Measuring Children's Early Vocabulary Across Languages Using a 'Swadesh' Word List

Anonymous CogSci submission

Abstract

Early language skill is predictive of later life outcomes, and is thus of great interest to developmental psychologists and clinicians. The Communicative Development Inventories (CDI), including a parent-reported inventory of early-learned vocabulary items, have proven to be valid and reliable instruments for measuring children's early language skill. CDIs have been painstakingly adapted to dozens of languages, and cross-linguistic comparisons thus far show both consistency and variability in language acquisition trajectories. However, hundreds of languages do not yet have CDIs, posing a significant barrier to increasing the diversity of languages that are studied. Here, we propose a method for selecting candidate words to include on new CDIs, leveraging analysis of psychometric properties of translation-equivalent concepts that are frequently included on existing CDIs. Leveraging 26 datasets from existing CDIs, we propose a list of 108 concepts that have low variability in their cross-linguistic learning difficulty. This pool of 'universal' concepts– analogous to the "Swadesh" lists used in glottochronology–can be used as a starting point for future CDI adaptations. We test how well the proposed list generalizes to data from 8 additional languages.

Keywords: early language learning; CDI; psychometrics; cross-linguistic comparison; Swadesh vocabulary;

Introduction

Tools that enable valid assessments of children's early language abilities are invaluable for researchers, clinicians and parents, as early language skill is predictive of educational outcomes years later (e.g., Bleses, Makransky, Dale, Højen, & Ari, 2016). The MacArthur-Bates Communicative Development Inventories [CDIs; Fenson et al. (2007); Marchman, Dale, & Fenson (2023)] are parent report assessments that provide reliable and valid estimates of children's early vocabulary size and other aspects of early communicative development, such as gesture use and early use of word combinations. Parent report is a relatively quick and low-cost method to assess early language skills since it takes advantage of the fact that parents are "natural observers" of their child's skills and does not depend on a child engaging with an (unfamiliar) experimenter.

Over the years, the CDIs have been adapted to dozens of languages, with forms now available in English, Spanish, French, Hebrew, and Mandarin, to name just a few. Recently, data from more than 85,000 CDIs in 38 languages have been archived in a central repository [Wordbank; Frank, Braginsky, Yurovsky, & Marchman (2017)]. These data have revealed both cross-linguistic consistency and variability in early language skills, with insights from these patterns informing theories of early language learning (Frank, Braginsky, Yurovsky, & Marchman, 2021). For example, cross-linguistic analyses indicate that measures of vocabulary size are tightly correlated with other aspects of early language skill, like gesture and grammatical competence. Thus, over development, the language system is "tightly woven" (Bates et al., 1994; Frank et al., 2021) and early vocabulary size serves as a good proxy measure of children's overall language skill.

On the CDIs, vocabulary size is assessed via a checklist format, which enables caregivers to scan and recognize words their child produces or understands, rather than relying on recall alone. For example, the American English CDI Words & Sentences (CDI:WS) form, targeting children 16-30 months of age, is comprised of 680 words from 22 semantic categories, including nouns (e.g., Body Parts, Toys, and Clothing), action words (e.g. verbs), descriptive words (e.g. adjectives), and closed-class words such as pronouns. The vocabulary checklist from the American English CDI Words & Gestures (CDI:WG) form, targeting younger children ages 8 to 18 months, is comprised of ˜400 words from a similar set of categories. Items on these original forms were chosen to reflect a range of difficulty levels (i.e., both easy, moderate, and more difficult), as well as capture the linguistic and societal contexts of (most) children living in the US. Short versions of each of these forms are also available, each with about 100 items (e.g., Fenson et al., 2000), consisting of a set of items that generate scores that more strongly correlate with scores on the long forms, while retaining representation across a broad set of semantic categories.

Creating a new CDI requires a lot of effort and resources, presenting a daunting barrier to increasing the diversity of languages studied. Following the guidelines[1] from the MacArthur-Bates CDI Advisory Board, the process of adapting a CDI for a language other than American English goes well beyond simply translating items on these forms to that new language. While the process can begin with identifying translation equivalents (i.e., items that capture the same general concept in both languages, e.g., "dog" in English, and "perro" in Spanish), the final item set must then be filtered so that all items appropriately reflect the linguistic and sociocultural context of the children learning that language. This

[1] `https://mb-cdi.stanford.edu/adaptations.htm`

process usually requires considerable time and effort by researchers who are native speakers of a language, to first select and identify translation equivalents and to then iteratively add, refine, and pilot the new CDI in the target language. Because the goal is to obtain the set of items that best capture general trends and individual differences in that language, the items across CDIs in different languages do not necessarily overlap to a great extent. For example, the American English CDI:WS and Mexican Spanish CDI:WS forms – two of the first CDIs created – each have 680 words, but only have 463 overlapping concepts (68%).

Given that it is well-established that, all over the world, early learned words reflect the people and things that children are likely to experience, that is, words for family members, animals, and common household objects (Frank et al., 2021; Tardif et al., 2008), it is reasonable to ask: Is there a single set of translation equivalents that would meet the criteria for inclusion on CDIs from multiple languages? To facilitate this effort, it is useful to leverage recent work using Item-Response Theory (IRT, Embretson & Reise, 2013) models. IRT models infer both the abilities of test takers and the difficulty of individual test items (i.e., words), along standardized dimensions. Recent work using IRT models have facilitated our understanding of the psychometric properties of CDI instruments. As such, they offer the potential to not only yield more accurate measures of children's language ability, but also to enable the construction of Computerized Adaptive Tests (CATs), which choose the next test item based on the responses to the previous items, and thus quickly hone in on the test-takers language ability. CAT-based CDIs presenting 50 or fewer items have been found to strongly correlate with scores on the full CDI:WS (Chai, Lo, & Mayor, 2020; Makransky, Dale, Havmose, & Bleses, 2016; Mayor & Mani, 2019). A general method for creating CDI CATs that work well across a broader age range (12-36 months) has been proposed, and tested for American English and Mexican Spanish (Kachergis, Marchman, Dale, Mankewitz, & Frank, 2022). However, the IRT model driving each CAT needs to be trained on a large and normative dataset, which may not be available in a given language. To date, the IRT models are also fitted separately for each language, and the fitted parameters (e.g., word difficulty) are likely to vary across languages.

The goal of the current study is to use IRT modeling in conjunction with data from Wordbank to examine whether there might be a core set of concepts that are frequently included on CDIs, and–importantly–whether many of them are of roughly equal difficulty across many languages. This work takes its inspiration from the fields of lexicostatistics and glottochronology, where researchers (notably, Swadesh, 1971) have proposed a list of 'universal concepts'–concepts that exist in all catalogued languages–in order to quantify the genealogical relatedness and dates of divergence of languages. For example, the original Swadesh list contains 100 words, comprised of categories including common pronouns ("I", "you", "we"), animals ("man", "fish", "bird", "dog"), objects ("tree", "leaf", "sun", "mountain"), and verbs ("kill", "die", "see", "sleep"). Extending this work to the development of a universal CDI, or "Swadesh CDI," would include all of the concepts that researchers have chosen to include on the vast majority of CDI:WS adaptations, and which have relatively similar difficulty across many languages.

More broadly, this work examines which types of words (and their corresponding concepts) are more or less similar across languages in terms of the ease with which they are learned, revealing commonalities as well as idiosyncrasies in children's early experiences. We can ask: Which semantic categories, e.g., animals, household objects, food and drink, or another category, are most consistently learned across languages? Which are more variable? These types of cross-linguistic comparisons may give new insight into theoretical questions surrounding the similarity and differences between language experience and development in different cultural and linguistic contexts.

In particular, our contributions are 1) to revise and extend a set of translation-equivalent concepts ("universal lemmas", i.e. uni-lemmas) in Wordbank, 2) to fit IRT models to 28 CDI:WS datasets, 3) to identify 108 candidate "Swadesh CDI" items from a cross-linguistic comparison of concept difficulty and inclusion, and 4) to test how well this Swadesh CDI generalizes to a set of 8 additional low-data languages. We end by making a concrete proposal for how this Swadesh CDI list could be used in creating future CDI adaptations, and by discussing the strengths and weaknesses of our approach.

Methods

Item Response Theory

A variety of IRT models targeting different types of testing scenarios have been proposed (see Baker, 2001 for an overview), but for the dichotomous responses that parents make for each item (word) regarding whether their child can produce a given word, we will use the popular 2-parameter logistic (2PL) model that is best justified for CDI data (see Kachergis et al., 2022).

The 2PL model jointly estimates for each child j a latent ability θ_j (here, language skill), and for each item i two parameters: the item's difficulty b_i and discrimination a_i, described below. In the 2PL model, the probability of child j producing a given item i is

$$P_i(x_i = 1|b_i, a_i, \theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

where D is a scaling parameter ($D = 1.702$) which makes the logistic more closely match the ogive function used in a standard factor analysis (Chalmers, 2012; Reckase, 2009). Children with high latent ability (θ) will be more likely to produce any given item than children with lower latent ability, and more difficult items will be produced by fewer children (at any given θ) than easier items. The discrimination (a_i) adjusts the slope of the logistic (in the classic 1-parameter logistic (1PL or Rasch) model, the slope is always 1). Items with higher discrimination (i.e. slopes) better distinguish children above vs. below that item's difficulty level, and hence are generally more useful. While other standard IRT models exist (e.g., the 3-parameter logistic model adds a 'guessing' parameter for each test item), a recent study found the 2PL model most appropriate for multiple Wordbank datasets (Kachergis et al., 2022).

Datasets

Language	items	N
Norwegian	731	9304
English (American)	680	8828
Danish	725	3714
Portuguese (European)	639	3012
Turkish	711	2422
Mandarin (Taiwanese)	696	1897
Spanish (Mexican)	680	1853
English (Australian)	558	1520
French (French)	690	1410
Korean	641	1376
Cantonese	804	1295
German	588	1181
Slovak	609	1066
Mandarin (Beijing)	799	1056
Russian	728	1037
French (Quebecois)	664	929
Swedish	710	900
Spanish (Argentinian)	699	784
Italian	670	752
Spanish (European)	588	593
Hebrew	605	518
Latvian	723	500
Czech	553	493
Croatian	717	377
Hungarian	802	363
Dutch	704	303
Greek (Cypriot)	815	176
Spanish (Peruvian)	600	105
Kigiriama	696	100
English (Irish)	660	99
Irish	691	99
Kiswahili	705	90
Finnish	581	70
Persian	558	50

Table 1: Number of CDI:WS items and subjects (N) per dataset.

We report IRT analyses for twenty-six languages from Wordbank (Frank et al., 2017), comprising production data from CDI:WS vocabulary checklists that have at least 200 administrations.[2] Data from the first twenty-six rows of Table 1 (Norwegian through Dutch) will be used to select a pool of words with approximately equal cross-linguistic difficulty. CDI:WS production data from an additional eight languages (bottom of Table 1: Greek (Cypriot) through Persian) had too few participants to be analyzed with IRT. Datasets for these languages will be used to test how well the selected pool of words generalize can be expected to new languages.

Uni-lemmas In order to do so, we need a method to map between words across languages that correspond to broadly similar concepts. As such, we mapped each item on the CDI:WS for each language onto a set of "universal lemmas" or "unilemmas", which are approximate cross-linguistic conceptual mappings of words. For example, "dog" (English) and "perro" (Spanish) both correspond to the same unilemma, dog. These mappings were recently updated for Wordbank 2.0 to improve their quality and systematicity, and to increase coverage across items and languages. This new set of unilemmas was constructed based on glosses provided by the original contributors of the Wordbank datasets, which were then verified by native or advanced proficient speakers of the language, and cleaned to increase their consistency across languages. All unilemmas are accessible from Wordbank; details about the recent update can be found at `https://github.com/langcog/update_unilemmas`.

Participants The CDI:WS production dataset consists of the combined Wordbank production data for 48272 children aged 16-30 months on 23020 items across 34 forms. Note that the distributions of demographic variables (age, sex, maternal education, etc.) of these datasets are not matched, so comparing overall language ability estimates across languages would be impossible. (See Frank et al. (2021) for a discussion of effects of demographic variables on vocabulary development.) Thus, we will focus only on the estimated item parameters, and in particular the variability of item difficulty (b_i).

Instruments When a CDI:WS forms was administered, caregivers were asked to indicate for each vocabulary item on the instrument whether or not their child can recognizably produce (say) the given word.

"Produces" responses were coded as 1 and all other responses were coded as 0. Our datasets consist of a dichotomous-valued response matrix for each language, of size N subjects $\times$ W words.

Results

All models, data, code for reproducing this paper are available on OSF[3]. Across the 26 IRT models for different CDI:WS forms, parameters for a total of 17715 items were obtained.

Figure 1 shows the average cross-linguistic difficulty of CDI items by semantic category. Sounds (e.g., animal noises), body parts, and common nouns tend to be early-learned, and thus have lower difficulty values, while abstract words such as time words and morphologically complex help-

[2] `http://wordbank.stanford.edu/contributors`

[3] OSF repository: https://osf.io/8swhb/.

ing verbs are later-learned (i.e., more difficult).

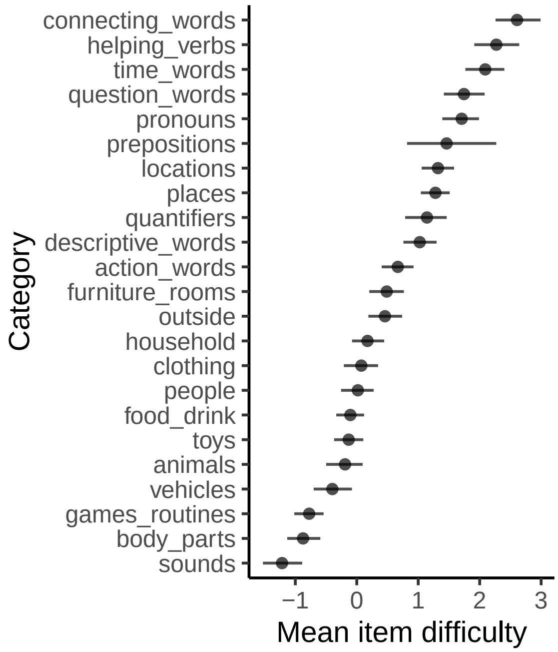

Figure 1: Mean cross-linguistic difficulty of CDI words by semantic category. Bars represent bootstrapped 95% confidence intervals.

Identifying Swadesh CDI Candidates CDI:WS forms have a median of 693 uni-lemmas defined (range: 553 (Czech) to 804 (Cantonese)), and there are a total of 1840 uni-lemmas defined across the 26 languages. Only 61 uni-lemmas appear on all 26 CDI:WS forms: too few to comprise a short form. To expand the pool of items, we consider thresholds on both 1) the variability of an item's difficulty across languages (lower is better), and 2) the number of CDI:WS forms on which it appears (more is better). Below we examine the standard deviation of uni-lemmas' cross-linguistic difficulty as a function of how many languages that uni-lemma is missing from. For now we consider items with less than median variability in difficulty that are included on 10 or more of the 26 CDI:WS forms.

Of these, a total of 1312 uni-lemmas are included on more than one form, and only 656 uni-lemmas are included on 10 or more of the CDIs. This restricted list is used as a starting point for selecting the Swadesh CDI candidates.

The next criterion for Swadesh candidates is to find uni-lemmas of consistent difficulty across languages: if an item has relatively similar difficulty across 10 or more languages, we expect that it is more likely to have similar difficulty in a new language, compared to an item that has more variability. To evaluate how variable uni-lemmas are in their cross-linguistic difficulty, we calculate the standard deviation (SD) of each uni-lemma's difficulty. The median SD is 1.2 (SD=0.42), so we consider the items with SD less than

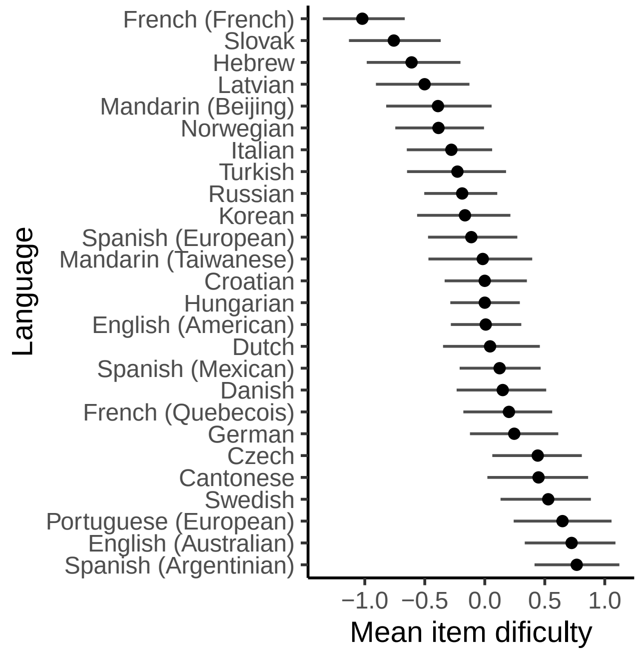

Figure 2: Mean item difficulty of Swadesh CDI items per language. Bars represent bootstrapped 95% confidence intervals.

half the median SD (i.e., SD < 0.99). These 108 candidate "Swadesh CDI" items are shown in Figure 3, grouped by semantic category. The Swadesh CDI candidates represent 17 of the 22 semantic categories on the original American English CDI:WS form: the five categories not represented are toys, pronouns, question words, quantifiers, and helping verbs. 54% of the Swadesh CDI words are nouns, 20% are predicates, 5% are function words, and 21% are other. This breakdown is comparable to the lexical category percentages on the 680-item English CDI:WS (46% nouns, 24% predicates, 15% function words, and 5% other), minimally suggesting that this method of selecting candidate items is not biased by lexical category. The Swadesh candidates were also present on more forms than typical in the selection set: on average, each item appeared on 20 forms.

Comparing the IRT parameters of the Swadesh CDI uni-lemmas to the rest of the items showed that the discrimination parameter (i.e., slope) of the Swadesh items did not significantly differ from the others. However, the candidate Swadesh items were significantly easier than the other uni-lemmas (mean Swadesh \$d=\$0, other items' mean \$d=\$0.5, $t(3123) = 15.63$, $p < .001$). It is worth noting that three of the five semantic categories not appearing on the Swadesh list are among the most difficult: helping verbs, question words, and pronouns. It may be prudent to consider expanding the Swadesh list to include some of these more difficult (and variable) items. Returning to the list of uni-lemmas that appear on at least four forms, we identified 10 uni-lemmas that cover the unrepresented difficult categories, and still fall below the variability threshold. The mean difficulty of these items is 3.13, far higher than that of other uni-lemmas. We will test the Swadesh candidates with and without this *difficult* list.

Comparing Swadesh CDI to Full CDI:WS

How well do raw scores from the Swadesh CDI items correlate with full CDI:WS scores? On average, for the 26 languages the IRT analysis was based on, the Swadesh CDI's sumscores were strongly related to the full CDI:WS scores (mean $r = 0.990$; $min = 0.976$, $max = 0.996$; full table on OSF). However, these correlations were slightly but significantly lower than scores based on randomly-selected items (mean $r = 0.993$, $min = 0.990$, $max = 0.996$, paired $t(25) = 5.67$, $p < .001$). This is likely due to some ceiling effects for older children on the Swadesh CDI, since the randomly-selected items are likely somewhat easier on average than the Swadesh items (see above results).

Adding the 10 difficult uni-lemmas improved the relationship (mean $r = .991$).

Testing Generality of the Swadesh CDI

For the eight low-data languages, a similar comparison revealed that the Swadesh CDI's raw scores were again strongly related to the full CDI:WS scores (mean $r = 0.978$; $min = 0.959$, $max = 0.992$). As with the training set, however, the Swadesh list correlations were slightly but significantly lower than scores based on randomly-selected items (mean $r = 0.985$; $min = 0.972$, $max = 0.992$; $t(7) = 3.22$, $p = .01$).

Adding the 10 difficult uni-lemmas improved the relationship (mean $r = .984$).

language	N	Rand r	Swadesh r
English (Irish)	82	0.989	0.984
Finnish	87	0.994	0.993
Greek (Cypriot)	77	0.988	0.975
Irish	85	0.989	0.980
Kigiriama	86	0.991	0.983
Kiswahili	83	0.989	0.969
Persian	60	0.977	0.960
Spanish (Peruvian)	68	0.990	0.981

Table 2: Generalization test results.

Actions	dance, drink, sing, blow, draw, swim, wait, catch, tickle, lick, stay, knock
Animals	cat, cow, *dog*, lion, animal, ant, horse, chicken, spider, goat, turkey
Body Parts	*nose*, *tongue*, *tooth*, head, lip, shoulder, penis, vagina
Clothing	button, necklace, belt, shorts, skirt
Connectives	and
Descriptives	blue, green, yellow, black, first, heavy, sticky, dirty, awake
Food/Drink	milk, banana, yogurt, peas, popcorn, potato chips, lollipop, hamburger
Furniture/ Rooms	door, washing machine, TV, drawer, garage
Games/ Routines	yes, shh, breakfast, shopping
Household	key, spoon, telephone, towel, broom, hammer, radio, toothbrush, brush, paper, glasses, nail, pacifier
Locations	in, under, out, behind, back
Outside	moon, *star*, sky, hose
People	mommy, daddy, doctor, pet's name, child's name, police, mailman
Places	school, church, gas station
Sounds	meow, cockadoodledoo, grrr, choo choo, yum yum
Time Words	yesterday, tonight, tomorrow
Vehicles	airplane, helicopter, firetruck

Figure 3: The 108 proposed Swadesh CDI words by semantic category.

Discussion

This study compared psychometric models fitted to 26 CDI datasets in order to find concepts that have low variability in their cross-linguistic difficulty, and that are frequently included on parent-reported measures of children's early vocabulary. We identified 108 concepts that appeared on at least 10 of the CDIs, and which have more consistent cross-linguistic difficulty than the majority of the concepts appearing on multiple CDIs. We showed that using this set of "Swadesh CDI" items as a short assessment would generate scores that are strongly related to full CDI:WS scores, both for the original 26 datasets, and in a generalization test to eight low-data languages. However, the Swadesh CDI items were also significantly easier than other items, meaning that older children may perform at ceiling if given only the Swadesh CDI items. (This may be unsurprising from the perspective that Swadesh words are meant to be universal, and are therefore more frequent and basic–both within and across individual children's experiences.)

Thus, our suggested use case for the Swadesh CDI list is as a starting point for researchers seeking to develop a CDI in a new language, rather than as a complete short-form CDI. In particular, researchers should seek to add contextually appropriate items from the categories that were notably absent from the Swadesh list: quantifiers, ..

Another potential limitation of this work is that most existing CDIs (and most datsets available in Wordbank) target languages in the Indo-European language family. It is not clear to what extent this bias in the existing data might interfere with generalizing to non-Indo-European languages.

Developing a list of appropriate vocabulary words is not the only challenge researchers face when seeking to develop and use parent report measures in a new language and culture. The pragmatics of language between children and adults can differ greatly across cultures, and has been found to interfere with administration of parent-report measures of early vocabulary, for example in Kiswahili (Alcock, 2017) and Wolof (Weber, Marchman, Diop, & Fernald, 2018).

Despite the myriad challenges that remain in creating new measures of early language development, we believe that the proposed Swadesh CDI list will give researchers a solid foundation to start from. Certainly, increasing the diversity of languages studied is a critical step towards developing a truly general understanding of how young children learn language.

Acknowledgements

We would like to thank all of the contributors to Wordbank, from the researchers who created and adapted the CDIs to those who collected the data (as well as the participants), to those who have created and maintained Wordbank over the years.

References

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

Baker, F. B. (2001). *The basics of item response theory*. ERIC.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P. S., Reznick, J. S., … Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*(1), 85–123.

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476.

Chai, J. H., Lo, C. H., & Mayor, J. (2020). A Bayesian-inspired item response theory-based framework to produce very short versions of MacArthur-Bates Communicative Development Inventories. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3488–3500.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. http://doi.org/10.18637/jss.v048.i06

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual (2nd ed.)*. Baltimore, MD: Brookes.

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, *21*, 95–116.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank project*. MIT Press.

Kachergis, G., Marchman, V. A., Dale, P. S., Mankewitz, J., & Frank, M. C. (2022). Online computerized adaptive tests of children's vocabulary development in English and Mexican Spanish. *Journal of Speech, Language, and Hearing Research*, *65*(6), 2288–2308.

Makransky, G., Dale, P. S., Havmose, P., & Bleses, D. (2016). An item response theory-based, computerized adaptive testing version of the MacArthur-Bates Communicative Development Inventory: Words & Sentences (CDI:WS). *Journal of Speech, Language, and Hearing Research*, *59*(2), 281–289.

Marchman, V. A., Dale, P. S., & Fenson, L. (2023). *MacArthur-bates communicative development inventories user's guide and technical manual, 3rd edition*. Baltimore, MD: Brookes Publishing Co.

Mayor, J., & Mani, N. (2019). A short version of the MacArthur-Bates Communicative Development Inventories with high validity. *Behavior Research Methods*, *51*(5), 2248–2255.

Reckase, M. D. (2009). Multidimensional item response theory models. In *Multidimensional item response theory* (pp. 79–112). Springer.

Swadesh, M. (1971). *The origin and diversification of language*. (J. Sherzer, Ed.). Chicago, IL: Aldine.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929.

Weber, A. M., Marchman, V. A., Diop, Y., & 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. http://doi.org/10.1017/S0305000917000605