HHS Public Access

Author manuscript

Dev Psychol. Author manuscript; available in PMC 2020 January 01.

Published in final edited form as:

Dev Psychol. 2019 January ; 55(1): 9–22. doi:10.1037/dev0000617.

Language Status at Age 3: Group and Individual Prediction from Vocabulary Comprehension in the Second Year

Margaret Friend,
San Diego State University

Erin Smolak,
San Diego State University/University of California, San Diego Joint Doctoral Program in Language and Communicative Disorders

Tamara Patrucco-Nanchen,
University of Geneva

Diane Poulin-Dubois, and
Concordia University

Pascal Zesiger
University of Geneva

Abstract

The present research extends recent work on the prediction of preschool language skills by exploring prediction from decontextualized vocabulary comprehension. Vocabulary comprehension was a stronger predictor than parent reported production, yielding a quadrupling of variance accounted for relative to prior studies. Parallel studies (Studies 1 and 2) are reported for two linguistically and geographically distinct samples. In both samples, decontextualized vocabulary comprehension late in the second year provided the best balance between model fit and parsimony in predicting language skills at age three. In Study 3, vocabulary comprehension prospectively identified children with low language status two years earlier than other prospective studies but with similar sensitivity and specificity. The present paper provides evidence on three questions of practical and theoretical significance: the relation between decontextualized vocabulary prior to 30 months of age and language outcomes, how prediction from decontextualized vocabulary compares to parent reported vocabulary and finally, how early stable predictions to language outcomes can be made.

Keywords

Vocabulary; Comprehension; Prediction; Practical Significance; French; English

Word learning develops rapidly in the first two years (Reznick & Goldfield, 1992; Fenson et al., 1994; Dapretto & Bjork, 2000; Mayor & Plunkett, 2010; Samuelson & McMurray, 2017) and vocabulary production (Morgan, Farkas, Hillemeier, Hammer & Maczuga, 2015), comprehension (Friend, Schmitt, & Simpson, 2012), and speed of word processing (Fernald & Marchman, 2012; Marchman & Fernald, 2008), are building blocks for later linguistic and cognitive development. The present paper evaluates the role of early vocabulary in

predicting language skill at age three in geographically and linguistically distinct samples of monolingual children at the group and individual levels. In so doing we attempt to reconcile evidence that word learning emerges from domain general processes that are expected to be stable with evidence that, in general, early vocabulary accounts for only a small proportion of variance at the group level and is unstable at the individual level.

Recent research reveals that domain general mechanisms can account for the pattern of vocabulary acquisition with age (Samuelson & McMurray, 2017; Vlach & DeBrock, 2017; Vlach & Johnson, 2013; Vlach & Sandhofer, 2011; Yu & Smith, 2012). For instance, young children's attentional biases reduce referential ambiguity (Samuelson & McMurray, 2017; Yurovsky, Smith & Yu, 2013) and both vocabulary size and memory contribute to cross-situational word learning (Smith & Yu, 2013) supporting the development of stable word-referent relations (Vlach & DeBrock, 2017; Vlach & Johnson, 2013; Vlach & Sandhofer, 2011). Initially weak word-referent relations may be strengthened over time through the iterative application of domain general learning (Bion, Borovsky & Fernald, 2013; Gershkoff-Stowe and Hahn, 2013; Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015; Hendrickson, Zesiger, Poulin-Dubois, & Friend, 2017; Yu & Smith, 2012). From this view, early vocabulary should evince stability with later abilities that build on these mechanisms (e.g., language, school readiness, and achievement; Duff, Reen, Plunkett, & Nation, 2015; Friend, Smolak, Liu, Poulin-Dubois, & Zesiger, 2018; Morgan et al., 2015). Since these mechanisms are presumed to be universal, this expectation also applies across languages.

In support of this idea, by 24 months of age, parent reported vocabulary predicts later language (Duff, et al., 2015; Ghassabian et al., 2014; Henrichs et al., 2011; Kemp et al., 2017; Reilly et al., 2010), literacy and reading (Bleses, Makransky, Dale, Højen, & Ari, 2016; Duff et al., 2015), and academic and behavioral functioning (Morgan et al., 2015) at the group level in English-, Dutch-, and Danish-speaking children. However, it accounts for a small to modest proportion of variance (Duff, et al., 2015; Morgan et al., 2015; Reilly et al., 2010). At the individual level, prediction is inadequate (Law et al., 2000a; Law & Roy, 2008). Across studies, parent report prospectively identifies only roughly one-half of children who develop language problems (e.g., Dale, Price, Bishop, & Plomin, 2003; Heilmann, Weismer, Evans, & Hollar, 2005; Westerlund, Berglund, & Eriksson; 2006). From these findings, Dale et al. (2003) concluded that supplemental assessment is necessary to identify developmental risk.

How can we explain the weak prediction from vocabulary prior to the third year to subsequent language and literacy at the group level and to language problems at the individual level? Imagine a child who produces the words "dog" and "milk." This child may have a strong association between the word dog and its referents and use it appropriately across contexts but a relatively weak association between milk and its referent, using it only in the context of breakfast. Indirect assessments such as parent report may assess this full continuum of word-referent associations from weak to strong. In both cases, parents should report these as words their child produces. In contrast, direct assessments that require active lexical retrieval and hypothesis testing (Is this a duck or is that a duck?) should preferentially tap strong, rather than weak, associations (Yu & Smith, 2012). Weak associations are

considered fragile, subject to interference, and context-bound (Bion, et al., 2013; Gershkoff-Stowe and Hahn, 2013) whereas strong associations, formed through iterative domain general processes, are stable across situations. We refer to these stable associations as decontextualized.

Decontextualized, in contrast to context-bound, associations may provide the substrate for subsequent word knowledge and conceptual development (Schmitt, 2014) and predict downstream language and cognitive ability. Friend, et al. (2018) found that decontextualized vocabulary in the second year predicted vocabulary comprehension and kindergarten readiness at age four in monolingual English and French-speaking children and in bilingual children acquiring French and English simultaneously. Decontextualized vocabulary was a stronger predictor than parent report with effect sizes comparable to or greater than those in prior studies.

The Present Research

This research builds on recent research on predicting language and literacy from early vocabulary. First, we assess prediction from decontextualized vocabulary at 16 and 22 months of age to preschool language in two groups (American English and Swiss French) that differ geographically, culturally, and in native language. Because these languages have distinct prosody, syntax, and grammar and differ both in age-related vocabulary and MLU (Bleses et al. 2008; Thordardottir, 2005) we conduct parallel analyses in Studies 1 and 2 to assess generalizability. We utilize spontaneous and elicited measures to capture the breadth of preschool language (vocabulary comprehension, expressive language complexity, sentence comprehension, grammar, and syntax) and extract a single factor to estimate core language ability and eliminate method variance (Bornstein, et al., 2016). We anticipate prediction from early vocabulary to core language ability that is statistically and practically significant at the group and individual levels.

Second, we evaluate the relative contributions of decontextualized vocabulary and parent report. We anticipate decontextualized vocabulary to account for unique variance beyond parent report, resulting in lower error and better fit. Third, consistent with guidelines from the Individuals with Disabilities Act (Wallace, et al., 2015), we contrast models across time points to find the age of earliest prediction (Bornstein, Plutnick, & Esposito, 2017) and expect models to become more stable with age (Bornstein, Hahn, & Putnick, 2016). Finally, in Study 3 we assess the sensitivity and specificity of decontextualized vocabulary to prospectively identify individual children with low language.

This research was conducted under the project, The Path from Language to Literacy, supported by the NICHD and approved by the Institutional Review Boards at San Diego State University (protocol #603057) and at the University of Geneva.

Study 1

Method

Participants.—Seventy-nine English-speaking monolingual children (41 girls) were recruited as part of a larger, multi-institutional longitudinal project investigating children's path to literacy. Participants were recruited from Women, Infant, and Children Centers, the YMCA, local churches, parenting groups, swap meets, child-oriented festivals, and birth records in a large city in the Southwestern U.S. Thirty-five children were excluded due to failure to complete at least one task across waves (n = 11), becoming bilingual (n= 1), or attrition (n = 23 or 30%). Roughly one-half of the attrition was due to parents moving out of state (n=10) with the rest due to lost contact (n=12), or a change in lab location (n=l). The final sample consisted of 44 children (26 girls). To test whether this sample size was appropriate to our aims, we conducted an a priori power analysis in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) using the observed correlations between the CCT, MCDI, and language sample MLU found in Friend et al. (2012). The analysis confirmed that the present sample size was sufficient to detect similar effect sizes with power = .8.

Participants visited the lab on three occasions: Wave 1 at 16 months of age (M= 16;21, range 15; 15 – 18;3), Wave 2 at 23 months (M= 23;0, range 21;6 – 25; 12), and Wave 3 at 36 months (M= 37;23; range 35;9 – 41;24). All infants were full term, had no diagnosed hearing or vision impairments, and were exposed at least 80% of the time to their native language. A $25 gift card to a major retailer and a small toy were provided as incentives at each Wave. See Table 1 for demographic data on the final sample.

Measures.

Language Exposure Assessment Tool (LEAT).: Language exposure was estimated on the LEAT (DeAnda, Bosch, Poulin-Dubois, Zesiger, & Friend, 2016) prior to the initial visit to insure monolingual status. This estimate derived from parent reports of the number of hours of language input by all interlocutors over the course of the child's life. Internal consistency is excellent (Cronbach's α = .96).

MacArthur-Bates Communicative Development Inventories (MCDI).: The MCDI is a parent report measure of early comprehension and production (Fenson et al., 1993, 1994). It consists of two forms: Words and Gestures (WG), for children from 8 to 18 months of age, and Words and Sentences (WS), for children from 16 to 30 months. The WG measure contains a checklist of 396 words, on which caregivers indicate words children understand and words they understand and say. At 16 months, receptive and expressive vocabulary were estimated from the WG form. The WS form contains a vocabulary checklist of 680 words and assesses only words that children say. At 23 months, expressive vocabulary was estimated from the WS form. The MCDI: WG and WS evince moderate to high internal consistency and good test-retest reliability. Six- month stability is moderate for both forms (Fenson et al., 1994).

Computerized Comprehension Task (CCT).: The CCT is a forced-choice measure of vocabulary comprehension administered on a touchscreen (Friend & Keplinger, 2003; 2008;

available at https://childes.talkbank.org). Paired images appear on the screen as an experimenter delivers a prompt in which the target word is embedded (e.g., Where is the *dog*? Who is *swimming*? Which one is *old*?). It is expected that retrieving word-referent associations upon hearing the prompt and selecting an association via haptic response correspond to processes of lexical retrieval and hypothesis testing. Correct responses are thought to reflect children's decontextualized word-referent associations. Each trial has a maximum duration of 7 seconds (sufficient to execute a haptic response) and trials are interleaved with a blank blue screen. The pace is experimenter-controlled to insure that trials are presented only when the child is quiet, alert, and looking at the screen. Administration followed Friend, et al. (2012) with the following additional criteria for repeating or terminating trials. Repetitions of trials were allowed under the following conditions: 1) the child attempts a response but does not complete the touch before the end of the trial, 2) the child becomes distracted and misses the trial, 3) the child accidentally touches the screen, or 4) the child has not made any attempts for the last 3 consecutive trials. In the last case, the experimenter attempts to re-engage the child by moving the child's hand to the target, or by touching the target to elicit the rewarding stimulus. If the child becomes fussy during the procedure, and 3 attempts to re-engage have failed, the experimenter terminates the procedure. If this is necessary and the child has completed one-third of the trials or less, they are excluded from analyses. Three children met this criterion (n=2 at Wave 1 and n= 1 at Wave 2).

Words on the CCT are derived from the MCDI vocabulary checklist (Fenson, Bates, Dale, Marchman, Reznick, & Thai, 2007). Referents are high-quality, colorful digital images that are prototypical exemplars. Pairs are matched on color, size, saliency, word class, and difficulty (see Table SI for full item set). Word difficulty was based on 16-month norming data (Frank, Braginsky, Yurovsky, & Marchman, 2017). There are approximately equal numbers of easy (comprehension ≥ 66 percent), moderate (comprehension = 33 to 66 percent) and difficult (comprehension ≤33 percent) words randomly distributed throughout the test. The inclusion of more difficult items allows the CCT to be extended up to 2 years of age. Two forms are counterbalanced across participants, such that each word serves as both target and distractor. Finally, target side is randomized, with the restriction that it not appear on the same side on more than two consecutive trials following Hirsh-Pasek and Golinkoff (1996).

The CCT has strong immediate test-retest reliability and moderate short-term stability over a 4-month period (Friend & Keplinger, 2008) and correlates concurrently with parent-report and predictively with a language sample (Friend et al., 2012). Internal consistency is strong across forms (Cronbach's α = .86 and .93, respectively; Friend, et al., 2018). It is the only measure of decontextualized vocabulary size prior to 30 months of age.

Peabody Picture Vocabulary Test (PPVT-III).: The PPVT is an adaptive measure of vocabulary comprehension appropriate from 30 months of age through adulthood (Dunn & Dunn, 1997). Vocabulary comprehension is associated with subsequent language, literacy, and academic success (Dickinson, Golinkoff, & Hirsh-Pasek, 2010; *Removed for blinding*, accepted with revision; Oakhill & Cain, 2012; Silva & Cain, 2015) and is therefore an important measure of 36-month language skill.

An experimenter displays four pictures and asks the child to point to the one that corresponds to a target word. Difficulty increases with age and the final score is the number of items to reach ceiling minus errors. Like the CCT, the PPVT yields a direct estimate of vocabulary comprehension. It was standardized on a sample representative of the U.S. population and has generally strong reliability for all age ranges tested and strong internal consistency (Dunn & Dunn, 1997).

Free play language sample.: Children and caregivers played with an extended Fisher-Price farm play set, which included several structures, vehicles, toy people, and animals for 15 minutes. The full session was recorded with a Zoom H2n Handy Recorder microphone. This allowed us to assess spontaneous, as opposed to elicited, language usage at 36 months. Child language samples were transcribed, coded for grammatical morphemes, and analyzed for Mean Length of Utterance (MLU) using the Systematic Analysis of Language Transcripts software (SALT; Miller & Iglesias, 2012). MLU reflects general language ability that is correlated with grammatical and semantic development (Dethorne, Johnson, & Loeb, 2005) and is lower in children with language impairment than in typically-developing peers (Hewitt, Hammer, Yont, & Tomblin, 2005).

To ensure stability across transcripts varying in child talkativeness, we restricted MLU analysis to the first 100 complete and intelligible utterances. Thirty-five children (80%) met this criterion leaving 9 children whose transcripts included less than 100 utterances (M = 85.78, range = 58 to 97). Case-by-case review indicated no systematic difference in language skills from the larger sample. We retained these cases and calculated MLU in morphemes (MLU) over the entire transcript.

Five trained assistants transcribed 13 language samples each using Express Scribe Transcription Software (available at: http://www.nch.com.au/scribe/) and coded them for lexical units, plurals, articles, tense markers, possessives, and contractions. The assistants were trained using the SALT video training module to an inter-rater agreement of .90. Reliability checks were performed by four trained transcribers/coders for 2-4 transcripts from each transcriber for a total of 12 transcripts or approximately 20% of the full sample. Morpheme-level inter-rater agreement was .90.

Sentence repetition (SR).: This task was based on Devescovi and Caselli (2007). The test included 27 sentences of varying complexity and length (see Table S2) accompanied by images depicting sentence-level meaning. SR improves significantly with age and correlates with concurrent spontaneous production. This elicited measure complements our measure of spontaneous production by tapping into diverse skills in language processing, sentence comprehension, production and syntax (Klem et al., 2015). It is also a cross-linguistic marker for language impairment (Conti-Ramsden, Botting & Faragher, 2001; Armon-Lotem & Meir, 2016).

The experimenter told the child to repeat after her. With the image covered, she modeled the sentence then revealed the image. Sentences were repeated up to three times. The child's first attempt at repetition was scored. The second author coded all SR data. The number of sentences repeated correctly ranged from 0 to 27. A correct repetition included all words and

morphemes in the correct order with no extraneous words. Twenty-seven percent of the data were reliability-coded by one additional coder. Sentence-level inter-rater agreement was .92.

Procedure.—Before each Wave, caregivers completed the LEAT interview over the phone. At the lab, following a brief warm-up period, caregivers and children were escorted to a playroom for testing. At Waves 1 and 2, this warm-up included a game to familiarize children with the touch-sensitive screen. Next, toddlers completed the CCT seated on their caregivers' laps approximately 30 cm from the screen; parents wore opaque sunglasses and listened to masking music over noise-cancelling headphones. Following the CCT, caregivers completed the MCDI. The WG form was completed at Wave 1, and the WS form was completed at Wave 2.

At Wave 3, dyads participated in three assessments: free play, PPVT, and SR. During free play, caregivers were instructed to play with their children as they would at home. Next, children were administered the PPVT and SR task. Free play always occurred first to facilitate optimal performance across tasks and the order of the other tasks was counterbalanced across participants.

Results

Descriptive data are reported below on the raw scores for all measures. English language exposure ranged from .87 to 1.00 (M=.99, SD=.03). At 16 months, expressive vocabulary on the MCDI ranged from 0 to 233 words (M= 43.80) corresponding to the 1st to the 99th percentile. Receptive vocabulary on the MCDI ranged from 63 to 355 words (M=184.16) corresponding to the 1st to the 99th percentile. CCT receptive vocabulary ranged from 0 to 29 words (M= 12.11) and internal consistency was excellent (Cronbach's α = .92 and .95 for forms A and B, respectively). Twenty-eight children completed reliability trials and immediate test-retest reliability was high (r(26) = .79, p< 001). Measures were approximately normally distributed with the exception of MCDI expressive vocabulary (*skewness* = 2.25, *kurtosis* = 5.51), indicating floor effects. A log transformation yielded equivalent results to the raw data in all analyses therefore raw data are presented.

At 23 months, MCDI expressive vocabulary ranged from 5 to 614 (M= 259.34), corresponding to the 1st to the 98th percentile. CCT receptive vocabulary ranged from 10 to 37 words (M= 27.93). Internal consistency was strong (Cronbach's α = .86 and .97 for forms A and B, respectively). Forty-one children completed reliability trials, and immediate test-retest reliability was moderate (r(39) = .55, p < .001). Both measures were approximately normally distributed.

At 36 months, PPVT receptive vocabulary ranged from 10 to 93 (M=51.43), corresponding to the 2nd to the 99th percentile. MLU in morphemes ranged from 1.23 to 5.04 (M = 3.40), within the expected range at this age (Miller & Chapman, 1981). SR scores ranged from 0 to 26 (M=15.48, SD=7.51) and mirrored Devescovi and Caselli's (2007) findings. All measures were approximately normally distributed.

We first evaluated the role of control variables (age, sex, and maternal education) on language skills. Maternal education served as a proxy for SES due to the relation between

maternal education and early vocabulary (Hoff, 2013). Age was not significantly correlated with any predictors or dependent measures (all *p*s >.18). There was a negative correlation between sex and SR scores ($r(42)=-.32$, $p=.03$) and between sex and MLU ($r(42)=-37$, $p=.013$): boys performed slightly more poorly than girls. Maternal education correlated with CCT receptive vocabulary at 23 months ($r(42)=.31$, $p=.04$) and with SR at 36 months ($r(42)=.35$, $p=.02$). Both maternal education and child sex were included as control variables in subsequent analyses.

We evaluated zero-order relations between predictor and 36-month language skill variables to provide context for our predictive analyses (see Table 2). At 16 months, MCDI comprehension and production and the CCT were significantly correlated with SR at 36 months but no 16-month measure correlated with MLU or PPVT (*p*s > .90). At 23 months, MCDI production was significantly correlated with SR and MLU but not with PPVT at 36 months ($p = .09$) whereas the CCT correlated significantly with the PPVT, SR, and MLU.

Prediction to 36-month language skills.—We transformed the PPVT, MLU, and SR to sample-specific z-scores and entered these into an exploratory factor analysis (see Table 3 for the component matrix). All participants contributed data for each indicator. A single factor explained 60.54% of the variance. This Language Factor was significantly correlated with 16- and 23-month MCDI production ($r(42)=.39$, $p=.009$ and $r(42)=.51$, $p<.001$, respectively), and 23-month CCT ($r(42)=.62$, $p<.001$). This composite is derived by removing unshared variance of the indicators to arrive at a more robust representation of broad language skill at 36 months (e.g., Bornstein, et al., 2016).

We conducted two stepwise, hierarchical linear regressions to independently assess prediction at 16 and 23 months of age. We took this approach because the more proximal measures may suppress the predictive power of the earlier measures (Bornstein et al., 2017). We used backward selection to remove non-significant predictors sequentially. This permits the unique contribution of the remaining variables to be more accurately estimated. The criterion for the removal was $p > .10$. In the first model, Language Factor was entered as the dependent measure with child sex, maternal education, 16-month CCT comprehension, and 16-month MCDI comprehension and production entered as predictors. The final model ($F(2,41)=7.09$, $p=.002$) included child sex and MCDI production. Tolerance was excellent at .999. Observed power = .92 (G*Power; Faul et al., 2007). A follow-up t-test of the effect of sex on Language Factor scores was not significant ($p=.05$). In the second model, child sex, maternal education, 23-month CCT comprehension, and 23-month MCDI production were entered using the step-wise method with the Language Factor as the dependent measure. Tolerance was good at .798. The final model ($F(1,42)=16.83$, $p < .001$) included CCT comprehension and MCDI production. Observed power = .99. See Table 4 for parameter estimates for the final model and excluded variables.

To identify the most parsimonious model with the best fit, we contrasted all possible models from significant predictors identified in the stepwise regressions using the Akaike Information Criterion (AIC; Posada & Buckley, 2004, see Table 5). The AIC evaluates the loss of information in each model in approximating the data using maximum likelihood estimation and imposes a penalty for model complexity. Lower AIC scores are associated

with higher quality. The lowest AIC values were obtained for 23- month CCT Comprehension alone (AIC = 108.49), 23-month MCDI production and CCT comprehension (AIC = 105.48), and the full complement of predictors at both ages (AIC = 107.03). For the current sample size, a 10-point spread in AIC scores would indicate a meaningful difference in the fit of the candidate models (Hilbe, 2011). Whereas these models cannot be distinguished in terms of fit, we can conclude that, of the candidate models, the one with 23-month CCT as the sole predictor provides the best balance of fit and parsimony.

Discussion

As predicted, decontextualized vocabulary comprehension at 23 months uniquely predicted language skills at 36 months. Parent reported vocabulary comprehension at 16 months of age was the earliest predictor of later language skills however, by 23 months, decontextualized vocabulary offered the best balance of fit and parsimony consistent with our expectation of increased stability with age. Before assessing the practical significance of these findings, we first assess the generalizability of our findings in a sample of French-speaking children in Switzerland.

Study 2

Method

Participants.—Sixty-six Swiss-French-speaking monolingual children (33 girls) were recruited through birth lists in a large city in Switzerland. Six children were excluded for not completing a task at one visit (n=4) or attrition (n=2; 3%). The final sample consisted of 60 toddlers (30 girls) all of whom had been carried to term and had normal hearing and vision. Participants made three visits to the lab: Wave 1 at 16 months (M= 16;0, range=15;6–17;l), Wave 2 at 22 months (M= 21;28, range=21;0–22;6), and Wave 3 at 36 months (M= 35;25, range=34;8–37;2). See Table 6 for demographic data on the final sample.

Measures.

Language Exposure Assessment Tool (LEAT).: Identical to Study 1.

L'Inventaire Français du Développement Communicatif (IFDC).: The IFDC (Kern, 2003; Kern, 2007; Kern & Gayraud, 2010) is the European-French adaptation of the MCDI. The IFDC: Mots et Gestes (MG) corresponds to the MCDF WG and the IFDC: Mots et Phrases (MP) corresponds to the MCDF WS. Vocabulary comprehension and production were estimated from the IFDC: MG at 16 months and production was estimated from the IFDC: MP at 22 months.

Computerized Comprehension Task (CCT).: The French CCT was adapted from the English CCT. The design and administration were the same as Study 1. Translation equivalents across languages were included whenever possible while maintaining the same distribution of word class and difficulty (see Table S3). Images were prototypical exemplars in the region where children were tested. The French CCT has moderate test- retest reliability and convergent validity with the IFDC (Friend & Zesiger, 2011). Internal

consistency is strong across forms (Cronbach's α = .92 and .91, respectively; Friend, et al., 2018).

Échelle de Vocabulaire en Images Peabody (EVIP).: The EVIP (Dunn, Thériault-Whalen, & Dunn, 1993) is the French adaptation of the PPVT normed on a large representative sample of French speakers in Canada.

Free play language sample.: Identical to Study 1. Fifty-nine children (98 percent) met the criterion of 100 complete and intelligible utterances leaving 1 child whose transcript contained only 89 utterances. We retained this case and calculated MLU over the entire transcript. Five trained assistants transcribed and coded child language samples. Morpheme-level inter-rater agreement for 25% of the total sample was .89.

Sentence repetition (SR).: Identical to Study 1, adapted to French. All SR data were reliability-coded. Sentence-level inter-rater agreement was .99. See Table S4.

Procedure.—Identical to Study 1.

Results

Exposure to French ranged from .80 to 1.00 (*M*=.96 *SD*=.06). At 16 months, IFDC expressive vocabulary ranged from 0 to 185 words (*M*= 26.26) corresponding to the 5th to the 90th percentile. IFDC receptive vocabulary ranged from 52 to 387 words (*M* = 200.45), corresponding to the 5th to the 90th percentile. CCT receptive vocabulary ranged from 2 to 32 words (*M*= 15.95). The internal consistency of the CCT was excellent across forms (Cronbach's α = .92 and .90, respectively). Thirty-three children completed reliability trials and immediate test-retest stability was moderate ($r(31) = .54$, $p=.001$). Measures were approximately normally distributed with the exception of expressive vocabulary on the IFDC (*skewness* = 2.82, *kurtosis* = 9.46). A log transformation yielded equivalent results so subsequent analyses are reported on the raw data.

At 22 months, IFDC expressive vocabulary ranged from 13 to 523 (*M*= 196.40), corresponding to the 5th to the 90th percentile; CCT receptive vocabulary ranged from 12 to 40 words (*M*= 28.71). Internal consistency was strong across forms (Cronbach's α = .92 and .87, respectively). Fifty-seven children completed the reliability trials and immediate test-retest stability was moderate ($r(55) = .49$, $p < .001$). All measures were approximately normally distributed.

At 36 months, MLU from the language sample ranged from 2.18 – 5.82 (*M*= 4.03), EVIP receptive vocabulary ranged from 7 to 61 (*M*= 27.71), corresponding to the 2nd to the 99th percentile, and sentences correct on the SR task ranged from 0 to 25 (*M* = 10.96). Measures were approximately normally distributed.

There was no relation between child sex or maternal education and language measures at any wave. However, to parallel Study 1, maternal education and child sex were included as control variables in the analyses. At 16 months, IFDC comprehension was significantly correlated with EVIP and MLU at 36 months, but not with SR ($p=.26$). IFDC production

was not correlated with any 36 month variable (*p*s>.53) and CCT comprehension was significantly correlated with EVIP and SR, but not with MLU (*p*=.70). At 22 months, IFDC production was significantly correlated with SR but not with EVIP or MLU (*p*s > .07). CCT comprehension was significantly correlated with EVIP and SR, but not with MLU (*p*=.06; see Table 7).

Prediction to 36 month language skills.—Following the procedure in Study 1, we computed a composite language score at 36 months (see Table 8 for the component matrix) using sample-specific z-scores. A single factor explained 54.01% of the variance. This Language Factor was significantly correlated with 16-month IFDC (*r*(58)=.36, *p*<.01) and CCT comprehension (*r*(58)=.27, *p*=.04), and 22-month IFDC production (*r*(58)=.32, *p*=.01), and CCT comprehension (*r*(58)=.56, *p*<.001).

We conducted two stepwise, hierarchical linear regressions to independently assess prediction at 16 and 22 months of age. At 16 months, the final model included only MCDI comprehension (*F* (1,58)=8.78, *p*=.004) and at 22 months, the final model included only CCT comprehension (*F*(1,58)=26.63, *p* < .001, see Table 9). Observed power was .84 and . 99 for the final models in regressions 1 and 2, respectively. We contrasted all possible models using the significant predictors at 16 and 22 months by calculating the AIC values for each model (see Table 10). The lowest AIC values obtained for 22-month CCT comprehension (AIC = 152.60) and 16-month MCDI production and 22-month CCT comprehension (AIC = 147.18). As in Study 1, these models cannot be distinguished in terms of fit, but we can conclude that, of the candidate models, the one with 23-month CCT as the sole predictor provides the best balance of fit and parsimony.

Discussion

Results for the French sample largely paralleled those for English: parent reported vocabulary was a stronger predictor of 36 month language skill than was decontextualized comprehension at 16 months and decontextualized comprehension at 22 months was a stronger predictor than parent report. The only difference was that, in the English sample at 16 months of age, parent-reported vocabulary *production* predicted language skill whereas in the French sample, parent reported *comprehension* was predictive. Consistent with our expectation, the balance of model fit and parsimony was superior for the model including only decontextualized vocabulary at 22 months.

Next we evaluate the practical significance of these findings for identifying children with low language skills at the individual level. Studies 1 and 2 each yielded a one-factor solution for the 36-month language variables with factor loadings that were remarkably similar suggesting that this factor has a similar underlying structure across samples. Therefore, we combine samples to take advantage of the increase in sample size.

Study 3

Method

Participants.—All English-speaking monolingual children from Study 1 and French-speaking monolingual children from Study 2 were included in Study 3, resulting in a final

sample of 104 children. Although a typically developing sample was recruited at 16 months, demographic data from the 22- and 36-month visits indicated that some children had received speech/language services or a diagnosis of language delay as a primary feature after the first visit. In the English sample, five children received services prior to 36 months of age for diagnosed phonological disorder (n=1), expressive language delay, sensory processing disorder, and mild autism spectrum disorder (ASD) (n=1), and low expressive language with no other identified deficits (n=3). In the French sample, three children received services prior to 36 months of age for universal dyslalia (n=1), dysphasia (n=1), and low expressive language (n=1).

Choice of Gold Standard.—Consistent with our interest in discriminating children with low language skills at 36 months from their average-to-high language peers, we chose the Language Factor score as our "gold standard" for signal detection analysis to assess the practical significance of Studies 1 and 2 for individual children. Choosing a single gold standard is difficult (Heilmann, Weismer, Evans, & Hollar, 2005; Eriksson, Westerlund, & Miniscalco, 2010; Westerlund et al., 2006). For example, spontaneous language is subject to contextual variation: length of session, time of day, and the conditions under which it was recorded can all influence the quality of the sample. On the other hand, standardized assessments may not capture the richness of child language and history of speech-language services can conflate children with primary language impairment with those who "catch up." We chose the Language Factor score as the gold standard because it takes into account vocabulary, grammar, and general language ability derived from both spontaneous language and standardized assessments. Vocabulary is associated with later language and literacy and SR and MLU are recognized markers of language impairment. Finally, the Language Factor is more robust than any individual measure since the process of factor construction removes unshared variance.

Procedure.—We used signal detection analysis to investigate the practical significance of decontextualized vocabulary in the second year for prospectively discriminating individual language skills. In this approach, a Receiver Operating Characteristic (ROC) curve plots accuracy in identifying low language children against accuracy in identifying their average-to-high language peers. The ROC analysis was conducted in R (R Core Team, 2016) with the pROC script (Robin et al., 2011). The CCT at 22–23 months was entered as the predictor and score on the Language Factor at 36 months was the dependent measure. We evaluated the empirical ROC curve and a binormal-smoothed curve. Binormal smoothing serves to normally distribute scores separately for the low and average-to-high vocabulary groups (Robin et al., 2011). This transformation is robust, provides a good fit to the empirical data, and provides a better estimate of the area under the curve (AUC) than raw data especially when there are few positive cases (i.e., low language). Estimated AUC and confidence intervals were performed on the smoothed curve.

The signal detection analysis yields several measures of discriminability. The measures of interest include AUC, sensitivity, specificity, positive likelihood ratio (LR+), and negative likelihood ratio (LR−). The AUC can be interpreted as the average sensitivity over all points on the curve. An AUC of .5 indicates no discrimination, whereas an AUC of 1 indicates

perfect discrimination. Sensitivity refers to the ability of the predictor to accurately detect children who will have low language skills, whereas specificity refers to the ability to discriminate these children from their average-to-high language peers. If a test has a sensitivity of .70 and a specificity of .80, it accurately captures 70% of children with low language skills and correctly rejects 80% of children with average-to-high language skills.

The positive likelihood ratio (LR+) is the likelihood that a child identified as low language by the predictor is also identified as low language at 36 months. For example, an LR+ of 5 indicates that a child identified as low language on the predictor is 5 times more likely to have low, than average-to-high, language skills at 36 months. Complementarily, LR− is the likelihood that a child identified as average-to-high language on the predictor has low language at 36 months. For example, an LR− of .2 indicates that a child identified as average-to-high language on the predictor variable is 1/5 times as likely to have a low, rather than average-to-high, language skills.

Children were classified as Low Language (LL) if their Language Factor score at 36 months was more than 1 SD below the mean and Average-to-High Language (HL) if their score was at or above 1 SD below the mean. This approach resulted in 13 LL children (8 French-speaking children and 5 English-speaking children; 12.5% of the total sample), and 91 HL children. This is within the expected incidence in the population (American Speech and Hearing Association, 2017). The LL sample had an average score of −1.55 SD below the mean (range = −2.81 to −1.03) whereas as the HL children had an average score of .22 SD above the mean (range= −.99 to +2.36) relative to their sample-specific Language Factor.

Other classification cutoffs (e.g., 1.25 SD below the mean, 1.5 SD below the mean) were considered: first, we visually inspected the resulting ROC curve for smoothness and symmetry and considered the number of children who were identified as LL, which affected the ROC curve and the reliability of the discriminability measures. We also examined multiple indices of discriminability (e.g., sensitivity, specificity, LR+, and LR−) to find the cutoff with the best balance across indices. See Figure 1 for a graphical representation of the discriminability measures across cutoff points. Based on these considerations, we concluded that 1 SD below the mean provided the best balance of smoothness and symmetry with maximum discrimination across indices and a sufficient number of children classified as LL. With more stringent cutoffs, although estimates of sensitivity, specificity, and likelihood ratios appear better, they are less likely to be reliable: the number of children identified is reduced and the curve becomes less smooth and symmetric.

Results

The AUC (see Figure 2) was .83 (95% confidence interval = .72 to .94), indicating good discriminability. Visual inspection suggested high specificity for low scores on the CCT. For example, a cutoff on the CCT predictor at approximately the 10th percentile (less than 21 words correct), yielded a specificity of .95, and a LR+ of 5.6, indicating a child who knew less than 21 words on the CCT was 5.6 times more likely to be LL at 36 months than HL. However, sensitivity and LR− at this cutoff were inadequate at .31 and .73, respectively. On the other side of the curve, higher scores on the CCT evinced excellent sensitivity. For example, a cutoff on the CCT predictor at approximately the 50th percentile (less than 29

words correct) yielded a sensitivity of .92 and a LR− of .12. At this cutoff, all but one LL child at 36 months were captured. However, specificity and LR+ were insufficient at .62 and 2.4, respectively. Because there is no prior evidence to suggest an optimal cutoff, we examined two statistics to maximize the balance of sensitivity and specificity.

First, we determined the cutoff on the CCT that produces the optimal balance of sensitivity and specificity using Youden's J (Youden, 1950). The cut point was 26.5 words (out of 41), which yielded a sensitivity of .85 and specificity of .76. The LR+ was 3.5, indicating that children who knew less than 26.5 words were 3.5 times as likely to be LL at 36 months than to be HL. Complementarily, the LR− was .20, indicating children who knew more than 26.5 words were 1/5 times as likely to be LL at 36 months than to be HL. Second, we evaluated the point closest to the top left corner of the ROC plot (perfect sensitivity and specificity). This yields an optimal cutoff of 24.5 words, with a sensitivity of .77 and a specificity of .84. The LR+ was 4.67, and the LR− was .28. These statistics varied primarily in the relative balance of sensitivity and specificity.

Using the first criterion, CCT scores at 22 months correctly identified 4 out of 5 children referred for services in the English sample and 1 out of 3 children in the French sample. Using the second criterion, CCT scores at 22 months correctly identified 3 out of 5 children referred for services in the English sample and 1 out of 3 children in the French sample. The one child who was consistently not identified in the English sample received services prior to 24 months of age but had language skills that were above average at 36 months. Of the two children not identified in the French sample, one received services for articulation difficulties and had language skills at the classification borderline (1 SD below the mean). The other was diagnosed with expressive, but not receptive, delay. However his score on the Language Factor at 36 months was 2 standard deviations below the mean indicating low language skills relative to his peers. This child was a true "miss" in signal detection terms.

In order to situate these results within the literature, we also conducted a signal detection analysis with MCDI production as the predictor. We evaluated both the empirical ROC curve and a binormal-smoothed curve. The AUC was .77 (95% confidence interval = .63 to .88), indicating fair discriminability. Next, we determined the cutoff on the MCDI that produces the optimal balance of sensitivity and specificity using Youden's J (Youden, 1950). The cut point was 119.5 words (between the 30^{th} and 35^{th} percentiles on the MCDI and approximately the 40^{th} percentile on the IFDC), with a sensitivity of .74 and specificity of . 77. The LR+ was 2.92 and the LR− was .31. The point closest to the top left corner of the ROC plot yielded the same cutoff (see Figure 3). It is noteworthy that this cutoff is considerably higher than cutoffs based on normative considerations, resulting in improved sensitivity but poorer specificity relative to previous reports (e.g., Heilmann, et al., 2005). Because the AIC model contrasts suggested that the 22-month model including both the MCDI and the CCT were equivalent in strength at the group level to the model containing only the CCT, we repeated this analysis using a composite MCDI/CCT approach. Children were classified based on whether they fell below the statistically determined cutoff on both measures. This approach yielded an improvement in specificity (90 to 93%, depending on the CCT criterion), but a reduction in sensitivity (61 to 69%).

Discussion

This study represents the first attempt to estimate the sensitivity and specificity of an assessment of decontextualized vocabulary for prospectively identifying children whose language at 36 months is a full standard deviation below their peers. These were typically developing samples at the time of recruitment with no known risks of language impairment. Therefore our findings should be interpreted as a preliminary indication of the practical significance of this approach.

Sensitivity and specificity were moderately strong although their relative strength depended on the cut point. In general, a cutoff between about 24 and 26 words comprehended (out of 41) yielded the best balance of performance. This corresponded well to both LR+ and LR− estimates. Sensitivity ranged between .76 and .85 and specificity ranged between .77 and .84. Vocabulary size on the CCT at 22 months correctly identified 77 to 85 percent of the "true" cases of low language and correctly rejected 76 to 84 percent of cases of average-to-high performance at 36 months of age. Compared to the MCDI/IFDC, the CCT yielded either higher sensitivity or specificity depending on the cutoff. Finally, sensitivity for the CCT was generally better than in prospective studies with other measures (Frisk, Montgomery, & Boychyn, 2009; Klee et al., 1998; Klee, Pearce, & Carson, 2000; McIntyre et al., 2017; McKean et al., 2016; McKean et al., 2017; Stott, Merricks, Bolton, & Goodyer, 2002; Westerlund, et al., 2006). Our findings are most comparable to McIntyre et al. (2017) and McKean et al. (2017) with the notable exception that we obtain comparable sensitivity and specificity a full two years earlier. Thus a significant empirical contribution is a potential screen for language difficulties that can be used as early as the second year.

General Discussion

This research follows from recent efforts to predict developmental achievements from early vocabulary (Duff, et al., 2015; Kemp et al., 2017; Morgan et al., 2015; Reilly et al., 2010;). In prior work, effect sizes are generally modest and prediction at the individual level is weak (e.g., Westerlund, Berglund, & Eriksson; 2006). Thus, an overarching goal was to overcome these limitations. With this in mind, this paper addresses three primary aims: to predict preschool language skills in typically developing children from a measure of decontextualized vocabulary, to contrast this measure with parent report of comprehension at 16 months and production at 22 months, and to determine how early predictions to preschool language can be made.

Predicting Preschool Language

Both parent reported and decontextualized vocabulary were associated with preschool language. At 16 months of age parent reported production was the strongest predictor of 36-month skills in English whereas parent reported comprehension was the strongest predictor in French. This finding reflects variation across samples in the underlying pattern of correlation between parent reported comprehension and production in the second year and diverse metrics of language ability (vocabulary, MLU, sentence processing) in the third year.

By 22 months, decontextualized vocabulary accounted for more variance in preschool language skills and provided a better balance of fit and parsimony than parent report in both French and English. The unique variance in language skills accounted for at the group level was 20 to 25%: four to five times that reported for large-scale studies using parent-reported vocabulary (Duff, et al., 2015; Morgan et al., 2015; Reilly, et al., 2010).

This observed stability may lie in underlying domain general processes rather than in language itself. This follows from the idea that a characteristic (e.g., language) might appear stable because some other characteristic mediates the relation between measures of language at different points in time (Bornstein et al., 2017). Simple associative processes, which themselves rely on attention and memory, can account for cross-situational word learning (Yu and Smith, 2012) and, hypothetically, yield graded word-referent associations that are strengthened over time and situations. When faced with a choice between a target referent and a perceptually similar distractor (i.e., color, size, saliency) from the same word class and conceptual category, weak, context-bound associations are not sufficient to elicit a correct response. By estimating the number of generalized word-referent relations, decontextualized vocabulary reflects the efficiency of domain general learning. By 22 months of age, the word-world relations that children recognize beyond the context in which they were acquired predict downstream language skills.

Stronger prediction at 22 relative to 16 months is in line with previous work: prediction to language samples has been reported by 18-24 months of age but not earlier (Friend et al., 2012; Westerlund et al., 2006), paralleling an acceleration in word learning late in the second year (Samuelson & McMurray, 2017). This may reflect richer semantic organization with stronger word-world relations that facilitate retrieval of word meaning late in the second year. Alternatively, better prediction at 22 relative to 16 months may be a function of proximity to the 36-month measures (Bornstein, et al., 2017) although previous research suggests stability in decontextualized vocabulary from the second year through at least the beginning of the fourth year (Friend et al., 2018).

Finally, decontextualized vocabulary evinced practical significance for prospectively identifying children with low language skills. Sensitivity and specificity compared favorably to other prospective studies (Frisk et al. 2009; Klee et al., 1998; 2000; Stott et al., 2002; Westerlund, et al., 2006; Wetherby et al., 2003) overcoming the longstanding difficulty of harnessing early vocabulary prior to 30 months of age to predict development at the group and individual levels. This is the first paper to show strong prospective sensitivity as early as two years of age with implications for the early assessment of developmental risk.

Limitations and Directions for Future Research.

First, whereas a strength of this paper is replication across two distinct samples, in Study 3, we collapsed across samples to obtain stable estimates of sensitivity, specificity and likelihood ratios. We expect these estimates to be consistent across languages, but this remains to be empirically tested. Second, we focused on prediction in a sample with no known risk factors thus we identified only a small number of children with low language skills at age three. Whereas the proportion of children identified is consistent with the population incidence, future work is needed to establish norms and cut points to estimate

sensitivity and specificity in high-risk samples (e.g., children from families with low-income/a history of language difficulties). Third, although one could question the component structure of the Language Factor, average scores on the component measures suggest that it adequately classifies low- relative to higher-language children (please see Table S5 and Figures S1 and S2). Finally, although we cannot know how well our research generalizes to other languages or language families, the fact that our findings parallel each other so well in English and in French is encouraging.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge Laura Alaria, Kristi Hendrickson, and Danielle Rosen for assistance in data collection and coding and all of our participant families. This research was supported by NICHD #R01HD468058 and NIDCD #T32DC00736 and does not necessarily represent the views of the National Institutes of Health.

References

Armon-Lotem S, & Meir N (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. International journal of language & communication disorders, 51, 715–731. [PubMed: 26990037]

Bleses D, Vach W, Slott M, Wehberg S, Thomsen P, Madsen TO, & Basbøll H (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. Journal of Child Language, 35, 619–650. doi: 10.1017/S0305000908008714 [PubMed: 18588717]

Bornstein MH, Hahn CS, & Putnick DL (2016). Long-term stability of core language skill in children with contrasting language skills. Developmental Psychology, 52, 704. doi: 10.1037/dev0000111 [PubMed: 26998572]

Conti-Ramsden G, Botting N, & Faragher B (2001). Psycholinguistic markers for specific language impairment (SLI). Journal of Child Psychology and Psychiatry, 42, 741–748. doi: 10.1111/1469-7610.00770 [PubMed: 11583246]

Dapretto M, & Bjork EL (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth. Child Development, 71, 635–648. http://www.jstor.org/stable/1132382 [PubMed: 10953930]

DeAnda S, Bosch L, Poulin-Dubois D, Zesiger P, & Friend M (2016). The language exposure assessment tool: Quantifying language exposure in infants and children. Journal of Speech, Language, and Hearing Research, 59, 1346–1356. doi: 10.1044/2016_JSLHR-L-15-0234

Dethorne LS, Johnson BW, & Loeb JW (2005). A closer look at MLU: What does it really measure? Clinical Linguistics & Phonetics, 19, 635–648. doi: 10.1080/02699200410001716165 [PubMed: 16147407]

Devescovi A, & Caselli M (2007). Sentence repetition as a measure of early grammatical development in Italian. International Journal of Language & Communication Disorders, 42, 187–208. doi: 10.1080/13682820601030686 [PubMed: 17365093]

Duff FJ, Reen G, Plunkett K, & Nation K (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? Journal of Child Psychology and Psychiatry, 56, 848–856. doi: 10.1111/jcpp.12378 [PubMed: 25557322]

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

Dunn LM, Thériault-Whalen CM, & Dunn LM, (1993). Échelle de vocabulaire en images Peabody: série de planches. Toronto: Psycan.

Eriksson M, Westerlund M, & Miniscalco C (2010). Problems and limitations in studies on screening for language delay. Research in Developmental Disabilities, 31, 943–950. doi: 10.1016/j.ridd.2010.04.019 [PubMed: 20483561]

Faul F, Erdfelder E, Lang AG, & Buchner A (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175–191. [PubMed: 17695343]

Fenson L, Bates E, Dale PS, Marchman VA, Reznick JS, & Thal DJ (2007). MacArthur-Bates communicative development inventories. Paul H. Brookes Publishing Company.

Fenson L, Dale PS, Reznick SJ, Bates E, Thal DJ, Pethick SJ, Tomasello M, Mervis CB, & Stiles J (1994). Variability in early communicative development Monographs of the Society for Research in Child Development, 59, i–185. doi: 10.2307/1166093 [PubMed: 8047076]

Fernald A, & Marchman VA (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. Child Development, 83, 203–222. doi: 10.1111/j.1467-8624.2011.01692.x [PubMed: 22172209]

Frank M, Braginsky M, Yurovsky D, & Marchman V (2017). Wordbank: An open repository for developmental vocabulary data. Journal of Child Language, 44, 677–694. doi:10.1017/S0305000916000209 [PubMed: 27189114]

Friend M, & Keplinger M (2003). An infant-based assessment of early lexicon acquisition. Behavior Research Methods, Instruments, & Computers, 35, 302–309. doi: 10.3758/BF03202556

Friend M, & Keplinger M (2008). Reliability and validity of the Computerized Comprehension Task (CCT): Data from American English and Mexican Spanish infants. Journal of Child Language, 35, 77–98. doi: 10.1017/S0305000907008264 [PubMed: 18300430]

Friend M, Schmitt SA, & Simpson AM (2012). Evaluating the predictive validity of the Computerized Comprehension Task: Comprehension predicts production. Developmental Psychology, 48, 136–148. doi: 10.1037/a0025511 [PubMed: 21928878]

Friend M, Smolak E, Liu Y, Poulin-Dubois D, & Zesiger P. (2018). A cross-language study of decontextualized vocabulary comprehension in toddlerhood and kindergarten readiness. Developmental Psychology, 54, 1317–1333. doi:10.1037/dev0000514 [PubMed: 29620386]

Friend M, & Zesiger P (2011). A systematic replication of the psychometric properties of the CCT in three languages: English, Spanish, and French. Enfance, 3, 329–344. doi: 10.4074/S0013754511003041

Frisk V, Montgomery L, Boychyn E, Young R, McLachlan D, & Neufeld J (2009). Why screening Canadian preschoolers for language delays is more difficult than it should be. Infants & Young Children, 22, 290–308. doi: 10.1097/IYC.0b013e3181bc4db6

Ghassabian A, Rescorla L, Henrichs J, Jaddoe VW, Verhulst FC, & Tiemeier H (2014). Early lexical development and risk of verbal and nonverbal cognitive delay at school age. Acta Paediatrica, 103, 70–80. doi: 10.1111/apa.12449. [PubMed: 24117532]

Heilmann J, Weismer SE, Evans J, & Hollar C (2005). Utility of the MacArthur-Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. American Journal of Speech-Language Pathology, 14, 40–51. doi: 10.1044/1058-0360(2005/006) [PubMed: 15966111]

Hendrickson K, Mitsven S, Poulin-Dubois D, Zesiger P, & Friend M (2015). Looking and touching: What extant approaches reveal about the structure of early word knowledge. Developmental Science, 18, 723–735. doi: 10.1111/desc.12250 [PubMed: 25444711]

Hendrickson K, Poulin-Dubois D, Zesiger P, & Friend M (2017). Assessing a continuum of lexical–semantic knowledge in the second year of life: A multimodal approach. Journal of Experimental Child Psychology, 158, 95–111. 10.1016/j.jecp.2017.01.003 [PubMed: 28242363]

Henrichs J, Rescorla L, Schenk JJ, Schmidt HG, Jaddoe VW, Hofman A, ... & Tiemeier H. (2011). Examining continuity of early expressive vocabulary development: The Generation R study. Journal of Speech, Language, and Hearing Research, 54, 854–869. doi: 10.1044/1092-4388(2010/09-0255).

Hewitt LE, Hammer CS, Yont KM, & Tomblin J (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. Journal of Communication Disorders, 38, 197–213. [PubMed: 15748724]

Hirsh-Pasek K, & Golinkoff RM (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension In McDaniel D, McKee C, & Cairns HS (Eds.), Methods for assessing children's syntax (pp. 105–124). Cambridge, MA: MIT Press.

Hoff E (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. Developmental Psychology, 49, 4–14. doi : 10.1037/a0027238 [PubMed: 22329382]

Kern S (2003). Le compte-rendu parental au service de l'évaluation de la production lexicale des enfants français entre 16 et 30 mois. Glossa, 85, 48–62.

Kern S (2007). Lexicon development in French-speaking infants. First Language, 27, 227–250. doi: 10.1177/0142723706075789

Kern S, & Gayraud F (2010). L'inventaire français du développement communicatif. Grenoble: La Cigale.

Klee T, Carson DK, Gavin WJ, Hall L, Kent A, & Reece S (1998). Concurrent and predictive validity of an early language screening program. Journal of Speech, Language, and Hearing Research, 41, 627–641. doi: 10.1044/jslhr.4103.627

Klee T, Pearce K, & Carson DK (2000). Improving the positive predictive value of screening for developmental language disorder. Journal of Speech, Language, and Hearing Research, 43, 821–833. doi: 10.1044/jslhr.4304.821

Klem M, Melby-Lervåg M, Hagtvet B, Lyster SAH, Gustafsson JE, & Hulme C (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. Developmental Science, 18, 146–154. doi: 10.1111/desc.12202 [PubMed: 24986395]

Law J, Boyle J, Harris F, Harkness A, & Nye C (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. Developmental Medicine & Child Neurology, 42, 190–200. doi: 10.1111/j.1469-8749.2000.tb00069.x [PubMed: 10755459]

Law J, & Roy P (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. Child and Adolescent Mental Health, 13, 198–206. doi: 0.1111/j.1475-3588.2008.00503.x

Marchman VA, & Fernald A (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. Developmental Science, 11, F9–F16. doi: 10.1111/j.1467-7687.2008.00671.x [PubMed: 18466367]

Mayor J & Plunkett K (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. Psychological Review, 117, 1–31. doi: 10.1037/a0018130 [PubMed: 20063962]

McIntyre LL, Pelham WE, Kim MH, Dishion TJ, Shaw DS, & Wilson MN (2017). A brief measure of language skills at 3 years of age and special education use in middle childhood. The Journal of Pediatrics, 181, 189–194. doi: 10.1111/j.1475-3588.2008.00503.x [PubMed: 27908645]

McKean C, Law J, Mensah F, Cini E, Eadie P, Frazer K, & Reilly S (2016). Predicting meaningful differences in school-entry language skills from child and family factors measured at 12 months of age. International Journal of Early Childhood, 48, 329–351. doi: 10.1007/s13158-016-0174-0

McKean C, Reilly S, Bavin EL, Bretherton L, Cini E, Conway L, … & Mensah F. (2017). Language Outcomes at 7 Years: Early Predictors and Co-Occurring Difficulties. Pediatrics, e20161684. doi: 10.1007/s13158-016-0174-0 [PubMed: 28179482]

Miller JF, & Chapman RS (1981). The relation between age and mean length of utterance in morphemes. Journal of Speech, Language, and Hearing Research, 24, 154–161. doi: 10.1044/jshr.2402.154

Miller J & Iglesias A (2012). Systematic Analysis of Language Transcripts (SALT), Research Version 2012 [Computer Software]. Middleton, WI: SALT Software, LLC.

Morgan PL, Farkas G, Hillemeier MM, Hammer CS, & Maczuga S (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. Child Development, 86, 1351–1370. doi: 10.1111/cdev.12398 [PubMed: 26283023]

Posada D & Buckley TR (2004). Model selection and model averaging in phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches over Likelihood Ratio Tests. Systematic Biology, 53, 793–808. doi: 10.1080/10635150490522304 [PubMed: 15545256]

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL https://www.R-project.org/.

Reilly S, Wake M, Ukoumunne OC, Bavin E, Prior M, Cini E, … & Bretherton L. (2010). Predicting language outcomes at 4 years of age: Findings from Early Language in Victoria Study. Pediatrics, 126, e1530–e1537. doi: 10.1542/peds.2010-0254 [PubMed: 21059719]

Reznick JS, & Goldfield BA (1992). Rapid change in lexical development in comprehension and production. Developmental Psychology, 28, 406–413. doi: 10.1044/jslhr.4003.556

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, & Müller M (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77. doi: 10.1186/1471-2105-12-77 [PubMed: 21414208]

Samuelson LK, & McMurray B (2017). What does it take to learn a word? WIREs Cognitive Science, 8:e1421. doi: 10.1002/wcs.1421

Stott CM, Merricks MJ, Bolton PF, & Goodyer IM (2002). Screening for speech and language disorders: The reliability, validity and accuracy of the General Language Screen. International Journal of Language & Communication Disorders, 37, 133–151. doi: 10.1080/13682820110116785 [PubMed: 12012612]

Schmitt N (2014). Size and depth of vocabulary knowledge: What the research shows. Language Learning, 64, 913–951. doi: 10.111/lang.12077

Thordardottir ET (2005). Early lexical and syntactic development in Quebec French and English: Implications for cross-linguistic and bilingual assessment. International Journal of Language and Communication Disorders, 40, 243–278. doi: 10.1080/13682820410001729655 [PubMed: 16195189]

Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, & Lohr KN (2015). Screening for speech and language delay in children 5 years old and younger: A systematic review. Pediatrics, 136, 1–14. doi: 10.1542/peds.2014-3889 [PubMed: 26077476]

Westerlund M, Berglund E, & Eriksson M (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. Journal of Speech, Language, and Hearing Research, 49, 237–247. doi: 10.1044/1092-4388(2006/020)

Wetherby AM, Goldstein H, Cleary J, Allen L, & Kublin K (2003). Early identification of children with communication disorders: Concurrent and predictive validity of the CSBS Developmental Profile. Infants & Young Children, 16(2), 161–174.

Youden WJ (1950). Index for rating diagnostic tests. Cancer, 3, 32–35. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3 [PubMed: 15405679]

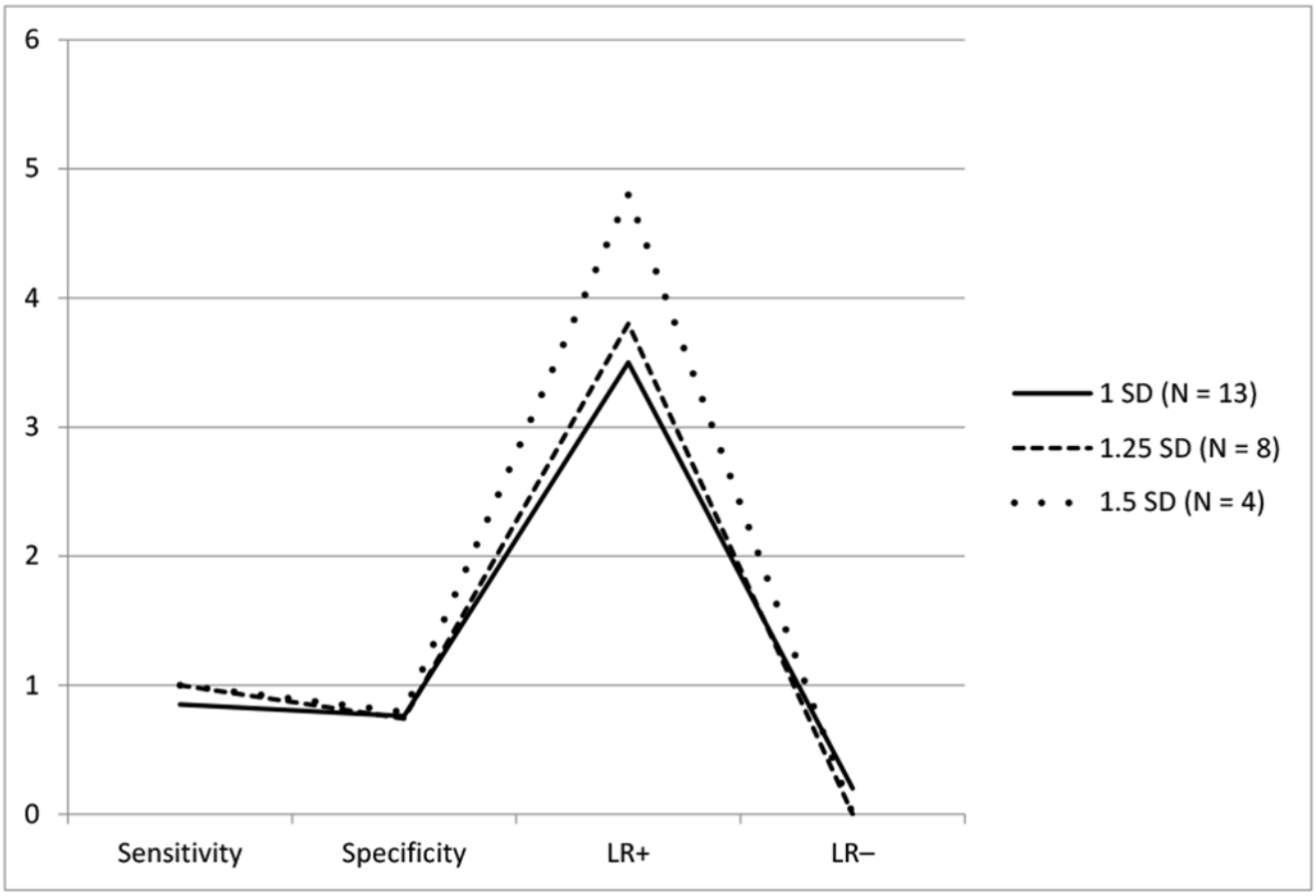

Figure 1.

Measures of sensitivity, specificity, positive and negative likelihood ratio for each of the potential cutoff determinations for low language skill.

Note: Whereas the parameters appear stronger at higher cutoffs, in fact these cutoffs eliminate many children with language skills sufficiently low to be markers of impairment. Further, the shape of the curve under these cutoffs suggests that these parameter estimates are unstable.

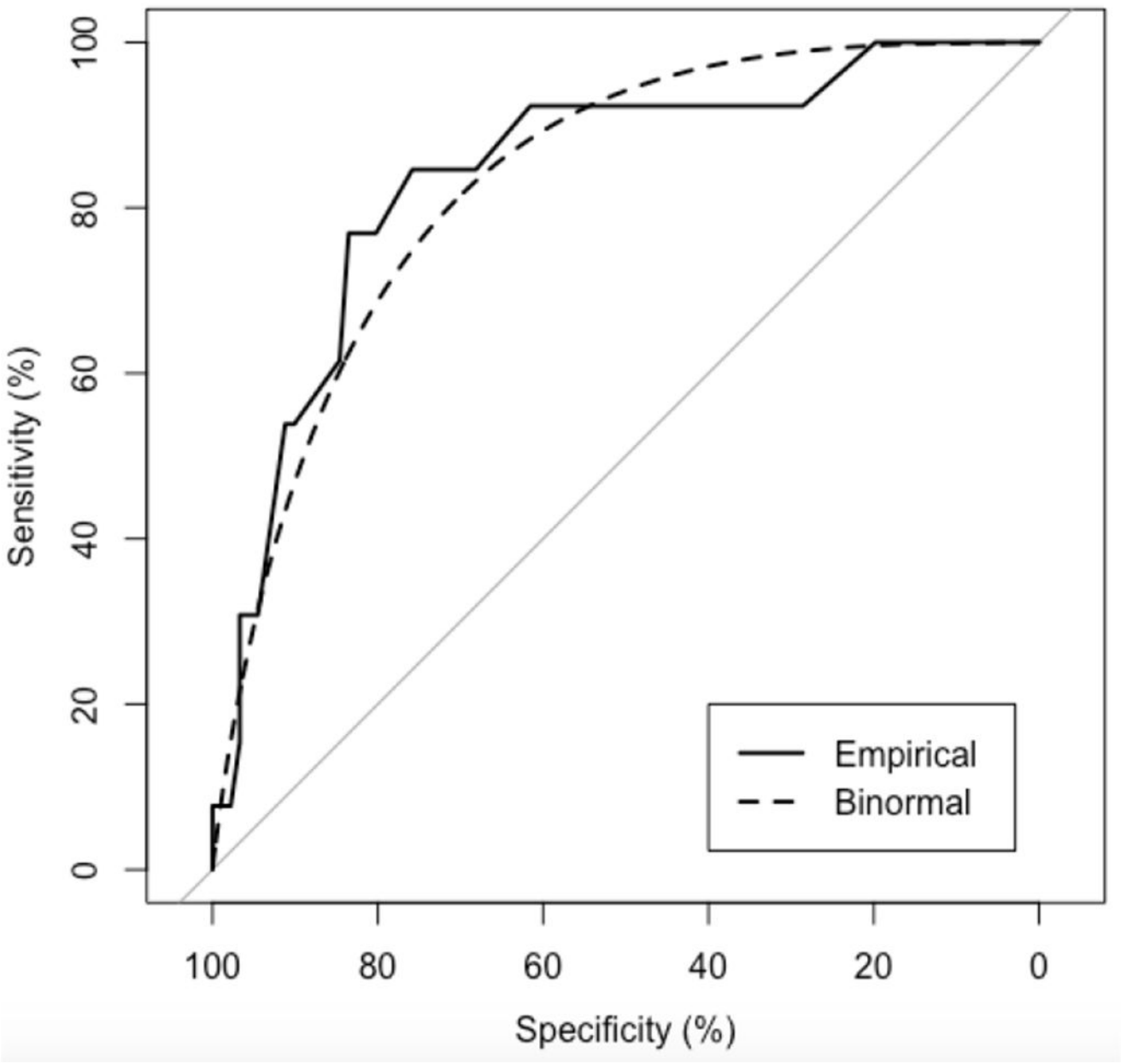

Figure 2.
ROC plot for empirical data and binormal-smoothed ROC curves from Study 3 using the CCT predictor ($N = 104$).

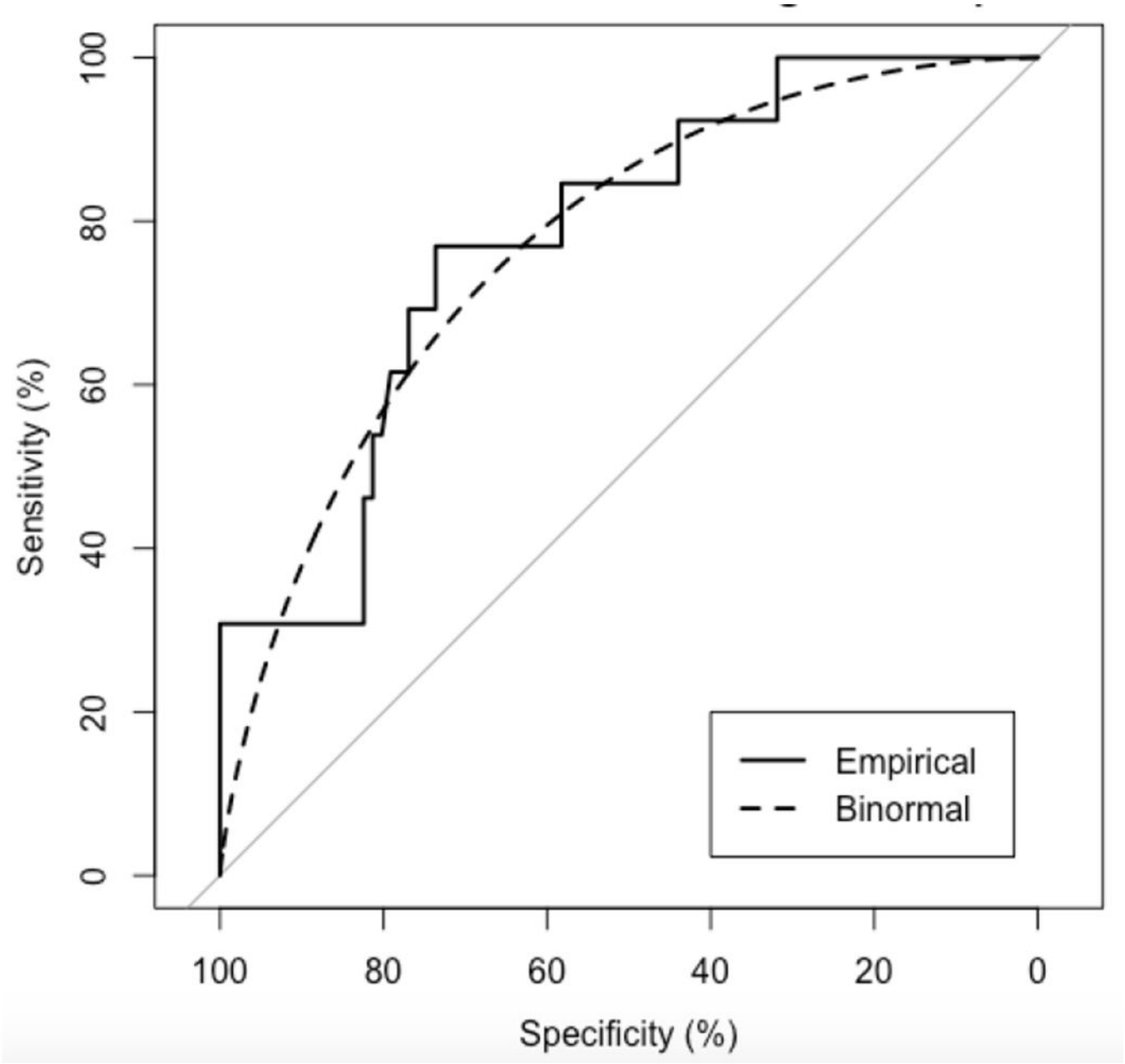

Figure 3.
ROC plot for empirical data and binormal-smoothed ROC curves from Study 3 using the MCDI predictor ($N = 104$).

Table 1.

Distribution of Selected Demographic Characteristics of Participants in Study 1.

	Number (%) of participants		
	Female	**Male**	**Total**
Maternal education			
High School or Less	2 (4.5)	4 (9.1)	6 (13.6)
Some College	7 (15.9)	1 (2.3)	8 (18.2)
College Graduate	7 (15.9)	6 (13.6)	13 (29.5)
Post-Baccalaureate	10 (22.7)	7 (15.9)	17 (38.6)
Family Income			
18,000-40,000	4 (9.1)	2 (4.5)	6 (13.6)
41,000-60,000	1 (2.3)	3 (6.8)	4 (9.1)
61,000-80,000	5 (11.4)	0 (0.0)	5 (11.4)
81,000-100,000	10 (22.7)	8 (18.2)	18 (40.9)
>100,000	6 (13.6)	5 (11.4)	11 (25.0)
Ethnicity			
Asian	0 (0.0)	2 (4.5)	2 (4.5)
Black/not Hispanic	1 (2.3)	0 (0.0)	1 (2.3)
Hispanic	6 (13.6)	1 (2.3)	7 (15.9)
White/not Hispanic	14 (31.8)	14 (31.8)	28 (63.6)
Mixed Race	5 (11.4)	1 (2.3)	6 (13.6)

Notes. Income reported in US dollars. Some values may not sum to 100 due to rounding error.

Table 2.

Bivariate correlations for all predictor and 36-month measures (z-scores) in Study 1 (N=44)

Measure	1	2	3	4	5	6	7	8	9	10
1. Maternal education										
2. Sex	−.05									
3. MCDI comprehension 16 months	.09	−.06								
4. MCDI production 16 months	.21	−.03	.66**							
5. CCT comprehension 16 months	.18	.05	.38*	.28						
6. MCDI production 23 months	.17	−.23	.63**	.65**	.06					
7. CCT comprehension 23 months	.31*	−.24	.47**	.35*	.24[a]	.45*				
8. PPVT comprehension 36 months	.04	−.07	.12	.15	.24	.26	.45**			
9. SR sentences correct 36 months	.35*	−.32*	.34*	.48**	.35*	.50**	.67**	.44**		
10. MLU morphemes 36 months	.10	−.37*	−.07	.26	−.04	.41**	.31*	.27	.51**	

Note.

*p<.05

**p<.01

[a]Three children improved dramatically on the CCT from Wave 1 to Wave 2 ; at Wave 1, they gave between zero and two correct responses and at Wave 2, they gave between 31 and 35 correct responses. These children are not outliers, however they are visually distinct from the rest of the sample and, when they are removed from the data, the test-retest correlation for the CCT at Wave 1 and Wave 2 is significant consistent with previous reports (r(41) = .360, p=.021).

Table 3.

Component Matrix of the Language factor extracted after the factorial analysis with language measures at 36 months in Study 1.

Measure	Component 1
PPVT comprehension 36 months	.711
SR sentences correct 36 months	.854
MLU morphemes 36 months	.763

Extraction Method: Principal Component Analysis

Note. 1 component extracted

Table 4.

Hierarchical regression parameters for 16- and 22-month models, Study 1(*N*=44)

Final Model	**Model 1**				**Model 2**			
Measure	R^2	***SE***	***β***	***p***	R^2	***SE***	***β***	***p***
	.22			**.002**	.42			**<.001**
Sex		.27	−.32	**.022**				
MCDI production 16 months		.14	.38	**.007**				
CCT comprehension 23 months						.13	.49	**<.001**
MCDI production 23 months						.13	.29	**.032**

Excluded Variables						
Measure	B^a	***t***	***p***	***B***	***t***	***p***
Maternal Education	.13	.93	.358	.02	.19	.848
Sex				−.16	−1.37	.177
CCT comprehension 16 months	.17	1.19	.241			
MCDI comprehension 16 months	−.16	−.89	.377			

[a]Unstandardized B values. *Note:* Follow-up test on the effect of sex at 16 months was not significant (*p*=.05

Table 5.

AIC values for models containing combinations of the significant predictors from regression analyses in Study 1

Model	Model ID	AIC
Child Sex + 16-month MCDI Production	1	118.79
23-month MCDI Production	2	116.68
23-month CCT Comprehension	3	108.49
Child Sex + 16-month MCDI Production + 23-month MCDI Production	4	116.75
23-month MCDI Production + 23-month CCT Comprehension	5	105.48
Child Sex + 16-month MCDI Production + 23-month MCDI Production + 23-month CCT Comprehension	6	107.03

Note. Child Sex was included only in models that contain 16-month variables since it did not reach significance in the 23-month model.

Table 6.

Distribution of Selected Demographic Characteristics of Participants in Study 2.

	Number (%) of participants		
	Female	**Male**	**Total**
Maternal education			
High School or Less	9 (15.0)	6 (10.0)	15 (25.0)
Some College	4 (6.7)	9 (15.0)	13 (21.7)
College Graduate	3 (5.0)	3 (5.0)	6 (10.0)
Post-Baccalaureate	14 (23.3)	12 (20.0)	26 (43.3)
Approximate Income			
18,000-40,000	2 (3.3)	0 (0.0)	2 (3.3)
41,000-60,000	1 (1.7)	1 (1.7)	2 (3.3)
61,000-80,000	2 (3.3)	2 (3.3)	4 (6.7)
81,000-100,000	2 (3.3)	3 (5.0)	5 (8.3)
>100,000	10 (16.7)	10 (16.7)	20 (33.3)
Ethnicity			
Asian	0 (0.0)	0 (0.0)	0 (0.0)
Black/not Hispanic	1 (1.7)	2 (3.3)	3 (5.0)
Hispanic	1 (1.7)	0 (0.0)	1 (1.7)
White/not Hispanic	28 (46.7)	28 (46.7)	56 (93.3)
Mixed Race	0 (0.0)	0 (0.0)	0 (0.0)

Notes. 27 participants declined to provide income information. Income reported in Swiss Francs. Some values may not sum to 100 due to rounding error

Table 7.

Bivariate correlations for all predictor and 36-month measures in Study 2 (*z*-scores) (*N*=60)

Measure	1	2	3	4	5	6	7	8	9	10
1. Maternal Education										
2. Sex	−.012									
3. IFDC comprehension 16 months	−.147	−.004								
4. IFDC production 16 months	−.095	.048	.369**							
5. CCT comprehension 16 months	.023	−.061	.236[a]	.058						
6. IFDC production 22 months	.006	.089	.136	.390**	.122					
7. CCT comprehension 22 months	−.042	.185	.149	−.088	.360*	.250[a]				
8. EVIP comprehension 36 months	−.039	.087	.402**	.041	.273*	.194	.509**			
9. SR sentences correct 36 months	−.038	.043	.147	.082	.261*	.276*	.484**	.325*		
10. MLU morphemes 36 months	.022	−.088	.264*	.044	.052	.272	.243	.254	.349**	

Note.

*p<.05

**p<.01

[a]one outlier was discovered at 3 standardized residuals from the regression line describing the relation between the CCT and IFDC comprehension at 16 months. This parent had reported that the child knew all of the words on the IFDC but this was not consistent with the child's performance on the CCT. With this outlier removed, the correlation between the CCT and the IFDC at each wave is significant consistent with previous reports (r(59)=.355, p=.006, and r(59)=.405, p=.001, respectively).

Table 8.

Component Matrix of the Language Factor extracted after the factorial analysis with language measures at 36 months in Study 2.

Measure	Component 1
EVIP comprehension 36 months	.703
SR sentences correct 36 months	.776
MLU morphemes 36 months	.724

Extraction Method: Principal Component Analysis

Note. 1 component extracted

Table 9.

Hierarchical regression parameters for 16- and 22-month models, Study 2 (*N*=60)

Final Model	**Model 1**				**Model 2**			
Measure	R^2	***SE***	***β***	***p***	R^2	***SE***	***β***	***p***
	.12			**.004**	.30			**<.001**
MCDI comprehension 16 months		.13	.37	**.004**				
CCT comprehension 23 months						.11	.56	**<.001**

Excluded Variables	**Model 1**			**Model 2**		
Measure	***B*** [a]	***t***	***p***	***B***	***t***	***p***
Maternal Education	.03	.23	.816	−.001	−.01	.992
Sex	.02	.16	.871	−.09	−.79	.431
CCT comprehension 16 months	.19	1.54	.129			
MCDI production 16 months	−.07	−.50	.623			
MCDI production 23 months				.19	1.74	.087

[a]Unstandardized B values are reported

Table 10.

AIC values for models containing combinations of the significant predictors from regression analyses in Study 1

Model	Model ID	AIC
16-month MCDI Comprehension	1	166.78
23-month CCT Comprehension	2	152.60
16-month MCDI Comprehension +23-month CCT Comprehension	3	147.18