A short version of the MacArthur–Bates Communicative Development Inventories with high validity

2 authors:

Julien Mayor
University of Oslo
84 PUBLICATIONS   **858** CITATIONS

Nivedita Mani
Georg-August-Universität Göttingen
127 PUBLICATIONS   **2,474** CITATIONS

CrossMark

A short version of the MacArthur–Bates Communicative Development Inventories with high validity

Julien Mayor[1] · Nivedita Mani[2]

Abstract

The MacArthur–Bates Communicative Development Inventories (CDIs) are among the most widely used evaluation tools for early language development. CDIs are filled in by the parents or caregivers of young children by indicating which of a prespecified list of words and/or sentences their child understands and/or produces. Despite the success of these instruments, their administration is time-consuming and can be of limited use in clinical settings, multilingual environments, or when parents possess low literacy skills. We present a new method through which an estimation of the full-CDI score can be obtained, by combining parental responses on a limited set of words sampled randomly from the full CDI with vocabulary information extracted from the WordBank database, sampled from age-, gender-, and language-matched participants. Real-data simulations using versions of the CDI-WS for American English, German, and Norwegian as examples revealed the high validity and reliability of the instrument, even for tests having just 25 words, effectively cutting administration time to a couple of minutes. Empirical validations with new German-speaking participants confirmed the robustness of the test.

Keywords Language assessment · Word learning · Psychometric testing

The MacArthur–Bates Communicative Development Inventories (CDIs) are among the most widely used evaluation tools for early language development (Fenson et al., 2007). CDIs are filled in by the parents or caregivers of young children, by indicating which of a prespecified list of words and/or sentences their children can already understand and/or produce. Three versions of the forms exist in American English, each targeting a different age: The CDI–Words and Gestures (CDI-WG) assesses early vocabulary comprehension and production along with communicative gestures, targeting infants 8 to 16 months of age; the CDI–Words and Sentences (CDI-WS) assesses productive vocabulary as well as early use of grammar for toddlers 16 to 30 months of age; and the CDI-III (Fenson et al., 2000) is a short form to assess language development for children between 30 and 37 months of age. With a focus on vocabulary development, the CDI-WG contains a checklist of 396 vocabulary items, whereas the CDI-WS includes 680 words, spanning 22 semantic categories (e.g., animals, body parts, action words, descriptive words, and pronouns). Short-form versions of the full CDIs also exist in other languages (e.g., in Galician: Pérez-Pereira & Resches, 2007).

Researchers typically find parents to be reliable and valid indicators of the vocabulary of their typically developing child (Fenson et al., 2007; Fenson et al., 1993), as well as when their child is deaf or hard of hearing (Mayne, Yoshinaga-Itano, Sedey, & Carey, 1998), uses American Sign Language (Anderson & Reilly, 2002) or British Sign Language (Herman, Woolfe, Roy, & Woll, 2010), has cochlear implants (Thal, DesJardin, & Eisenberg, 2007), or is diagnosed with Down syndrome (Galeote, Checa, Sánchez-Palacios, Sebastián, & Soto, 2016) or autism spectrum disorder (Luyster, Lopez, & Lord, 2007).

Initiatives to centralize data collected from the vocabulary checklists of CDIs across languages and instruments (in the WordBank [Frank, Braginsky, Yurovsky, & Marchman, 2017] and its predecessor for cross-language norms, CLEX [Jørgensen, Dale, Bleses, & Fenson, 2010]) have enabled researchers to make full use of the wealth of data to constrain models of language acquisition and test developmental hypotheses (Amatuni & Bergelson, 2017; Braginsky, Yurovsky, Marchman, & Frank, 2015, 2016; Mayor & Plunkett, 2011, 2014; Mollica & Piantadosi, 2017;

✉ Julien Mayor
julien.mayor@psykologi.uio.no

1 University of Oslo, Oslo, Norway

2 University of Göttingen, Göttingen, Germany

Springer

Schneider, Yurovsky, & Frank, 2015). As of January 2018, WordBank contained a repository of over 70,000 fully digitized forms across 25 languages.

Yet, despite the widespread use of CDI data, their collection suffers from the time it takes parents to respond to all the questions, which may considerably limit the instrument's use in clinical settings (when multiple assessments must be conducted in a limited time frame), in multilingual environments (where time spent filling forms is multiplied by the number of languages being assessed), and when parents have limited literacy skills. To address this issue, short forms of the CDIs have been developed (Fenson et al., 2000), which consist of 100 words sampled from multiple semantic categories. Two lists of 100 words have been developed, to allow for repeated testing—since, otherwise, parents would have to respond to identical lists of words across different sessions. These short forms have shown strong reliability and validity and correlate strongly with the full CDIs. Unfortunately, these short forms are subject to ceiling effects when toddlers produce many words (typically from around 27–28 months of age). Furthermore, lists of 100 words may still pose a challenge to parents with limited literacy skills and when the assessment must be conducted rapidly.

In a recent effort aimed at designing a time-efficient test without compromising precision, Makransky, Dale, Havmose, and Bleses (2016) developed an item-response-theory-based computerized adaptive test (hereafter referred to as the *CDI-CAT*), based on the American English CDI-WS. In the CDI-CAT, items sampled from the full CDI-WS are chosen in light of previous responses, to optimize the estimate of the full CDI-WS score. Makransky et al. conducted real-data simulations, based on a norming sample consisting of 1,461 children from 16 to 30 months of age assessed with the CDI-WS, using CDI-CATs of 5, 10, 25, 50, 100, 200, 400, and 680 items. The quality of this model was assessed by computing the correlations of the CDI-CAT estimates to the full CDI-WS scores, the average standard errors (*SE*s), and the reliability of the test in terms of precision ($1 - SE^2$). The CDI-CAT approach showed promising results, even when only a relatively small number of words were used. Based on these real-data simulations, correlations over .95 and *SE*s less than .20 were obtained when the CDI-CAT contained 50 words or more, thereby suggesting that the CDI-CAT may provide a principled and robust alternative to the longer CDI-WS. Nevertheless, Makransky et al. also acknowledged that the acceptability of the test by caregivers/parents had yet to be demonstrated. Furthermore, participants might act differently when responding to a limited number of items that are not grouped thematically or semantically (as in the CDI-CAT) than when they fill in the more structured but longer CDI-WS forms. Consequently, it is likely that performance on the CDI-CAT with new participants would be inferior to that reported in the real-data simulations.

Our approach shares the objectives of Makransky et al. (2016), of producing a time-efficient early vocabulary assessment without compromising validity. In addition, we aim to provide a generic tool that can be applied to multiple languages beyond American English, and to provide a validation of the method with new participants. In our method, an estimation of the full-CDI score is obtained by combining the responses given by the parents/caregiver on a limited set of words sampled randomly from the full CDIs with vocabulary information extracted from the WordBank database, sampled from age-, gender-, and language-matched participants. We first evaluated our method using real-data simulations and the CDI-WS versions for American English (Fenson et al., 2007), German (Szagun, Stumper, & Schramm, 2009), and Norwegian (Simonsen, Kristoffersen, Bleses, Wehberg, & Jørgensen, 2014) as examples. The method was then validated by comparing the vocabulary scores obtained with a new set of German-speaking participants with the full CDI. We present the results and discuss the implications of this novel method for researchers and practitioners, especially in terms of the possibilities of repeated testing.

Method

The model

Our approach has a Bayesian flavor. A series of words, selected randomly from the full CDI, is presented to participant *j*. For each word, the histogram of full-CDI scores from age-, gender-, and language-matching participants is extracted from Wordbank: That is, if participant *j* is a 24-month-old Norwegian-speaking boy who produces word *i* (or does not produce word *i*), the histogram of all 24-month-old Norwegian-speaking males producing word *i* (or respectively not producing word *i*) will be extracted from Wordbank (see Fig. 1, left panels). Each histogram is fitted with a normal distribution, to obtain a smooth and continuous distribution. This histogram, once normalized, can be thought of as the probability distribution of full-CDI scores, given that participant *j* produced word *i*. This procedure is repeated as many times as there are items on the word list. Multiplication of the distributions associated with each new item selected in the short-form CDI, shown in the right panels of Fig. 1, results in a distribution whose mode is measured (Fig. 1, bottom panel). A short-CDI score is produced by the linear transformation of this mode,[1] to be compared with the full-CDI score.

---

[1] This linear transformation is needed in order to reach the full range of CDI scores, since the mode resulting from the multiplication of the distributions associated with all words on the CDI spans a reduced range of CDI scores. The maximal value *max* for the mode is reached for a virtual participant producing all *N* words on the full CDI, whereas the minimal value *min* can be measured using a virtual participant who does not produce any word on the full CDI. The short-CDI score *s* is obtained via affine transformation of the mode *m*: $s = N * (m - min)/(max - min)$.
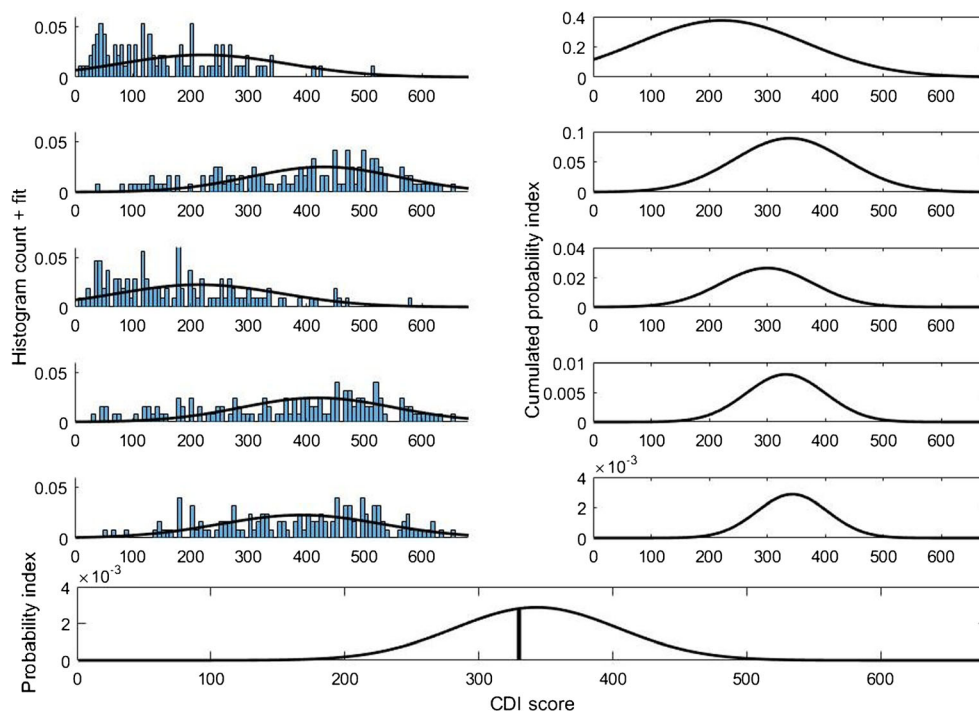
Fig. 1 Sample real-data simulation with a five-item short-form CDI. For each word, the participant's knowledge of that word is examined and the corresponding distribution of matching CDI-WS scores is retrieved from the database, to which a normal distribution is fitted (left panels). The right panels depict the cumulated distributions, updated after each word presentation. The bottom panel depicts the resulting (nonnormalized) distribution of full-CDI score probabilities. The mode of this distribution defines the short-CDI score, to be compared with the full-CDI score (vertical bar)

Real-data simulations

To assess the validity of the model, our approach was two-fold. First, real-data simulations were carried out using the Wordbank database (Frank et al., 2017). For each age group and each sex, correlations were computed between the scores obtained for each participant on Wordbank from short versions of the CDI (covering, respectively, 5, 10, 25, 50, 100, 200, and 400 words, in addition to all words on the CDI) to their corresponding full-CDI scores. In line with Makransky, Dale, Havmose, and Bleses (2016), correlations, reliability, and *SE*s were computed and are reported for American English (Fenson et al., 2007), as well as for German (Szagun et al., 2009) and Norwegian (Simonsen et al., 2014), in production, as assessed by their respective Words and Sentences CDIs.

Empirical validation

Second, the model was validated on new German-speaking participants, recruited in Göttingen, Germany, on 25-word and 50-word versions of the test. The parents of 23 children were recruited for the 25-word version; four of these children had to be excluded due to a missing full-CDI score (two) or because the children were too old for application of the test (two). In the remaining sample

of 19 participants (seven boys and 12 girls), the age range of the children was from 18 to 30 months (M = 22.7). The parents of 33 children were recruited for the 50-word version; eight of the children had to be excluded because they were too old (seven) or because the parents failed to complete the short-CDI version of the test (one). In the remaining sample of 25 participants (16 boys and nine girls), the ages of the children ranged from 18 to 30 months (M = 24.4).

The parents were sent the full-length "FRAKIS" German CDI (Szagun et al., 2009) by post prior to their visit and were asked to complete this CDI before coming to the lab. During their visit to the lab, the parents were given the short, computerized version of the questionnaire (either the 25- or the 50-word version), programmed using the Qualtrics software. The questionnaire was presented on an iPad, and the researchers entertained the children while their parents completed the test. The parents were informed briefly prior to starting the online questionnaire that the procedure was similar to that of the full-length CDI, and they were asked to indicate by pressing the corresponding button whether or not their child produced a certain word. Prior to parents being given the iPad, the researchers filled in an anonymized participant code that ensured that the short CDI was correctly linked to the same child's long CDI form. Completing the questionnaire took, on average, 2 min.

Results

Real-data simulations

Real-data simulations were run on the Wordbank database scores on the Words and Sentences CDIs for American English (Fenson et al., 2007), German (FRAKIS; Szagun et al., 2009), and Norwegian (Simonsen et al., 2014). All data reported consist of the averages of ten simulations.

American English CDI-WS

Real-data simulation were run, for each age group (16–30 months of age) and each sex, on the different short-CDI sizes (5, 10, 25, 50, 100, 200, or 400 words) and on the full list. Figure 2 depicts the short-CDI scores as a function of the full-CDI scores for different short-CDI sizes associated with 24-month-old English-speaking boys.

Correlations between the short lists and the full CDI are reported in Table 1, along with the average SEs and reliability ($1 - SE^2$). To provide a comparison to a parallel approach, the data from Makransky et al. (2016) using their CDI-CAT approach are reported in parentheses.

Our approach outperformed the CAT approach, in terms of correlations, SEs, and reliability for all list sizes. For tests containing 25 words, our SEs were three to four times smaller

than those achieved by the CDI-CAT, and our reliability was accordingly superior (.995, as opposed to either .938 [males] or .956 [females] for the CDI-CAT). Reliability greater than .99 was already achieved with our 25-word test, whereas 200 to 400 words were required for the CDI-CAT approach to reach similar performance levels. It is also noteworthy that our model performed at the same level for both male and female participants, whereas performance on the CDI-CAT was greater for females than for males.

A similar analysis was conducted per age group, since Makransky et al. (2016) reported uneven performance across ages (see Table 2). The CDI-CAT was shown to work best for the 22- to 24-month-olds. Our test captured the same overall pattern—with reduced performance for the youngest age group (16–18 months) and, to a lesser extent, for the oldest age group (28–30 months). This suggests that longer tests may be required in order to obtain sufficient validity when testing young toddlers, since their vocabularies are still limited in size.

German CDI-WS

The model was then applied on the "FRAKIS" German CDI-WS (Szagun et al., 2009) for children 18 to 30 months of age. The results are reported in Table 3. The level of performance achieved by the model for German matched the levels
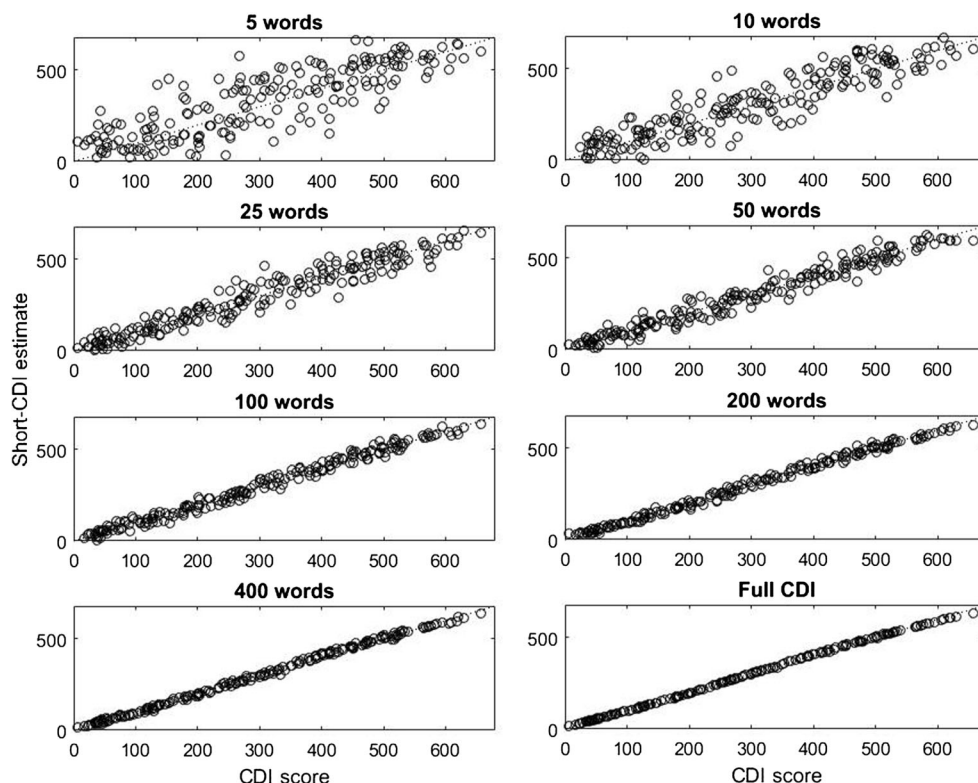
Fig. 2 Real-data simulations of short-CDI estimates when the tests contained, respectively, 5, 10, 25, 50, 100, 200, 400, and 680 words, as a function of the full CDI scores. The data depicted here correspond to 24-month-old English-speaking boys listed in Wordbank

Table 1 Results of real-data simulations with different short-CDI sizes on the American English CDI-WS

Length	Females			Males		
	r With Full CDI	Avg. SE	Rel.	r With Full CDI	Avg. SE	Rel.
Full CDI-WS	.997 (1.000)	.02 (.05)	1.000 (.998)	.999 (1.000)	.01 (.06)	1.000 (.996)
400	.996 (.995)	.03 (.06)	.999 (.996)	.997 (.994)	.02 (.07)	.999 (.995)
200	.992 (.987)	.03 (.09)	.999 (.992)	.994 (.984)	.02 (.11)	.999 (.989)
100	.986 (.977)	.04 (.12)	.999 (.986)	.986 (.970)	.03 (.14)	.999 (.980)
50	.972 (.959)	.05 (.17)	.997 (.972)	.971 (.946)	.05 (.19)	.998 (.963)
25	.947 (.930)	.07 (.21)	.995 (.956)	.943 (.918)	.06 (.25)	.995 (.938)
10	.888 (.869)	.11 (.27)	.988 (.925)	.872 (.828)	.11 (.30)	.988 (.909)
5	.798 (.729)	.15 (.32)	.976 (.896)	.792 (.687)	.15 (.33)	.978 (.890)

Performance from the CDI-CAT approach by Makransky et al. (2016) is reported in parentheses, for comparison

achieved for American English, with correlations of .97 for 50-word lists, SEs of .05, and reliability of .998. No systematic differences in performance were observed across sexes.

Norwegian CDI-WS

Real-data simulations were run on the Norwegian CDI-WS (Simonsen et al., 2014) for 18- to 36-month-olds. The results, displayed on Table 4, reached levels of accuracy comparable with those in the German and American English samples: with correlations of .96 to .97, SEs of .04 to .05, and reliability of .998 for 50-word lists. No systematic differences in performance across sexes were observed. Reliability of .99 was already reached for 10-word tests.

Empirical validation

The data obtained from the 25-word lists administered to the parents of German-learning children correlated with their full CDIs at r = .957, with an SE of .14 and a reliability of .982. Although the correlation level was even higher than in the real-data simulations, the SEs and reliability levels were lower than expected.

Inspection of the data revealed that, frequently, parents responded differently to the same items in the short list of words and in the full CDI. On average, this concerned 10.5% of the items, with about half of the participants (nine out of 19) having one or more inconsistencies. Such inconsistencies reached 36% of the items for the most unreliable case. The equivalent of a test–retest evaluation, between the short list of words and the responses associated with the same items on the full CDI, suggested that the responses correlated to a level of r = .979, with an SE of .09 and a reliability of .992. These values should be considered ceiling performance when considering parental reliability when responding to CDIs. Further inspection of the inconsistencies, using a two-sided paired-sample t test, revealed that participants reponded more positively to a word when it was on the short list than when it was on the full CDI [$t(18) = 2.606$, $p = .018$].

A subsample in which the participants who had over 15% inconsistencies were removed led to a significant improvement (r = .980, SE = .07, and a reliability of .994), thus suggesting that beyond participants' intrinsic unreliability, the method appears to perform well.

The same procedure was repeated with a 50-word model on new participants. The short-CDI scores correlated with the

Table 2 Correlations for different short-CDI sizes on the American English CDI-WS, by age

	16–18	19–21	22–24	25–27	28–30
Full CDI-WS	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
400	1.00 (.99)	1.00 (.99)	1.00 (.99)	1.00 (.99)	1.00 (.99)
200	.99 (.96)	.99 (.98)	.99 (.98)	.99 (.98)	.99 (.98)
100	.98 (.93)	.99 (.95)	.99 (.97)	.99 (.97)	.99 (.97)
50	.95 (.87)	.97 (.91)	.98 (.94)	.98 (.94)	.98 (.94)
25	.90 (.81)	.95 (.87)	.96 (.91)	.96 (.90)	.95 (.89)
10	.79 (.61)	.89 (.74)	.91 (.83)	.91 (.82)	.90 (.78)
5	.68 (.29)	.80 (.57)	.84 (.68)	.84 (.74)	.81 (.65)

Performance from the CAT approach by Makransky et al. (2016) is reported in parentheses, for comparison

Table 3 Results of real-data simulations with the different short-CDI sizes on the German CDI-WS

Length	Females			Males		
	r With Full CDI	Avg. SE	Rel.	r With Full CDI	Avg. SE	Rel.
Full CDI-WS	.994	.03	.999	.999	.01	1.000
400	.993	.03	.999	.998	.01	1.000
200	.990	.04	.999	.994	.02	1.000
100	.985	.04	.998	.986	.03	.999
50	.973	.05	.998	.972	.05	.998
25	.951	.07	.995	.945	.07	.996
10	.891	.11	.988	.878	.11	.989
5	.820	.15	.977	.803	.14	.980

full CDI at $r = .939$, with an SE of .14 and a reliability of .982. Although performance did not reach the levels expected from the real-data simulations, inspection of the data revealed that response inconsistencies (as indexed by differences in responses to the same items between the short list and the full CDI form) reached 14.9%; just seven out of 25 participants responded identically for all items on the short list of words and on the full CDI. Inconsistencies reached 40% of the items, in the most unreliable case. Test–retest measures between the short list and the matching words on the full CDI revealed that the two measures correlated at $r = .974$, with an SE of .09 and a reliability of .992. Again, participants responded more positively in the short list of words than on the full CDI [$t(24) = 3.063, p = .005$].

Similarly, a subsample in which the participants who had over 15% inconsistencies were removed led to a considerable improvement ($r = .991$, $SE = .07$, and a reliability of .996), thus confirming that the method is robust.

Discussion

Communicative Development Inventories are among the most widely used language assessment tools for children up to about 30 months of age. Their adaptations in multiple languages makes them especially important for languages other than English, because several alternative tests are not available. Furthermore, data collected from tens of thousands of CDI administrations have offered scientists unique glimpses into vocabulary acquisition within and across a variety of languages. Making use of the richness of this repository, we have introduced here a method to estimate full-CDI scores from the administration of a much smaller set of words, thus cutting the time down to a few minutes per administration of the test, from up to 45 min when assessing older toddlers. Despite this dramatic reduction of the number of items in the test, the combination of parental responses with data mined from WordBank makes the test both valid and reliable. Importantly, similar levels of performance were obtained for American English, Norwegian, and German, used here as examples, thus demonstrating that this method can successfully be generalized to multiple languages.

Makransky et al. (2016) suggested that a test can be considered of acceptable validity when it correlates to a level of .95 with full CDIs and when the SE is less than .20, a level reached with 50 words in their item-response-theory-based computerized assessment when evaluated with real-data simulations. Although our method, when also evaluated with real-

Table 4 Results of real-data simulations with different short-CDI sizes on the Norwegian CDI-WS

Length	Females			Males		
	r With Full CDI	Avg. SE	Rel.	r With Full CDI	Avg. SE	Rel.
Full CDI-WS	.999	.01	1.000	.999	.01	1.000
400	.996	.01	1.000	.998	.01	1.000
200	.990	.02	1.000	.994	.02	1.000
100	.979	.03	.999	.986	.03	.999
50	.958	.05	.998	.972	.04	.998
25	.919	.06	.996	.945	.06	.996
10	.829	.10	.990	.876	.10	.990
5	.723	.14	.980	.795	.14	.980

data simulations, reached correlation levels of .97 for the same number of words, our *SE*s were three to four times smaller (.05) than those obtained by Makransky et al. An analysis per age group further suggested that the present approach is superior for all ages and all test sizes.

Validation with new German-speaking participants revealed similar levels of performance: The administration of a 25-word test already fell well above the cut for test acceptability (*SE*s of .14, lower than the threshold of .20 suggested by Makransky et al., 2016; and correlations of .96, better than the cutoff of .95). The 50-word administration led to results slightly less valid than we expected from the real-data simulations, yet they still reached high levels of performance (correlation of .94, *SE* of .14). Inspection of the data revealed within-participant inconsistencies (i.e., parents responded differently when the same item was presented in both lists), on the order of 10%–15% of their responses. Such inconsistencies accounted for much of the degradation in performance relative to real-data simulations, since artificially excluding participants with high degrees of inconsistencies made the test extremely valid, in both the 25-word version (*r* = .98, *SE* = .07) and the 50-word version (*r* = .99, *SE* = .07). Such inconsistencies are, of course, an inevitable part of the parental report approach and will de facto establish a baseline performance level for all tests relying on third-party information about a child's vocabulary development. It is noteworthy that participants responded more positively in the short list of words than on the full CDI (i.e., in the case of inconsistent responses, participants more often responded "produced" in the short list of words and "not produced" in the full CDI). The reason underlying this imbalance is as yet unknown, since both tests differ from each other along several dimensions (e.g., full CDIs are administered in a paper format, whereas the short lists of words were administered on an iPad, and full CDIs consist of a semantically structured list of 580 words, whereas the short list sampled words randomly). These differences in response behavior will be subject to further investigation, as well as to validation with new participants in Norwegian and American English.

Performance can nevertheless be improved by implementing a principled selection of words from the full CDI. In its present form, the test features randomly selected words, and hence establishes a lower bound to the level of validity for this general framework of merging big data—as collected from WordBank—with parental reports. A careful selection of words that are maximally informative, such as was implemented by Makransky et al. (2016), could only improve the model further, producing higher performance levels. However, the simplicity of the present approach also means that the test can be run multiple times, such as in intervention studies or longitudinal approaches, while maintaining limited item repetitions across administrations. For example, if the same participant were tested ten times on the 25-word

version, 85% of the words, on average, would have been presented only once. Although Makransky et al. mentioned that the CDI-CAT "can be constrained to not use any of the words used previously," its validity is still to be demonstrated with such a constraint.

Finally, this method, or similar methods of merging "big" data stored in repositories with abridged versions of the full-sized test, can easily be adapted, not just to CDIs in other languages, or for assessing comprehension using the CDI-WG forms as a basis, but also to other psychometric tests. The only requirement is that sufficient data be available, collected on participants with matching key demographics (e.g., depending on the test: age, gender, and using a normative sample), so as to attain high levels of validity and reliability. The considerable effort of collecting data and of sharing them publicly should then be seen as particularly relevant not just for fundamental research, but also in providing a platform for building the next generation of psychometric tests.

References

Amatuni, & Bergelson (2017). *Semantic networks generated from early linguistic input*. bioRxiv preprint. bioRxiv:157701; 10.1101/157701

Anderson, D., & Reilly, J. (2002). The MacArthur communicative development inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education*, *7*, 83–106.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2015). Developmental changes in the relationship between grammar and the lexicon. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 256–261). Austin, TX: Cognitive Science Society.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2016). From *uh-oh* to *tomorrow*: Predicting age of acquisition for early words across languages. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 1691–1696). Austin, TX: Cognitive Science Society.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur–Bates communicative development inventories*. Baltimore, MD: Brookes.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., ... Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular.

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*, 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*, 677–694.

Galeote, M., Checa, E., Sánchez-Palacios, C., Sebastián, E., & Soto, P. (2016). Adaptation of the MacArthur–Bates Communicative Development Inventories for Spanish children with Down Syndrome: Validity and reliability data for vocabulary. *American Journal of Speech-Language Pathology*, *25*, 371–380.

Herman, R., Woolfe, T., Roy, P., & Woll, B. (2010). Early vocabulary development in deaf native signers: A British Sign Language

adaptation of the communicative development inventories. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*, 322–331. https://doi.org/10.1111/j.1469-7610.2009.02151.x

Jørgensen, R. N., Dale, P. S., Bleses, D., & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of Child Language*, *37*, 419–428

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur–Bates Communicative Development Inventory (CDI). *Journal of Child Language*, *34*, 623–654.

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*, 281–289.

Mayne, A. M., Yoshinaga-Itano, C., Sedey, A. L., & Carey, A. (1998). Expressive vocabulary development of infants and toddlers who are deaf or hard of hearing. *The Volta Review*, *100*(5), 1–28

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*, 769–785.

Mayor, J., & Plunkett, K. (2014). Shared understanding and idiosyncratic expression in early vocabularies. *Developmental Science*, *17*, 412–423.

Mollica, F., & Piantadosi, S. T. (2017). How data drive early word learning: A cross-linguistic waiting time analysis. *Open Mind: Discoveries in Cognitive Science*, *1*, 67–77. https://doi.org/10.1162/OPMI_a_00006

Pérez-Pereira, M., & Resches, M. (2007). Elaboración de las formas breves del Inventario do Desenvolvemento de Habilidades Comunicativas. Datos normativos y propiedades psicométricas. *Infancia y Aprendizaje*, *30*, 565–588.

Schneider, R., Yurovsky, D., & Frank, M. (2015). Large-scale investigations of variability in children's first words. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2110–2115). Austin, TX: Cognitive Science Society.

Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, *34*, 3–23. https://doi.org/10.1177/0142723713510997

Szagun, G., Stumper, B., & Schramm, A. S. (2009). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform)*. Frankfurt, Germany: Pearson Assessment.

Thal, D., DesJardin, J. L., & Eisenberg, L. S. (2007). Validity of the MacArthur–Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants. *American Journal of Speech-Language Pathology*, *16*, 54–64.