)
Male	459	(49%)
Birth Order		
Firstborn	300	(32%)
Secondborn	312	(33%)
Laterborn	325	(35%)

Note. n and % give the total number of children and the percentage of the total sample.

Table 3

Summary of surveyed caregiver and household information

	n	(%)
Total caregivers		
Caregiver education		
High school degree	312	(33%)
No high school degree	625	(67%)
Caregiver primary language		
Akan	784	(84%)
Ghanaian English	33	(4%)
Nzema	26	(3%)
Ewe	19	(2%)
Hausa	13	(1%)
GaDangme	11	(1%)
Other	51	(5%)
Number of reported languages		
One	411	(44%)
Two	474	(51%)
Three	49	(5%)
Four	3	(0%)
Activities with children in the last month		
Storytelling	238	(25%)
Playing	888	(95%)
Singing	761	(81%)
Reading	358	(38%)
Other	509	(54%)

Note. n and % give the total number of caregivers and the percentage of the total sample.

However, to ensure the measure is still sensitive in this age range, we examined both vocabulary sub-scores for evidence of ceiling effects. We found no evidence for these even in the oldest children, suggesting that our measure captures meaningful differences in children's vocabularies across this age range. We also conducted separate analyses of item-based performance for production and comprehension. This analysis could reveal whether some items performed appropriately as measures of productive vocabulary but not receptive vocabulary.

Before the first round of testing, the English CDI-WG was piloted with 50 caregivers of young children in the Volta region of Ghana, to ensure there was no overlap with families participating in the intervention. These families were a convenience sample, assumed to be similar to the intervention sample with slightly less education. Local surveyors were also consulted on the appropriateness of each vocabulary item for Ghanaian children. This information allowed researchers to eliminate items that did not appear to be relevant in this context. Words without appropriate translations into Akan (e.g., *away*) or those that lacked proper cultural equivalents (e.g., *applesauce*) were removed. Others were replaced with analogous items that were likely to be familiar to children (*candy* to *toffee*, *ouch!* to *ajeii!*). The vocabulary list was first drafted in English (i.e., the language of education and government). Later, this list was translated into Akan (with Twi and Fante variants, where applicable) beginning in round 12 to increase the consistency of surveyor translation. A total of 387 caregivers were administered the survey with these translations.

The final checklist consisted of 88 lexical items divided into 17 syntactic-semantic categories. As in previous studies, words were classified into one of four syntactic groups based on which semantic category they belonged to (Bates et al., 1994; Caselli et al., 1995; Coffey & Snedeker, 2025). A full comparison of the checklist with the CDI-WG wordlist can be found in Supplementary Materials. These included 24 social words (sound effects, games and routines, and people), 42 nouns (animals, vehicles, toys, food and drink, clothing, body parts, furniture, household items, and outside things), 14 predicates (action words and descriptive words), and 8 closed-class items (time words, pronouns, and question words). Due to improper coding of three food and drink nouns (*meat*, *eggs*, *beans*), we only examine 85 lexical items in this analysis.

Procedure

Surveyors conducted 17 rounds of testing between 2017-2023. For each round, surveyors were assigned to a geographical area in and around Accra to conduct surveys of families (based on their last known address) for 3-4 months. The vocabulary checklist was administered to children's primary caregivers as a part of the survey on their child's development. Caregivers were asked to report on the conceptual vocabulary of a given child when the child reached the eligible age range, such that each child had a single report. Due to mixed literacy and language use in the sample, the checklist was read verbally to caregivers in the language they reported speaking to the child, as has been done in other similar settings (e.g., Alcock et al., 2015). In cases where the primary caregiver spoke a language other than Akan or English, the surveyor gave translation equivalents for each of the words in the language spoken to the child at home. As each word was read from the list, the caregiver was asked to indicate whether the child understood or could say the word in any of the languages spoken in the home, producing a single measure of conceptual vocabulary across languages.

Results

Summary

The analyses below were conducted in R (v 4.4.1) (R Core Team, 2024). On average, children could produce 21 words (25%) and understand 40 words (47%) (Table 4). As a first pass check of reliability, we checked the correlation between productive and receptive vocabulary sub-scores. We believed these scores would not be correlated if the caregiver responses were not properly logged, caregivers did not understand the instructions given by the surveyor, or caregivers were not reliable reporters of their children's vocabularies. Productive and receptive sub-scores were found to be highly correlated ($r = 0.80$, $p < .001$). Examining the distribution of these scores, we see that productive vocabulary is right-skewed, while receptive vocabulary is roughly normally distributed (Figure 1). We find significant associations between child age and vocabulary size (receptive $r = 0.57$, $p < .001$; productive $r = 0.68$, $p < .001$) (Figure 2).

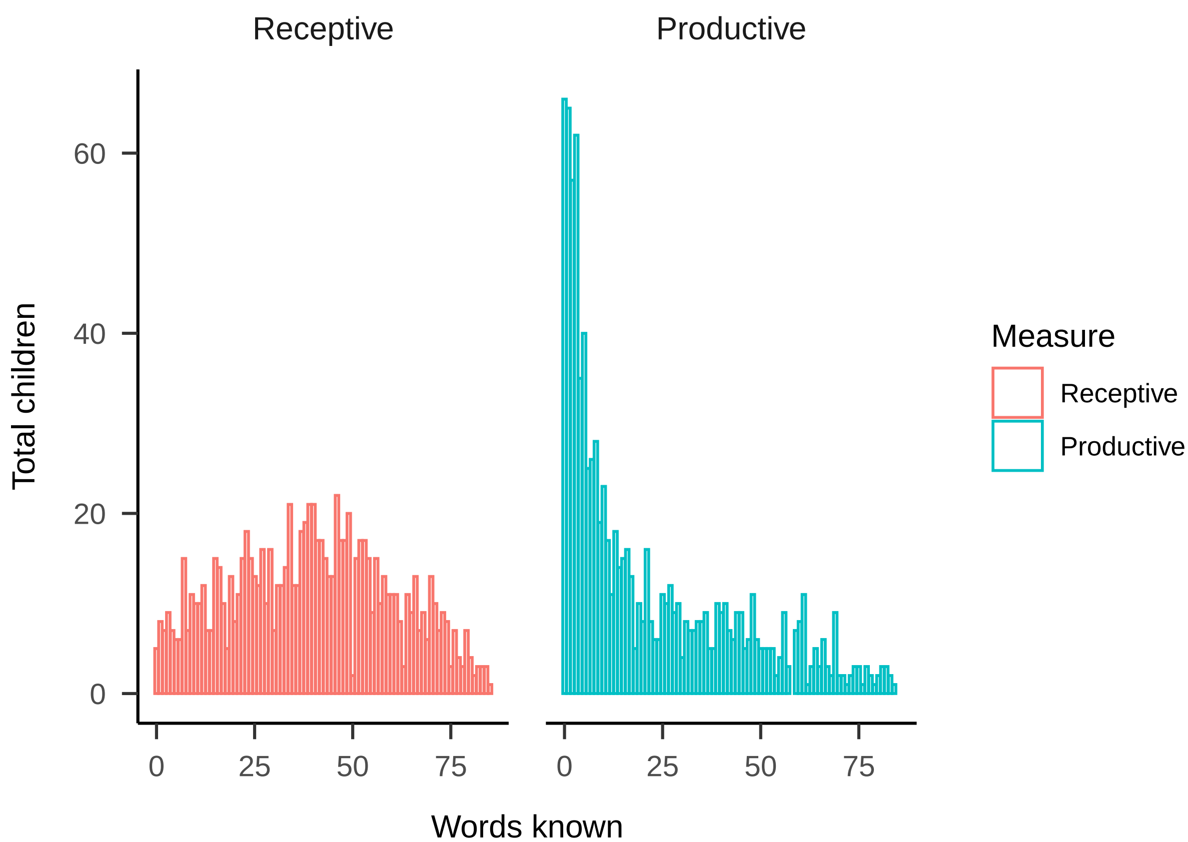

Figure 1. The distributions of children's receptive (left) and productive (right) conceptual vocabulary size. The plot is a histogram where each column signifies the number of children with that vocabulary size. Productive vocabulary is more left-skewed than receptive vocabulary, with many caregivers reporting their children do not speak yet.

Table 4

Summary of mean vocabulary scores by child age

Ages	n	%	Production	Comprehension
14mo	111	12%	5	24
15mo	144	15%	8	28
16mo	121	13%	11	31
17mo	111	12%	15	37
18mo	103	11%	20	41
19mo	95	10%	26	44
20mo	73	8%	29	50
21mo	39	4%	37	53
22mo	51	5%	40	53
23mo	35	4%	49	59
24mo	23	2%	56	64
25mo	31	3%	61	67
Total	937		21	40

Note. n and % give the total number of children at each age and the percentage of the total sample. Production and comprehension give the respective mean checklist sub-scores of children in each age group.

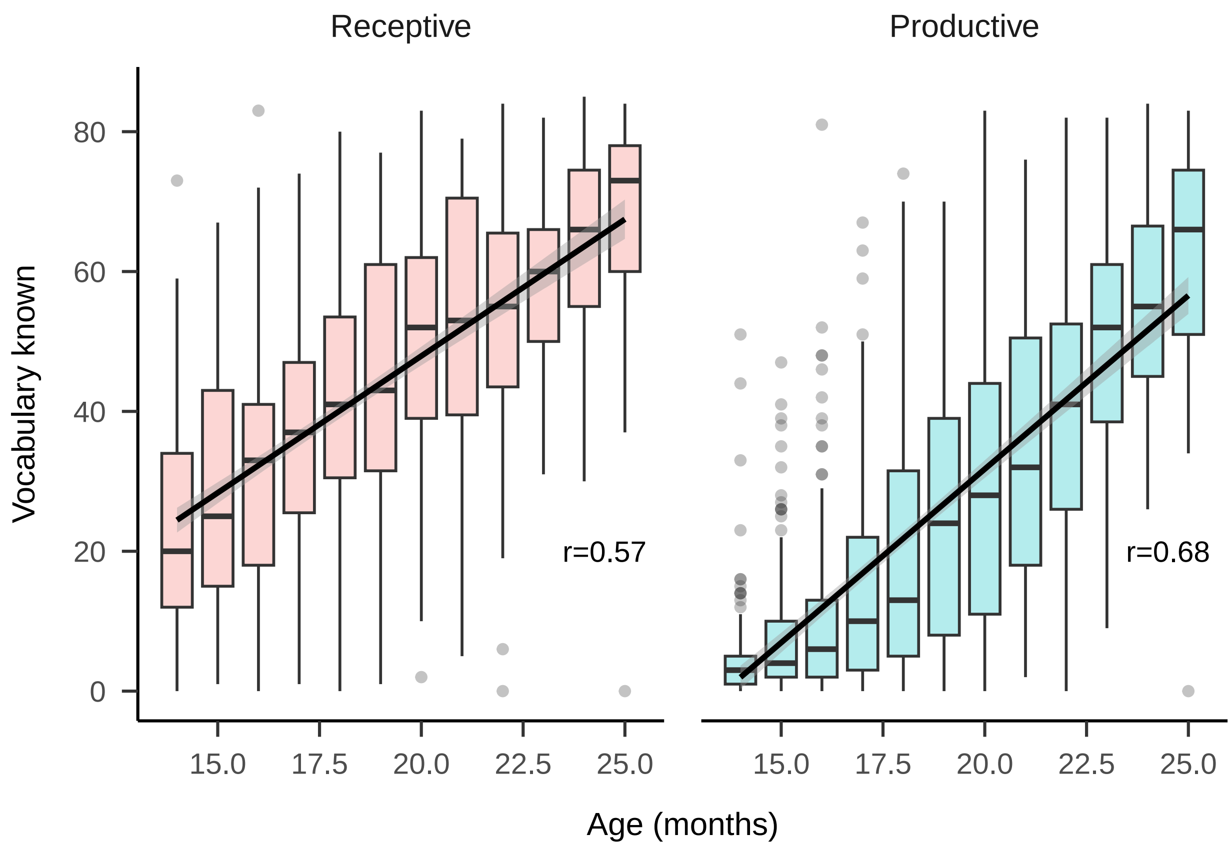

Figure 2. Children's receptive (left) and productive (right) conceptual vocabulary. The plot is a series of box plots where each box gives the median and interquartile range of vocabularies in that age group, with a line fit to indicate the overall correlation between age and vocabulary size given by r.

Psychometrics properties of the checklist

We fit the data to a Rasch model, a single parameter item response theoretic model that simultaneously estimates each child's latent ability and the difficulty of each item, using the *eRm* package (Mair & Hatzinger, 2007). Previous evaluations of CDI measures have used the Rasch models as simple and flexible models of vocabulary knowledge (Weber et al., 2018; Kachergis et al., 2023). With this model, we examined the psychometric properties of the assessment by examining instrument reliability and item-level performance.

Instrument-level Reliability. To measure instrument-level reliability, we calculated Cronbach's alpha, which is the equivalent of taking split-half reliability for each possible division of the trials. Both sub-scores (productive and receptive) demonstrated high internal consistency as measured by Cronbach's α (productive = 0.98; receptive = 0.98). From the Rasch model, we also calculated another measure of reliability, person-separability, which measures the amount of person parameter variance not attributable to error. The sub-scores each had high person separability (productive = 0.95; receptive = 0.96). These converging results indicated that productive and receptive vocabulary measures were reliable.

Item-level performance. We examined how well children's production and comprehension of each word fit our Rasch model by calculating its information-weighted mean square fit, or infit, which has been previously used to evaluate item-level performance in varied assessments of child development and psychology (e.g., Weber et al., 2018; Weisz et al., 2019; Duran et al., 2019). Infit is expressed as the ratio of observed residuals by the expected residuals of the model. Ideally, vocabulary items will have an infit of 1, when observed and expected variance are equal. An infit greater than 1 indicates higher variance than expected, or underfitting, which suggests the presence of noise or unmodeled sources of variance. These items are less able to discriminate between higher- and lower-performing children. An infit less than 1 indicates lower variance than expected, or overfitting to the model, such that these items overestimate differences in performance.

To identify the items with poor fit, we set acceptability cut-offs between 0.75 and 1.33 (following Weber et al., 2018, citing Wilson, 2005). For production, we found that 6 words had infits above threshold, signaling low discriminability (*meow*, *ajeii*, *grandma*, *grandpa*, *shh*, *thank you*) and 4 had infits below threshold (*leg*, *cup*, *spoon*, *mouth*), signally over discrimination. For receptive vocabulary, no words fell outside of this range.

Every word with high infits was a social word, which are generally acquired by children in the first few months after their first word. These words have been found to have lower discriminability across different CDIs (Frank et al., 2021). There are many possible explanations for this. Caregivers might overestimate children's knowledge of these words. Knowing these words could depend on characteristics of the household unrelated to language ability, like the presence of certain family members (i.e., grandparents) or routines (i.e., *meow*, *thank you*, *ajeii*, *shh*). Usage of some words may reflect individual and family stylistic preferences: children might know what ajeii! (*ouch!*) means but use a different word or phrase when expressing pain. In contrast, all words with lower infit and higher discriminability are common nouns, referring to body parts or things found in the kitchen. It is possible that some parents teach these words as sets or use them in the same routines (*leg* and *mouth*) or (*cup* and *spoon*), and thus production of these words is more indicative of which routines the parent uses than it is of the child's language level. It is also possible that words like cup and spoon might only have been produced by older children who are weaned and would use dishware and thus not be related to overall language level. This might also explain why these words perform well in receptive vocabulary.

Vocabulary

We examined the degree to which demographic characteristics that have been associated with language development in other studies predicted productive and receptive conceptual vocabulary size in the dataset. We constructed linear mixed-effects models with *lme4* (Bates et al., 2015), using the number of words on the checklist known as the response variables. Each model was constructed with a random intercept for the surveyor who administered the checklist. This was done to control for any variance in children's vocabulary size introduced by individual surveyors, whose choice of translation of certain words may have differed. Using stepwise model comparison, we determined whether predictors improved overall model fit by conducting likelihood ratio tests (reported by χ^2 statistic) and comparing corrected Akaike Information Criterion. We introduced predictors into the model according to the amount and consistency of evidentiary support for their effects on vocabulary: age (months from 14 months onward), child gender (treatment coded, male reference), primary caregiver education (treatment coded, "did not attend high school" reference), birth order (treatment coded, firstborn reference), caregiver-reported child-centered stimulating activities in the last month (summed and converted to z-score), and multilingualism in the home (treatment coded, monolingual reference). All continuous variables were converted to z-scores; in our models we report standardized coefficients. In addition, a quadratic term for age was also introduced, as previous work has found that the rate children's vocabulary growth accelerates as they grow older (Pan et al., 2005). Because information about urbanness (treatment coded, rural reference) was missing for 71 children, we refit our best-fitting models on a subset of participants with this information to determine whether there was any vocabulary advantage present in children living in urban areas.

For receptive vocabulary, the best-fitting model included child age (main ß $= 0.18$, $t = 20.81$, $p < .001$), gender (ß $= 0.18$, $t = 3.61$, $p < .001$), stimulating activities (ß $= 0.18$, $t = 6.64$, $p < .001$), and multilingualism (ß $= 0.16$, $t = 2.96$, $p = 0.003$) (Figure 3). Nakagawa's R^2 indicated that fixed effects terms explained 36.50% of variance and the random intercept for surveyor explained 3.80%. Adding a term for checklist translation did not improve model fit ($\chi^2(3) = 3.37$, $p = 0.34$). Fitting our best-fitting model to the subset of data with urbanness, we did not find it to be a significant predictor of receptive conceptual vocabulary size (main ß $= 0$, $t = -0.06$, $p = 0.954$).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	-0.03	-0.76***	-0.79***	-0.85***	-0.85***	-0.87***	-0.85***	-0.93***	-0.90***
	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)
Age (months)		0.19***	0.21***	0.19***	0.19***	0.19***	0.18***	0.18***	0.18***
		(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age^2			-0.00						
			(0.00)						
Child is female				0.19***	0.19***	0.18***	0.18***	0.18***	0.18***
				(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Caregiver graduated high school					0.02				
					(0.06)				
Child is secondborn						-0.03			
						(0.07)			
Child is laterborn						0.08			
						(0.06)			
Frequency of child-centered activity							0.19***	0.18***	0.18***
							(0.03)	(0.03)	(0.03)
Multilingual home								0.16**	0.16**
								(0.05)	(0.05)
Surveyed in Akan (after translations)									-0.08
									(0.06)
Surveyed in English									-0.16
									(0.14)
Surveyed in another language									0.04
									(0.08)
SD (Intercept Surveyor)	0.28	0.18	0.18	0.17	0.17	0.17	0.20	0.20	0.18
SD (Observations)	0.97	0.81	0.81	0.80	0.80	0.80	0.78	0.77	0.77
No. Observations	937	937	937	937	937	937	937	937	937
Marginal R^2	0.000	0.317	0.317	0.327	0.327	0.329	0.359	0.365	0.368
Conditional R^2	0.079	0.350	0.350	0.355	0.355	0.358	0.399	0.403	0.402
AICc	2636	2288	2289	2277	2279	2279	2224	2217	2220
LRT (χ^2)		350.62***	0.57	11.69	0.08	2.98	55.4***	8.72**	3.37

* p < 0.05, ** p < 0.01, *** p < 0.001

AICc: corrected Akaike Information Criteron; LRT: likelihood ratio test; Best-fitting model boxed

Figure 3. Mixed effects linear regression models predicting receptive conceptual vocabulary

For productive vocabulary, the predictors of the best-fitting model were the same: child age (main $ß = 0.14$, $t = 5.59$, $p < .001$; quadratic $ß = 0.01$, $t = 3.84$, $p < .001$), gender ($ß = 0.18$, $t = 3.94$, $p < .001$), stimulating activities ($ß = 0.15$, $t = 6.16$, $p < .001$), and multilingualism ($ß = 0.10$, $t = 2.05$, $p = 0.041$) (Figure 4). Notably, the child age quadratic improved model fit, indicating that productive (but not receptive) conceptual vocabulary size increases non-linearly. Fixed effects terms explained 50.90% of variance and the random intercept for surveyor explained 1.20%. Checklist translation did not improve model fit ($\chi^2(3) = 2.91$, $p = 0.41$). Urbanness was also not significantly related to conceptual productive vocabulary (main $ß = -0.01$, $t = -0.16$, $p = 0.873$).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	0.04	-0.86***	-0.75***	-0.83***	-0.83***	-0.85***	-0.82***	-0.88***	-0.91***
	(0.07)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
Age (months)		0.23***	0.14***	0.14***	0.14***	0.14***	0.14***	0.14***	0.13***
		(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Age^2			0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Child is female				0.18***	0.18***	0.18***	0.18***	0.18***	0.18***
				(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Caregiver graduated high school					0.00				
					(0.05)				
Child is secondborn						-0.00			
						(0.06)			
Child is laterborn						0.05			
						(0.06)			
Frequency of child-centered activity							0.16***	0.15***	0.14***
							(0.02)	(0.02)	(0.02)
Multilingual home								0.10*	0.10*
								(0.05)	(0.05)
Surveyed in Akan (after translations)									0.08
									(0.06)
Surveyed in English									-0.07
									(0.13)
Surveyed in another language									0.05
									(0.07)
SD (Intercept Surveyor)	0.26	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.11
SD (Observations)	0.97	0.72	0.72	0.71	0.71	0.71	0.69	0.69	0.69
No. Observations	937	937	937	937	937	937	937	937	937
Marginal R^2	0.000	0.467	0.474	0.482	0.482	0.483	0.506	0.509	0.511
Conditional R^2	0.066	0.481	0.488	0.496	0.496	0.496	0.518	0.520	0.522
AICc	2642	2070	2059	2047	2049	2050	2003	2000	2004
LRT (χ^2)		574.76***	12.4***	14.68***	0.01	1.02	46.01***	4.18*	2.91

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

AICc: corrected Akaike Information Criteron; LRT: likelihood ratio test; Best-fitting model boxed

Figure 4. Mixed effects linear regression models predicting productive conceptual vocabulary

Syntactic composition

Finally, we examined how the syntactic composition of children's vocabularies changed over time. It is well-established across many languages that children learn nouns and social words like kinship terms and common routines earliest, followed by verbs and adjectives, and then closed-class items like such as articles or prepositions (Fenson et al., 1994; Caselli et al., 1995; Gentner et al., 2001; Bornstein et al., 2004; Braginsky et al., 2019). We used the approach described in Frank et al. (2021) using the *langcog* package to compare the proportion of words in each category that children know to their overall vocabulary size.

We first capture the trends in vocabulary composition by constructing a generalized linear model predicting the proportion of words known in each category to the total proportion of the checklist known, fitting third-order polynomials to the model to capture nonlinear relationships in the data. The regression is constrained such that the coefficients of the regression sum to 1. This ensures that the model treats children that are reported as knowing the entire checklist (i.e., x=1) as also knowing every word within a given syntactic category (i.e., y=1). Assuming that we would expect the proportion of words known in a given category to increase monotonically with total knowledge of the checklist, a line running above the diagonal would indicate over-representation of a category in children's vocabulary, and a line running below the diagonal would indicate under-representation.

Next, we quantify the degree of over- or under-representation by calculating the total area under the curves, with positive values indicating over-representation and negative values indicating under-representation. We employ bootstrapping to estimate uncertainty using the *boot* package (v1.3.30; Canty & Ripley, 2021), randomly resampling participants from the population with replacement and recalculating these area values to produce bias-adjusted confidence intervals (R = 1000).

For children's receptive vocabulary, we find that nouns are relatively over-represented in vocabulary (mean = 0.06, 95% CI [0.05, 0.06]). There is also under-representation of predicates (mean = -0.03, 95% CI [-0.04, -0.03]) and closed-class items (mean = -0.12, 95% CI [-0.13, -0.11]). Social words are roughly proportionately represented at smaller vocabularies and become less represented as vocabulary grows. Overall, this results in under-representation (mean = -0.03, 95% CI [-0.04, -0.03]) (Figure 5).

The same patterns can be observed in productive vocabulary, with over-representation of nouns (mean = 0.06, 95% CI [0.05, 0.07]) and under-representation of predicates (mean = -0.05, 95% CI [-0.06, -0.04]) and closed-class items (mean = -0.16, 95% CI [-0.17, -0.14]). In comparison to social words in receptive vocabulary, social words in productive vocabulary exhibit a more noticeable S-shaped curve, indicating over-representation at lower vocabularies and under-representation at high vocabularies, resulting again in under-representation overall (mean = -0.02, 95% CI [-0.03, -0.01]).

Taken together, these patterns indicate that there is a persistent noun bias in children's reported vocabulary, and that children's relative knowledge of predicates and closed-class items grows with vocabulary size.

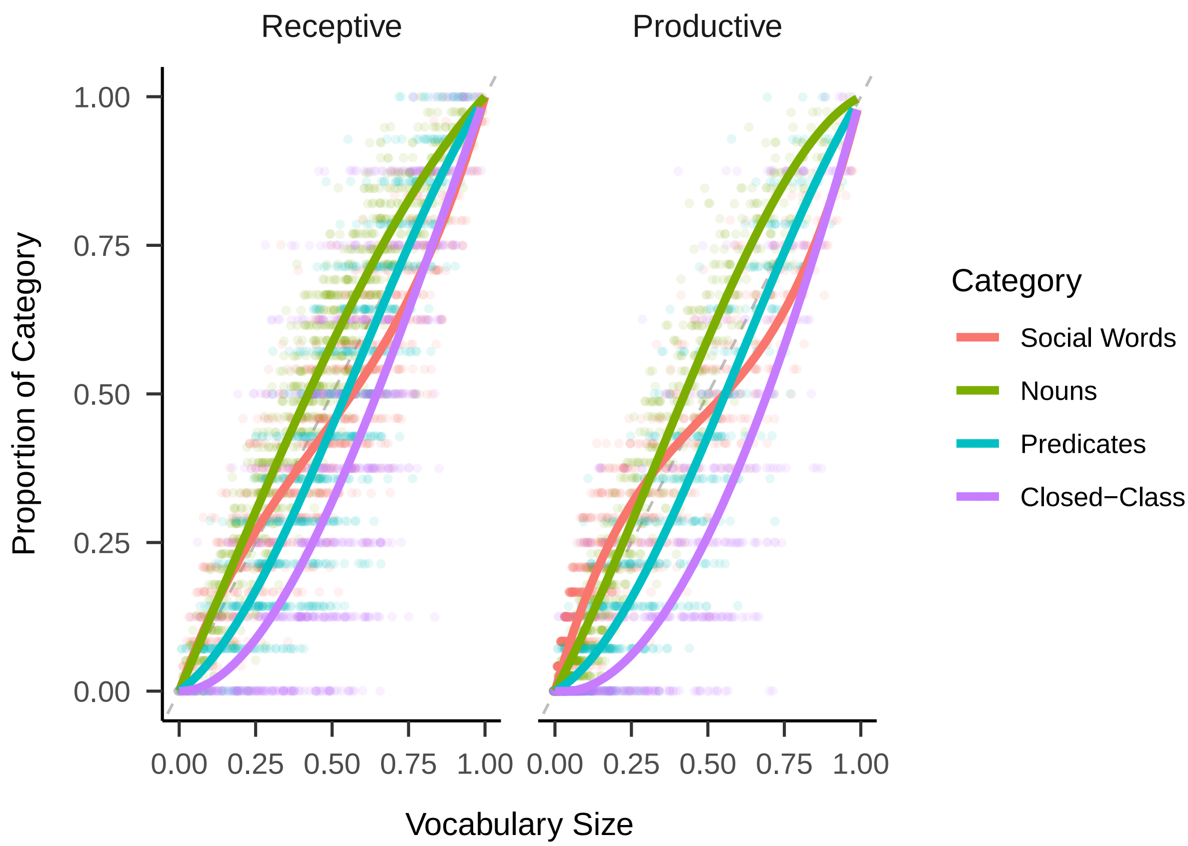

Figure 5. Changes in the syntactic composition of children's receptive (left) and productive (right) conceptual vocabulary as a function of growing total vocabulary size. There is consistent over-representation of nouns and under-representation of predicates and closed-class items. Social words are over-represented at lower vocabularies, but under-represented at higher vocabularies.

Discussion

In this study, we examined trends in vocabulary growth in a large sample of children growing up in Ghana using a caregiver reported vocabulary checklist. Caregiver reports are practical as means of assessing language development in low- and middle-income countries. They require little training to administer, do not require private or quiet spaces, and can be delivered verbally to caregivers in communities with low rates of literacy. Short-form methods are particularly useful, as they are quick and easier to implement at a large scale, while also providing a detailed snapshot of children's early vocabularies. This checklist was designed as part of a battery to measure the downstream effects of a high school education intervention parents participated in on their children's development. We will first discuss the psychometric properties of the checklist, focusing on its effectiveness for tracking children's language. Next, we will review significant predictors of children's vocabulary growth. Finally, we will examine changes in the composition of children's vocabulary across development.

Establishing validity and reliability

We found evidence that the checklist was effective at tracking individual differences in vocabulary size in the sample. Caregiver responses exhibited high internal consistency and person separability, indicating good reliability. Reported vocabulary size grew with child age, and receptive and productive vocabulary were highly correlated. The vast majority of items on the checklist were effective at distinguishing low and high-vocabulary children. In sum these findings provide support for the reliability and validity of the checklist for measuring children's vocabulary in the population. There were a few items on the checklist that were either under or over discriminating. These words were mostly in productive vocabulary, and included some social words, names for body parts, and common items found in eating spaces. Future research might seek out different translations for these terms, replace them with others similar words, or omit them entirely. Social words in particular have been reported as relatively less discriminating across different CDIs, possibly because caregivers use laxer criteria in determining whether their children know these words (Frank et al., 2021). Removing these words or replacing them with other items might improve the ability of the checklist to accurately measure children's language ability. However, eliminating them also risks distorting our understanding of word learning and vocabulary composition. For instance, social words follow different developmental trajectories than other categories, being acquired earlier than common nouns (Bates et al., 1994; Caselli et al., 1995; Caselli et al., 1999). In our own analysis of syntactic composition, social words followed an expected developmental trajectory, being over-represented in smaller vocabularies and under-represented in larger vocabularies. Thus, it may be desirable for some researchers to ask about these words regardless of fit.

Demographic predictors of vocabulary growth

The present study also explored several potential demographic predictors of vocabulary growth. For both receptive and productive vocabulary, we found that girls and children from multilingual homes had higher conceptual vocabularies on average than boys and children from monolingual homes. Girls have been observed to have an early advantage in word learning across a multitude of languages and cultures (Huttenlocher et al., 1991; Bornstein et al., 2004; Frank et al, 2021). This advantage has also been documented in various African languages (Vogt et al., 2015; Weber et al., 2018; Southwood et al., 2021).

We also found that caregiver reported activities with their children were positively related to conceptual vocabulary. Caregivers were surveyed on items selected from the Multiple Indicators Cluster Survey, which have been related to children's cognitive development, including their receptive and productive language, across dozens of different countries around the world (Jeong et al., 2016; Cuartas et al., 2020). A common explanation is that these activities afford children the opportunity to learn from caregiver input. Children's vocabulary as measured by caregiver checklist has been previously associated with caregiver input produced during e.g., play, reading, and object-naming (Salo et al., 2016). Parents who engage children in these activities might also attend more closely to children's vocabulary growth, allowing them to more fully report on their children's total vocabulary. Another possibility is that the pace of children's cognitive development and their elicitation of stimulating activities from their caregivers might each be influenced by the same genetic influences. Starr et al. (2024) found that longitudinal associations between caregiver reports of stimulation and cognitive development in a sample of 15,000 twins growing up in the U.K. were not causal in nature but instead were explained by genetic and shared environmental factors.

In contrast, our finding that the children from multilingual households have larger conceptual vocabularies than the children from monolingual households is surprising and highlights the degree to which the relationship between multilingualism and early word learning varies across cultural, socioeconomic, and linguistic contexts. Studies that focus on total vocabulary size (the sum of both languages) sometimes find that young bilingual children know more words than monolingual children (see e.g., Byer-Heinlien et al., 2024) and sometimes find the opposite (see e.g., Cote & Bornstein, 2014). However, studies that measure conceptual vocabulary size (the total number of concepts for which the child has a word, calculated by only counting translation equivalents once) either find no difference between the bilinguals and monolinguals (see e.g., Byers-Heinlein et al., 2024 for comprehension; Junker & Stockman, 2002) or find that the bilinguals have smaller conceptual vocabularies than the monolinguals (see e.g., Byers-Heinlein et al., 2024 for production; Core et al., 2013).

What might account for the larger conceptual vocabulary size of the multilingual children in the sample? It seems unlikely that this is a direct effect of multilingualism per se. To learn a word you must hear it, and the multilingual child necessarily splits their time between two languages, typically resulting in slower growth in each individual language. Under some circumstances total vocabulary may grow faster in multilingual homes because it is easier to learn words for some concepts than for others (see e.g., Coffey & Snedeker, 2025) and the child has the opportunity to expand their vocabulary by learning two words for these concepts (rather than having to acquire the harder ones). There is, however, no current theory that predicts that multilingualism should result in a larger conceptual vocabulary. Presumably this could happen if different words were difficult to produce in each language or when the early vocabulary in each language (or culture) has less overlap. We do not believe that either of these factors is accounting for this pattern in the current data set: we see parallel effects in comprehension and production, the lexical items were chosen because they were believed to have translation equivalents that would be relevant to a child in the languages spoken by these families, and there is no reason to believe that the language pairs spoken by these children would be more likely to show these effects than the other language pairs that have been studied.

This effect might also be an artifact of how the tool was designed and deployed. Caregivers were administered the checklist verbally by the surveyor in the language they were most comfortable conversing in. However, the checklist text the surveyors read from was written out in English. Thus, if surveyors defaulted to English during the administration of the survey, it may have resulted in more reporting from families who used English more frequently. If this were true, we would expect more noise in the data set for children whose families didn't speak English at home. While we did not find the language of the checklist was associated with overall vocabulary size, there may still be noise at the item level. To test this, we reexamined the reliability and validity of the checklist, dividing the sample into two subgroups of children who had some exposure to English ($n = 416$) and children who had none ($n = 521$). We found no differences in internal consistency or person separability between the groups. We found 7 additional items that exhibited poor fit in the sample with no English exposure (5 in productive vocabulary, 2 in receptive vocabulary). Each of these belonged to the same categories as the poorly fitting items in our original analysis (i.e., kinship: *uncle*; body parts: *face*, *hand*; kitchen items or furniture: *chair*, *plate*; games and routines: *hide and seek*), with the exception of *seesaw* (outside words). We also found 3 additional items that exhibited poor fit in the English exposed sample, all in receptive vocabulary. As with the non-English exposed sample, these words belonged to the same categories as the poorly fitting items in the combined analysis (body parts: *leg*; games and routines: *bye bye*) with the exception of an outside word *water*. Notably, none of the additional poorly fitting words are shared across English and non-English exposed groups. It is possible that a small amount of additional noise was introduced to productive conceptual vocabulary size in non-English-speaking households compared to English-speaking homes. However, the consistency of the effect across both productive and receptive conceptual vocabulary measures, as well as the lack of association with checklist language translation, suggests that this did not account for the affect of multilingualism.

A second possibility is that these effects arise from an association between multilingualism and other social or economic factors that affect word learning (or parental reports). For example, Vogt et al. (2015) report larger vocabularies in urban areas, where bilingualism is more common than in rural areas. The bilinguals in their sample were more often exposed to Portuguese, which is also the language of government and education in Mozambique. Around three-fourths of multilingual children in our sample were exposed to English as a second or third language, which could also be correlated with being in a more urban or educated household. We conducted a chi-squared test to determine whether there was an association between these variables and English language use in the home. As with our reanalysis of reliability and validity, we split children into two groups by their reported exposure to English. We found no association between urbanness and exposure to English ($\chi^2(1) = 2.29$, $p = 0.13$), but we found a significant association between primary caregiver education and English exposure in the home ($\chi^2(1) = 89.96$, $p < .001$), such that primary caregivers reporting English language use were more likely to have post-secondary education (207 vs. 105 with no English) and less likely to have no high school education (209 vs. 416 with no English).

Despite this association, we found no main effect of caregiver education on vocabulary. Still, it is possible English language use may index other differences in the home environment relevant to children's language development. For instance, when schooling is conducted in English, pedagogical activities may become associated with that language, such that they are more likely to occur if English is used in the home. Thus, English use might be a proxy for the presence of activities that positively predict vocabulary growth. We conducted t-tests comparing the frequency of caregiver-child activities between children with and without English language exposure. We found that households with English language use are more likely to report engaging in these activities with children (3.38 vs. 2.59 activities) ($t = 9.99$, $p < .001$). The effects of multilingualism on vocabulary may result in part from the association between English usage and caregiver-child interaction. However, for both receptive and productive conceptual vocabulary, multilingualism remained significant even after the introduction of stimulating activities into each respective model. Additionally, we conducted an exploratory analysis fitting our same models to the data set only including children who were exposed to Akan and/or English. We found that multilingualism was no longer a significant predictor of productive conceptual vocabulary but remained a significant predictor of receptive vocabulary. These findings suggest other unmeasured factors might also be at play.

We did not find a significant main effect of caregiver education on vocabulary size. Household education, particularly maternal education, has been widely found to be associated with differences in early language development in studies conducted in North America, Europe and East Asia (Frank et al., 2021). In an U.S. context, education has been associated with increased willingness and opportunity to engage in conversation and other forms of verbal engagement with children, which has been found to mediate the relationship between language development and socioeconomic status broadly (e.g., Huttenlocher et al., 1991; 2010; Hoff, 2003; Rowe, 2008). In studies using the parental reports in African countries, the pattern is less clear, with some studies reporting positive effects of maternal education or educational interventions (Vogt et al., 2015; Weber et al., 2017), and some reporting none (Southwood et al., 2021).

It is possible that effects of education and income were not observable due to limited variation within the sample. One of the parents of each of the infant participants had been selected for the study because 1) they were from a low-income household in a rural area of southern or central Ghana, and 2) they had qualified for high-school admission based on their performance on the entrance exam but failed to enroll in high school by the deadline due to financial constraints. As a result, the vast majority of caregivers reported having at least finished junior high school with 9 years of education (n = 835, or 89%). By comparison, Vogt et al. (2015) followed households in Mozambique that reported mostly either primary schooling or no formal schooling. There may not be substantial differences in the home environment on the basis of education within this range in our sample.

We also found no evidence of differences in vocabulary by birth order. Previous CDI studies in North America, Europe and East Asia, have generally found that firstborns have larger productive vocabularies than laterborns but not larger receptive vocabularies (Frank et al., 2021). Some studies have linked these effects to differences in caregiver input, finding that parents direct longer utterances to firstborn children compared to laterborn children (Hoff-Ginsberg, 1998; Huttenlocher et al., 2010). Firstborn children are also more likely to receive one-on-one attention from caregivers than laterborn children, who must share their parents' attention with their siblings. None of the African-adapted CDIs examined birth order as a predictor of vocabulary. One possible explanation for these findings is that firstborn status may not result in more input from adult caregivers. For instance, smaller intervals between firstborn and laterborn children might result in a shorter period where firstborns receive individualized attention: firstborns may need to compete for attention earlier than children in other countries. This seems unlikely in the dataset: the median age gap between siblings in the data set is about 3 years, which is well in-line with norms in the U.S. (Martinez & Daniels, 2023). Other studies in other African communities have found that young children spend much of their daily lives supervised by older child caregivers, who talk less and produce less complex speech than adult caregivers (Harkness, 1977; Nwokah, 1987; Loukatou et al., 2022). Vogt et al. (2015) found that children in Mozambique who have a child secondary caregiver have lower vocabularies than children with another adult caregiver. Therefore, it is possible that birth order may not affect the amount of speech children hear from adults.

Lastly, we found no association between living in an urban area and reported vocabulary. Prior results using parent-reported vocabulary measures have been mixed (Vogt et al., 2015; Southwood et al., 2021; Ma et al., 2024), which appears to reflect the different contexts in which urbanization occurs and the different effects it has on parenting. It is possible that there are differences that are typically present in urban vs. rural samples that were eliminated in our study because the parents were drawn from a highly uniform sample of adolescents from similar regions and socioeconomic strata.

Noun bias in cross-linguistic studies of word learning

There has been considerable theoretical interest in which words children learn first and why and how this changes across languages and cultural contexts. To explore these patterns, we constructed predictive models to analyze trends in the syntactic composition of children's vocabulary. We found that the children in the sample had a noun bias in both reported vocabularies. Predicates and closed-class items were also under-represented in children's vocabularies. Social words were more likely to be over-represented early but were under-represented later in development, especially in productive vocabulary. This pattern parallels the shifts in vocabulary composition that have been observed in North America, Europe and many parts of East Asia (Braginsky et al., 2019). While the noun bias is clearly absent in some studies and some languages (see e.g., Tardiff et al., 1999; Tardif et al., 2009; Childers et al., 2007; Casillas et al., 2024), our findings demonstrate that it is not limited to Western cultures but instead may reflect deeper differences in the concepts that words encode, the roles that they play in sentences, and how children isolate their forms and meanings (see e.g., Gentner, 1982; Gillette et al., 1999; Coffey & Snedeker, 2025).

Limitations

The current validation of the checklist has limitations that could be addressed in future work. Due to the age and language background of the children in the sample, there was no gold-standard, standardized test against which we could standardize our measure.

Future work, however, could compare performance on the vocabulary checklist to other caregiver-reported measures, such as speech diaries or parent's spontaneous recall (e.g., Alcock et al., 2015). Our study also provides no direct evidence of the predictive validity of our measure. A previous analysis found that reported receptive vocabulary predicted performance on a novel receptive picture vocabulary test administered to the same children at 3 years (Coffey & Spelke, 2024). This vocabulary test was also created to evaluate the effects of the intervention; thus, as with concurrent validity, we are unable to make comparisons to other existing standardized measures. Because vocabulary knowledge in late infancy is correlated with later language and literacy skills in school, checklists such as these are especially valuable to educators and policymakers for their predictive power. Such patterns have also been found in Ghana, with Prado et al. (2018) finding that their checklist predicted later performance on the IDELA (Pisani et al., 2018).

Another limitation comes from the lack of survey information about language exposure. Although caregivers were asked to report on languages children were exposed to, no measure of children's relative exposure to different languages was collected. It is unclear for instance whether a caregiver reporting Akan, English, and Ewe exposure uses these languages equally often, whether the child has relatives who occasionally visit and use one or more of these languages, or if caregivers only use particular words from some of these languages in the context of code-switching. In an environment like Ghana, children may be exposed to up to 6 languages in daily life (Omane et al., 2023). In a U.S. context, multilingualism is often defined in terms of the proportion of the child's speech exposure coming from each language, with 25% or 30% exposure to another language often being used as a cutoff for distinguishing monolinguals from multilinguals (e.g. Core et al., 2014; Byers-Heinlein et al., 2024). We believed that it would be difficult for caregivers to accurately report exposure proportions. Furthermore, there was not sufficient time for surveyors to interview parents about the child's vocabulary in each language. For these reasons, we focused our surveys on capturing the child's total conceptual vocabulary.

While studies of relative input and vocabulary sizes within each language provide critical insight into language acquisition in multilingual (Pan, 2011; Hoff & Luz Rumiche, 2011), assessing total conceptual vocabulary was sufficient for our goal of characterizing individual differences in the pace of language development.

Our limited knowledge about the children's input is particularly relevant to interpreting the finding that multilinguals had larger conceptual vocabularies than their monolingual peers. We cannot directly ascertain which languages contributed to this effect because we did not directly assess vocabulary in each language. Because this effect has not been previously reported, we believe that it reflects factors that are correlated with multilingualism in this environment rather than being a direct consequence of multilingual exposure. For example, we know that caregiver use of English (the most common second language) is related to their engagement in stimulating activities with their children.

We also failed to locate any prior diary studies of early vocabulary acquisition in this context on which to base the selection of words. As a result, the checklist was adapted directly from the CDI-WG short form checklist. It is possible then that the checklist missed words that might be more important to children's acquisition. These include parts of speech that might appear in Akan or other languages that would not appear in English. The checklist might also have missed more culturally appropriate terms within the 17 semantic categories selected.

Finally, there is also a need to establish the ecological validity of the checklist, the degree to which what caregivers report on the checklist generalizes to children's real life deployment of their language skills. There are several reasons for questioning the ecological validity of caregiver reports in research on cultural variation in patterns of engagement between parents and their children. Caregiver reports assume that the child spends most of their time with an adult caregiver who tracks the child's developmentally relevant behaviors. This assumption is fairly accurate in Western households, where children spend most of their day with their parents and are spoken to from birth. However, it has been observed in many African cultures that parents do not speak to children until the children begin to speak themselves, and preschool-aged children may spend most of their time away from their parents with siblings or peers (Harkness, 1977; Nwokah, 1987; Loukatou et al., 2022). Thus, it is important to establish the relationship between caregiver reports and other, more ecologically valid, measures of child language use. One way this has been done in previous studies is by comparing the checklist to children's naturalistic production of language (Alcock et al., 2015; Vogt et al., 2015; Prado et al., 2018; Weber et al., 2018). The children in the present study were filmed during a short 5-minute play session with their primary caregiver, which was transcribed by a trained surveyor. Thus, there is a path to establishing the ecological validity of the parent report measure in the future.

Acknowledging these limitations is critical for ensuring the ethical use of this instrument. Assessments with poor sensitivity can underestimate the effects of interventions and fail to identify children with language impairments. Assessments that are not well adapted for the culture or language of the participants can reinforce stereotypes that all African children suffer from widespread cognitive delay. Although field research in African communities, particularly rural communities, comes with unique challenges that are not encountered in most research in North America and Europe, it is essential that researchers work closely with local communities to create ecologically appropriate measures and that these measures are subject to formal evaluations of their validity and reliability. We are uniquely positioned to provide such an evaluation. Of the six African checklist adaptations we located in our review, only Prado et al. (2018) reports a sample of similar size to our own. Because the families in our study came from diverse regions, languages and levels of education, there are good reasons to believe that these findings might generalize to other similar populations. Our finding that the language of the survey did not affect reported vocabulary is clearly relevant to future groups studying language development in Ghana and similar settings.

Conclusion

We examined trends in vocabulary growth using a short-form caregiver reported vocabulary checklist adapted for use with 14- to 25-month-old children in Ghana. We find a small but reliable effect of gender on vocabulary size, such that girls had larger vocabularies than boys on average. We also find that children whose parents reported their children being exposed to multiple languages also reported higher vocabularies. This association is linked to a relationship between English exposure and the frequency of reading, play, and other activities that have been found to support language development. No association is found between vocabulary size and education, suggesting a more complex relationship between parental education and early child development than has previously been posited. Finally, we find that the syntactic composition of children's vocabulary changed in predictable ways throughout development, supporting theories that attribute these shifts to universal features of language learning that generalize across contexts and languages (Gentner, 1982; Snedeker & Gleitman, 2004; Coffey & Snedeker, 2024). We also suggest reasonable guidelines for evaluating the validity and reliability of vocabulary assessments developed for use in contexts where such measures are difficult to develop, such as in studies with wide age ranges or studies conducted in highly multilingual societies.

References

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. Journal of Child Language, 42(4), 763-785.

Apenteng, M.A., & Amfo, N.A.A. (2014). The form and function of English loanwords in Akan. Nordic Journal of African Studies, 23(4), 22-22.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., . . . Hartung,

J. (1994). Developmental and stylistic variation in the composition of early vocabulary. Journal of Child Language, 21(1), 85-123.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48.

Bialystok, E., Luk, G., Peets, K. F., & Sujin, Y. A. N. G. (2010). Receptive vocabulary differences in monolingual and bilingual children. Bilingualism: Language and cognition, 13(4), 525-531.

Black, M.M., Walker, S.P., Fernald, L.C., Andersen, C.T., DiGirolamo, A.M., Lu, C., McCoy, D.C., … Devercelli, A.E. (2017). Early childhood development coming of age: science through the life course. The Lancet, 389(10064), 77–90.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., Pêcheux, M.G., … Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. Child development, 75(4), 1115-1139.

Bornstein, M. H., & Cote, L. R. (2005). Expressive vocabulary in language learners from two ecological settings in three language communities. Infancy, 7(3), 299-316.

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Suwalsky, J. T. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. Child Development, 85(4), 1346-1356.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. Open Mind, 3, 52-67.

Byers-Heinlein, K., Gonzalez-Barrero, A. M., Schott, E., & Killam, H. (2024). Sometimes larger, sometimes smaller: Measuring vocabulary in monolingual and bilingual infants and toddlers. First Language, 44(1), 74-95.

Can, D. D., Ginsburg-Block, M., Golinkoff, R. M., & Hirsh-Pasek, K. (2013). A long-term predictive validity study: Can the CDI Short Form be used to predict language and early literacy skills four years later?. Journal of Child Language, 40(4), 821-835.

Canty, A., & Ripley, B. D. (2021). boot: Bootstrap R (S-Plus) Functions.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. Cognitive Development, 10(2), 159-199.

Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. Journal of child language, 26(1), 69-111.

Casillas, M., Foushee, R., Méndez Girón, J., Polian, G., & Brown, P. (2024). Little evidence for a noun bias in Tseltal spontaneous speech. First Language, 44(6), 600-628.

Childers, J. B., Vaughan, J., & Burquest, D. A. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. Journal of Child Language, 34(2), 199-225.

Coffey, J. R., & Snedeker, J. (2024). Disentangling the roles of age and knowledge in early language acquisition: a fine-grained analysis of the vocabularies of infant and child language learners. Cognitive Psychology, 153, 101681.

Coffey, J., & Spelke, E. (2024). Direct cognitive testing in the field at scale: leveraging core knowledge to assess the development of young children in Ghana. psyArxiv. Accessed November 1, 2024.

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: Implications for assessment. Journal of Speech, Language, and Hearing Research, 56(5), 1637-16.

Cote, L. R., & Bornstein, M. H. (2014). Productive vocabulary among three groups of bilingual American children: Comparison and prediction. First Language, 34(6), 467-485.

Cuartas, J., Rey-Guerra, C., McCoy, D. C., & Hanno, E. (2020). Maternal knowledge, stimulation, and early childhood development in low-income families in

Colombia. Infancy, 25(5), 526-534.

Dupas, P., Falezan, C., Jayachandran, S., & Walsh, M. P. (2023). Informing mothers about the benefits of conversing with infants: Experimental evidence from Ghana. NBER Working Paper Series, 31264.

Duflo, E., Dupas, P., & Kremer, M. (2021). The Impact of Free Secondary Education: Experimental Evidence from Ghana. NBER Working Paper Series, 28937.

Duflo, E., Dupas, P., Spelke, E., & Walsh, M. P. (2024). Intergenerational impacts of secondary education: experimental evidence from Ghana. NBER Working Paper Series, 32742.

Durán, L. K., Wackerle-Hollman, A. K., Kohlmeier, T. L., Brunner, S. K., Palma, J., & Callard, C. H. (2019). Individual growth and development indicators-Español: Innovation in the development of Spanish oral language general outcome measures. Early Childhood Research Quarterly, 48, 155-172.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C.B., & Stiles, J. (1994). Variability in early communicative development. Monographs of the society for research in child development, i-185.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. Applied psycholinguistics, 21(1), 95-116.

Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.) Language development, Vol. 2: Language, thought and culture, (pp. 301-334). Lawrence Erlbaum Associates Inc.

Gentner, D. & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S.C. Levinson (Eds.) Language Acquisition and Conceptual Development (pp. 215-256). Cambridge University Press.

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 medicine, 7(5), e1000273.

Harkness, S. (1977). Aspects of social environment and first language acquisition in rural Africa. In Catherine E. Snow & Charles A. Ferguson (eds.), Talking to Children. Cambridge University Press. pp. 309-316.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. Applied psycholinguistics, 19(4), 603-629.

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. Child Development, 74(5), 1368–1378.

Hoff, E., & Luz Rumiche, R. (2011). Studying Children in Bilingual Environments. In Erika Hoff (ed.), Research Methods in Child Language, 300–316. Hoboken, New Jersey: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444344035.ch20.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. Developmental psychology, 27(2), 236.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. Cognitive Psychology, 61(4), 343–365.

Jeong, J., McCoy, D. C., & Fink, G. (2017). Pathways between paternal and maternal education, caregivers' support for learning, and early child development in 44 low-and middle-income countries. Early Childhood Research Quarterly, 41, 136-148.

Kachergis, G., Francis, N., & Frank, M. C. (2023). Estimating demographic bias on tests of children's early vocabulary. Topics in Cognitive Science, 15(2), 303-314.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? First Language, 42(6), 703–735.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. Applied Psycholinguistics, 32(1), 69-92.

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2022). Child-directed and overheard input from different speakers in two distinct cultures. Journal of Child Language, 49(6), 1173-1192.

Macours, K., Williams, J., & Wolf, S. (2023). Measurement of holistic skills in RCTs: Review and guidelines. Abdul Latif Jameel Poverty Action Lab. https://www.povertyactionlab.org/publication/measurement-holistic-skills-rcts-review-and-guidelines

Mair, P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. Journal of Statistical Software, 20(9), 1–20.

Marchman, V. A., & Dale, P. S. (2023). The MacArthur-bates communicative development inventories: updates from the CDI advisory board. Frontiers in Psychology, 14, 1170303.

Martinez, G.M. and Daniels, K. (2023). Fertility of Men and Women Aged 15–49 in the United States. National Survey of Family Growth 2015–2019, 179.

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: a naturalistic observation study of Mozambican infants from 1;1 to 2;1. Journal of Child Language, 43(2), 235–264.

Nwokah, E. E. (1987). Maidese versus motherese—Is the language input of child and adult caregivers similar? Language and Speech, 30(3), 213-237.

Omane, P. O., Benders, T., Duah, R. A., & Natalie, B. A. (2023). Diversifying language acquisition research can be (partly) achieved in urban societies and with simplified methodologies: Insights from multilingual Ghana. Journal of Child Language, 50(3), 532-536.

Osam, E. K. (1993). The loss of the noun glass system in Akan. Acta linguistica hafniensia, 26(1), 81-106.

Osam, E.K. (2003). An Introduction to the Verbal and Multi-verbal systems of Akan. In Beermann, D. & Hellen, L. (eds.). Proceedings of the Workshop on Multi-verb Constructions. Trondheim Summer School 2003.

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. Child Development, 76(4), 763-782.

Pan, B.A. (2011). Assessing Vocabulary Skills. In Erika Hoff (ed.), Research methods in child language, 100–112. Hoboken, New Jersey: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444344035.ch7.

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). International Journal of Educational Research, 91, 1–15.

Prado, E.L., Abbeddou, S., Adu-Afarwuah, S., Arimond, M., Ashorn, P., Ashorn, U., Bendabenda, J., … Lartey, A. (2017). Predictors and pathways of language and motor development in four prospective cohorts of young children in Ghana, Malawi, and Burkina Faso. Journal of Child Psychology and Psychiatry, 58(11), 1264-1275.

Prado, E.L., Phuka, J., Ocansey, E., Maleta, K., Ashorn, P., Ashorn, U., Adu-Afarwuah, S., … 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, 1-10.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rogoff, B., Coppens, A. D., Alcalá, L., Aceves-Azuara, I., Ruvalcaba, O., López, A., & Dayton, A. (2017). Noticing learners' strengths through cultural research. Perspectives on Psychological Science, 12(5), 876-888.

Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. Journal of Child Language, 35(1), 185–205.

Sania, A., Sudfeld, C.R., Danaei, G., Fink, G., McCoy, D.C., Zhu, Z., Fawzi, M.C.S., ... Fawzi, W. (2019). Early life risk factors of motor, cognitive and language development: a pooled analysis of studies from low/middle-income countries. BMJ, 9(10), e026449.

Singh, L., Yeung, W. J. J., Cheng, Q., & Heng, E. Y. T. (2023). The home literacy environment mediates effects of socio-economic status on infant vocabulary development. Developmental Science, 26(4), e13349.

Southwood, F., White, M.J., Brookes, H., Pascoe, M., Ndhambi, M., Yalala, S., ... Alcock, K. (2021). Sociocultural factors affecting vocabulary development in young South African children. Frontiers in Psychology, 12, 642315.

Snedeker, J., & Gleitman, L. R. (2004). Why it is hard to label our concepts. In D. Geoffrey Hall & S. R. Waxman (Eds.), Weaving a lexicon (pp. 257–294). MIT Press.

Starr, A., Oginni, O., & von Stumm, S. (2024). Do children cause the cognitive stimulation they receive? Modelling the direction of causality. Behavior Genetics, 54(6), 443-455.

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., ... Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: a randomized controlled pilot study. Journal of Child

Language, 43(2), 366–406.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the “noun bias” in context: A comparison of English and Mandarin. Child Development, 70(3), 620-635.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. Journal of child language, 36(5), 1115-1144.

UNICEF. (2005). “Monitoring the situation of children and women: multiple indicator cluster survey manual.” New York: United Nations Children’s Fund.

Vogt, P., Mastin, J. D., & Aussems, S. (2015). Early vocabulary development in rural and urban Mozambique. Child Development Research, 2015(1), 189195.

Weber, A., Fernald, A., & Diop, Y. (2017). When Cultural Norms Discourage Talking to Babies: Effectiveness of a Parenting Program in Rural Senegal. Child Development, 88(5), 1513–1526.

Weber, A. M., Marchman, V. A., Yatma, D. I. O. P., & Fernald, A. (2018). Validity of caregiver-report measures of language skill for Wolof-learning infants and toddlers living in rural African villages. Journal of Child Language, 45(4), 939-958.

Weisz, J. R., Vaughn-Coaxum, R. A., Evans, S. C., Thomassin, K., Hersh, J., Ng, M. Y., . . . Mair, P. (2019). Efficient Monitoring of Treatment Response during Youth Psychotherapy: The Behavior and Feelings Survey. Journal of Clinical Child & Adolescent Psychology, 49(6), 737–751.

Wilson, M. (2005). Constructing measures: an item response modeling approach (2nd Ed.). New York: Routledge.

Wong, K., Thomas, C., & Boben, M. (2020). Providence talks: A citywide partnership to address early childhood language development. Studies in Educational Evaluation, 64, 100818.