NIH Public Access
Author Manuscript
Child Dev. Author manuscript; available in PMC 2006 January 25.

Published in final edited form as:
Child Dev. 2005 ; 76(4): 856–868.

Concurrent and Predictive Validity of Parent Reports of Child Language at Ages 2 and 3 Years

Heidi M. Feldman
University of Pittsburgh and Children's Hospital of Pittsburgh

Thomas F. Campbell
University of Pittsburgh and Children's Hospital of Pittsburgh

Marcia Kurs-Lasky and **Howard E. Rockette**
University of Pittsburgh

Philip S. Dale
University of Missouri-Columbia

D. Kathleen Colborn
Children's Hospital of Pittsburgh

Jack L. Paradise
University of Pittsburgh and Children's Hospital of Pittsburgh

Abstract

The MacArthur–Bates Communicative Development Inventories (CDI; Dale, 1996; Fenson et al., 1994), parent reports about language skills, are being used increasingly in studies of theoretical and public health importance. This study (N = 113) correlated scores on the CDI at ages 2 and 3 years with scores at age 3 years on tests of cognition and receptive language and measures from parent–child conversation. Associations indicated reasonable concurrent and predictive validity. The findings suggest that satisfactory vocabulary scores at age 2 are likely to predict normal language skills at age 3, although some children with limited skills at age 3 will have had satisfactory scores at age 2. Many children with poor vocabulary scores at 2 will have normal skills at 3.

The optimal method for assessing language in infants and toddlers for both research and clinical purposes remains a matter of intense debate. Standardized clinical or laboratory assessments have the advantage of being based on observable behavior. However, such assessments must be brief to be administered within the attention span of the young child and therefore are likely to prove un-representative of the child's abilities. Moreover, young children may have difficulty cooperating for formal measures administered by strangers in unfamiliar clinical settings. Alternatively, language samples recorded in conversation with an examiner or a parent may be highly influenced by personality and social factors and are difficult and time consuming to

Correspondence to: Heidi M. Feldman.

Correspondence concerning this article should be addressed to Heidi M. Feldman, Children's Hospital of Pittsburgh, One Children's Place, Pittsburgh PA 15213-2583. Electronic mail may be sent to Heidi.Feldman@chp.edu..

This work was supported by grant HD26026 from the National Institute of Child Health and Human Development and the Agency for Healthcare Research and Quality, by grants MCJ 429414 and T73 MC 00036-04 from the Maternal and Child Health Bureau, and by gifts from GlaxoSmithKline and Pfizer Inc. The following individuals provided valuable assistance in the study: Tracy L. Balentine, RN; Beverly S. Bernard, RN, BS; Kathleen A. Cecotti, BS; Sharon A. DiBridge, BS; Irene A. Fabian, CRNP; Nancy J. Guerra, CRNP; Toni L. McKeever; Evelyn S. Palochak, RN; Dayna L. Pitcairn, MA; Karen S. Pourboghrat, BA; Robin E. Rice, BS; Diane L. Sabo, PhD; Joan E. Schiebel, RN; Clyde G. Smith, MS; and Ida Smith. We are indebted to the practicing pediatricians who participated in executing the study and to the parents who completed the questionnaires. We also are grateful to Larry Fenson and the late Elizabeth Bates for helpful consultation regarding the MacArthur-Bates Communicative Development Inventories.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

analyze. For these reasons, parent reports of language and communication are an appealing option because parents have extensive experience with their children under a wide variety of naturalistic situations.

The MacArthur–Bates Communicative Development Inventories (CDI) are a set of parent-report inventories about child language and communication (Fenson et al., 1994; Fenson et al., 1993), designed to yield information on the typical course of language development and on the extent of variation within the population (Fenson et al., 1994). The CDI–Words and Gestures (CDI–WG), appropriate for children ages 8 to 16 months, assesses comprehension and production of vocabulary, communicative gestures, symbolic behavior, and nonverbal imitation. The CDI–Words and Sentences (CDI–WS) evaluates children ages 16 to 30 months and assesses productive vocabulary, knowledge of irregular word forms (e.g., nouns such as *children* and verbs such as *bought*), overgeneralization of word endings to irregular nouns and verbs (e.g., *feets* and *eated*), and syntactic complexity. The CDI–III (Dale, Reznick, Thal, & Marchman, 2001), designed for children ages 30 to 42 months, evaluates productive vocabulary, syntactic maturity, and language use. These inventories are being used increasingly, particularly in large-scale research studies, to address important theoretical issues, such as estimating the relative contributions of genetic versus environmental factors to rates of language development (Dionne, Dale, Boivin, & Plomin, 2003; Price et al., 2000), evaluating the link between early speech perception and later language development (Tsao, Liu, & Kuhl, 2004), and determining the prevalence and predictors of language delays (Horwitz et al., 2003). They have been used in samples of children with developmental disorders to describe the populations (Caselli et al., 1998; Miller, Sedey, & Miolo, 1995) and to address clinical and theoretical issues, such as the stability of developmental profiles over time (Mervis & Robinson, 2000). In addition, they have served as the dependent variables in studies with major clinical and public health implications, such as the impact of maternal prenatal diet on child development (Daniels, Longnecker, Rowland, & Golding, 2004) and the degree of association between duration of otitis media and the development of language skills (Feldman et al., 2003; Feldman et al., 1999) The inventories have been translated into many languages, including Spanish, Italian, and Hebrew. Despite this extensive use, the accuracy of parent reports has been questioned, particularly in studies of children from low-income households (Roberts, Burchinal, & Durham, 1999). Given the widespread use of these measures and the implications for theory and practice, it is important that tests be conducted to examine both the concurrent and the predictive validity of the instruments.

We previously studied measurement properties of the CDI–WG in children ages 10 to 13 months and of the CDI–WS in children ages 22 to 25 months (Feldman et al., 2000). As part of a prospective study of speech, language, cognitive, and psychosocial development in relation to otitis media in the first 3 years of life (Paradise et al., 2000; Paradise et al., 2001; Paradise et al., 1997), we obtained CDIs on a large, sociodemographically diverse sample of healthy children at ages 1, 2, and 3 years (Feldman et al., 2003; Feldman et al., 1999; Feldman et al., 2000). We were able to analyze scores on the CDI in relation to age, sex, race, and maternal education and health insurance status as proxies for socioeconomic status (SES). We found that the continuous scales on both tests showed increasing scores with increasing monthly age, thereby capturing developmental progress in this age range. However, for three of the four continuous scales of the CDI–WG (Phrases Understood, Vocabulary Comprehension, Vocabulary Production) and two of the five continuous scales of the CDI–WS (Word Forms–Irregular, Word Forms–Overgeneralized), children in our study who had less favorable socioeconomic indicators had higher mean scores than those with more favorable indicators. Such findings contradicted the frequent observation that children of low SES typically develop language more slowly than do children of middle or high SES (Dollaghan et al., 1999; Hoff, 2003; Landry, Smith, & Swank, 2002). We therefore recommended caution, both in research and in clinical practice, in using those CDI scales to identify individual children at risk for

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

language deficits, to compare groups of children with different sociodemographic profiles, or to evaluate the effects of interventions.

The goal of the present study was to assess the concurrent validity of the CDI–III and the predictive validity of the three continuous scales of the CDI–WS that have favorable measurement properties (Vocabulary Production, Three Longest Sentences, and Sentence Complexity). A subset of the children from the larger otitis media study referenced earlier had been assessed at age 3 years using two types of data: (a) standardized tests, namely, the McCarthy Scales of Children's Abilities (McCarthy, 1972) as a measure of cognitive ability and the Peabody Picture Vocabulary Test–Revised (PPVT–R; Dunn & Dunn, 1981) as a measure of receptive language, and (b) an analysis of parent–child conversation. The larger study design thus afforded us the opportunity to assess the CDI–III and CDI–WS in relation to these other measures of cognition and language as indexes of concurrent and predictive validity, respectively.

Previous Studies of Parent Reports on Language Development

Available research on the validity of the CDI and other parent report measures has been generally encouraging. In the toddler age group (18 to 30 months), studies have reported substantial correlations between scores on a variety of parent-report measures and scores on measures from concurrent language samples and structured tests (Clark, Jorgensen, & Blondeau, 1995; Dale, 1991, 1996; Dale, Bates, Reznick, & Morisset, 1989; Fenson et al., 1994; Rescorla & Alley, 2001). This finding has been replicated in children speaking different languages (Thal, Jackson-Maldonado, & Acosta, 2000), children with developmental disabilities (Miller et al., 1995), and children who stutter (Ratner & Silverman, 2000). However, other studies have called into question the accuracy of parental report, particularly in under-represented minority groups (Roberts et al., 1998).

Less information is available about the predictive validity of the parent-report measures. Strong predictive validity has been reported for children around age 2 years who are developing typically (Bornstein & Haynes, 1998; Reese & Read, 2000) and for those who have developmental disorders (Miller et al., 1995). Children who score very low on parent-report measures (“late talkers”) are at substantially elevated risk for continuing language impairments (Rescorla, 1989; Rescorla & Alley, 2001), although many low-scoring children catch up in the later preschool years (Dale, Price, Bishop, & Plomin, 2003; Thal, Tobias, & Morrison, 1991; Thal, O'Hanlon, Clemmons, & Fralin, 1999).

Given the challenges of assessing language in toddlers and the importance of careful evaluations of early intervention for children with language delays, further studies of concurrent and predictive validity are warranted. Most studies have concerned parent reports for children below age 2.5 years (or this developmental level for children with language delays). Investigators have assumed that as the child's language becomes more extensive and complex, it will become increasingly more difficult for parents to monitor it accurately.

Methods for Assessing Validity

Several procedures may be used to assess the validity of a new assessment procedure. The first is to determine the degree of association between scores on the new assessment and scores on other accepted evaluations of the same underlying constructs. In this case we can correlate the scores on the parent-report measures with scores on standardized tests of language and cognition or measures of language in conversation. The correlational approach considers the full range of scores in the assessment of validity. If the scores on CDI–III have strong and statistically significant positive associations with scores on other previously validated measures at the same age, these findings would constitute strong evidence of concurrent

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

validity. Similarly, if scores on the CDI–WS have strong and statistically significant positive associations with scores on validated measures at a later age, the findings would constitute strong evidence for predictive validity.

For some research and clinical purposes, such as the evaluation of intervention techniques, parent reports are used to determine which children are likely to develop persistent language delays or disorders over time. To use them in this manner, as a screening tool, it is important to determine the extent to which children who score poorly on the parent reports are the same children whose performance on later standardized or accepted measures meets definitions of language or cognitive delay. Rather than assessing the full range of scores, the focus is on the extent to which the lowest scores predict future problems. For these purposes, the continuous scores on both the predictive measures and the outcome measures must be converted to binary variables, that is, performance below a cutoff on the potential screening measure, and low scores on the later measure that would be considered indicative of the presence of the clinical disorder, in this case, language or cognitive delay. If the screening test were a perfect instrument, all children below the cutoff would develop persistent problems and all children above the cutoff would continue to function in the normal range. However, screening instruments are rarely so successful. In reality, any given cut-point results in proportions of true positives, true negatives, false positives, and false negatives.

A receiver operator characteristic (ROC) plot or curve is a useful technique for summarizing the global diagnostic accuracy of the screening test, irrespective of any specific cutoff point. For a given definition of delay or disorder, the ROC demonstrates how changing the pass–fail criterion for the screening measure, in this case the CDI–WS, alters the proportions of true positives in relation to false positives and true negatives in relation to false negatives (Altman & Bland, 1994c). An ROC plot is generated by first calculating the sensitivity of the screening test (the proportion of children with the disorder who have a positive test or the true positives) and the specificity of the test (the proportion of children without the disorder who have a negative test or the true negatives) at each of a range of potential pass–fail cutoff points, in this case, the various CDI–WS scores (Altman & Bland, 1994a). One then plots sensitivity against (1–specificity), that is, the false positive rate, for each of the various cutoff points. The ROC curve of an excellent screening test rises rapidly toward the upper left corner of the graph, whereas the ROC curve of a test that does not distinguish between children who do and do not develop the disorder courses along a 45-degree diagonal line. The area under the resulting ROC curve provides a global assessment of the performance of the test, or its diagnostic accuracy. If the area under the ROC curve approaches 1.0, the test has excellent diagnostic accuracy. If the area under the ROC curve approaches 0.5, the test is unable to differentiate between those with and without the disorder. Conventionally, the diagnostic accuracy of a test with an area under the curve of 0.60 to 0.70 is considered poor; of 0.70 to 0.80, fair; of 0.80 to 0.90, good; and of ≥ 0.90, excellent.

When a test is used for screening purposes, the cutoff score considered a failure, or positive screen, must be determined. Choice of an appropriate cutoff score depends on the intended clinical or research uses as well as on the test characteristics. In some cases, it is important to identify all individuals with the disorder, even if doing so results in many false positive identifications. In other cases, it is important to reduce the rate of false positives even if doing so results in missing some individuals with the disorder. Also relevant in evaluating a screening test are its predictive values; the positive predictive value is the proportion of children with positive or abnormal parent reports who later are found to have language or cognitive delay, whereas the negative predictive value is the proportion of children with negative or normal parent reports who are later found to score above the selected pass–fail cutoff in language development and in the normal range for cognitive development (Altman & Bland, 1994b).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The present study extends previous work on the validity of the CDI in several ways: It evaluates the concurrent validity of the CDI–III; it evaluates the predictive validity of the CDI–WS over 1 year, longer than in most reported studies; it involves a sample of 113 children from diverse sociodemographic backgrounds, a larger sample than in many (though not all) reported studies (Dale et al., 2003); and it compares the results of parent reports with results of several other measures, including scores on standardized tests and measures from analysis of conversation. In addition to providing correlations between scores on the parent report and scores on the other outcome measures, this study presents summary results of ROC curves for scores on the Vocabulary Production scale of the CDI–WS in relation to language delay as variously defined, and in relation to cognitive delay. The study also presents sensitivities, specificities, positive predictive values, and negative predictive values for different CDI–WS Vocabulary Production pass–fail cutoff scores. We chose to consider the Vocabulary Production scale rather than the other two CDI–WS scales because its range of scores is much broader, thus affording potentially more precise discrimination between levels of performance.

Method

Participants

Participants in this study were apparently healthy children enrolled in the child development/otitis media study referred to previously. Details of the exclusion criteria are presented in detail in previous reports from the larger study (Paradise et al., 2000; Paradise et al., 2001; Paradise et al., 1999; Paradise et al., 1997). In brief, all children presented for primary care at one of eight participating pediatric practice sites in the greater Pittsburgh area between May 1991 and December 1995, and were enrolled by age 2 months. Children were excluded from enrollment if they met any of the criteria that could adversely affect global or language developmental outcome, such as low birth weight, small-for-gestational-age status, a serious neonatal illness or congenital malformation, mother less than 18 years of age, mother known to abuse drugs or alcohol, or placement in foster care. Children were also excluded if a language other than English was spoken at home.

Parents of all participants of the larger study were asked to complete the appropriate version of the CDI when their children reached approximately ages 1 and 2 years. The CDI–III was introduced into the study procedures in March 1994 and therefore was completed by only a subset of all participants.

A major component of the larger study was a randomized clinical trial of early versus delayed tympanostomy tube placement in a subgroup of the participants who met specified criteria regarding the development of persistent middle-ear effusion in the first 3 years of life. The children in the present study (N = 113) all met the criteria for the randomized clinical trial (although the parents of 4 of the children withheld permission for randomization). Table 1 shows selected characteristics of the sample. (Among the children in the larger study who did not meet randomization criteria, only a few had undergone all of the tests required for the present analysis; to avoid potential confounding we did not include those children in the present sample.) Information about maternal education was obtained at enrollment by asking the parent to indicate the mother's highest level of academic achievement. Of the 113 mothers, 12.4% had failed to complete high school, and 11.5% were college graduates. Over half of the children qualified for Medicaid, and approximately one third were Black. This sample contained a greater proportion of children of low SES than did the larger study population as a whole because all of the children in the present sample had had sufficiently persistent middle-ear effusion to qualify them for the larger study's randomized trial and because chronic otitis media is particularly prevalent among socioeconomically disadvantaged children (Paradise et al., 1997). The present sample also contains a greater proportion of children of low SES than did the norming sample for the CDI (Fenson et al., 1993).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Results from the larger study suggest strongly that chronic otitis media within the duration limits we studied does not adversely affect language development. In the randomized clinical trial, we found no differences in indexes of language development at ages 3 and 4 years between children who received tympanostomy tubes promptly after reaching the threshold of otitis media required for randomization and children assigned to undergo a period of watchful waiting, in most of whom otitis resolved and tubes were never inserted (Paradise, Dollaghan, et al., 2003; Paradise, Feldman, et al., 2001; Paradise et al., 2003). Although, in the associational component of the larger study, we found weak negative correlations between the duration of otitis media on the one hand, and scores on certain measures of language development, on the other hand (Feldman et al., 2003; Feldman et al., 1999; Paradise et al., 2000), we believe that those negative associations reflect the likelihood that both otitis media and relatively unfavorable developmental outcomes may be influenced by other common underlying factors. For example, certain parenting practices, such as bottle propping in lieu of personal interaction, might lead to both increased cumulative duration of otitis media and low scores on tests of language functioning. Because of our negative findings at ages 3 and 4 years in the randomized clinical trial, we do not consider the duration of otitis media as a variable in the present study. With respect to the generalizability of the results reported here, we note that most studies of children's language development have failed to consider the children's prior histories of otitis media. Apart from the otitis media histories of the children in the present sample and their relatively low SES, these children seem representative of healthy Pittsburgh-area preschoolers. We therefore think that the results are likely to be generalizable to other healthy preschoolers.

Obtaining the CDIs

The appropriate version of the CDI was either mailed or handed to families 1 to 2 months before each child's first, second, and third birthdays, and the completed forms were usually obtained on or near their birthdays. Families received a bonus for having completed all or most of the study visits during the preceding year, in most instances, after the CDI was completed and returned to study personnel. Although CDI forms are designed to be self-explanatory (Fenson et al., 1993), study personnel assisted any parent who had difficulty in completing the forms independently because of low literacy, additional children in attendance, or a noisy environment. Rarely, it was necessary for study personnel to read the items to the parent or to mark the forms themselves. The date on which a questionnaire was received was considered the date of completion for the purpose of calculating the child's age, and fractions of a month were discounted.

Measures

CDI–WS. Parents or other caregivers completed the CDI–WS when children were two years of age. Analysis of the CDI–WS was restricted to three of the measure's five continuous interval scales: (a) Vocabulary Production, the number of vocabulary items that the child says from a list of 680 words; (b) Three Longest Sentences, the average number of words and morphemes (meaningful parts of words, such as *s* or *ed* at the end of words) in the child's three longest sentences; and (c) Sentence Complexity, the number of mature forms typical of the child's sentences, chosen by the parent from a list of 37 sentence pairs (such as “Kitty sleeps” vs. “Kitty sleep”). The two remaining scales (Word Forms—Irregular, Word Forms—Overgeneralized) had been included in the CDI because of theoretical interests in how children learn correct use of irregular nouns and verbs. However, how the rate of acquisition of these forms, particularly the overgeneralization of morphological structures (e.g., *feets* and *eated*), relates to later developmental standing is unclear. In addition, we previously found that on these scales, children with unfavorable sociodemographic profiles had higher scores than did children with favorable sociodemographic conditions. For these reasons, we decided against evaluating these scales in terms of predictive validity.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

CDI–III. Parents or other caregivers completed the CDI–III when children were age 3 years. The CDI–III was constructed using principles similar to those used for the CDI–WS. The CDI–III has three parts, each yielding a continuous interval scale: (a) Vocabulary, the number of words the child says from a list of 100 words, comprising 45 words from the original CDI–WS and 55 more advanced words selected on the basis of literature review and pilot testing; (b) Sentences, the number of mature forms typical of the child's sentences from a list of 12 pairs, 5 from the original CDI–WS, and 7 new pairs; and (b) Using Language, the total number of affirmative answers to a list of 12 questions about language concepts and use such as, “Does your child ask questions with more than one word that begin ‘what’ or ‘where’?”

Scoring the CDIs. Two trained research assistants completed the scoring independently and entered the data into two separate computer files. A computer program was used to compare the results from the two files, and any scoring discrepancies were reconciled by discussion. Hand scoring rather than optical scanning was used after it was found that the latter resulted in unacceptably high levels of errors, missing data, or both, primarily because adults completing the form had not made pencil marks large enough or dark enough for reliable scanning.

Comprehensive Assessment at Age 3 Years

Formal testing. Children were tested individually at the Children's Hospital of Pittsburgh under standardized conditions. Examiners were qualified professionals or graduate students whose interrater reliability for the tests concerned had been established previously. All examiners were unaware of the sociodemographic status of children's families, children's CDI scores, and children's cumulative duration of middle-ear effusion. The McCarthy Scales of Children's Abilities (McCarthy, 1972) were used to assess cognition and the PPVT–R (Dunn & Dunn, 1981) was used to assess receptive language. (To reduce the length of the testing session at Children's Hospital, some of the children from two of the participating practices who had to travel a substantial distance to participate were administered the PPVT–R at the participating practice office within 2 weeks of their original testing.)

Samples of conversation at age 3. After completion of each child's formal testing, a sample of the child's spontaneous conversation was audio-recorded for about 15 minutes during play with a set of toys, including kitchen utensils, food items, appliances, a miniature play house, furniture, and people. A parent or caregiver familiar to the child was present during the conversational sessions; the examiner was also present at some of the sessions. The adults were instructed to play and talk with the child as they would at home.

The conversations were recorded onto portable cassette recorders (Marantz PMD 201, Marantz Co Inc., Newton, MA) using both a wireless FM microphone (Telex WT-25, Telex Communications Inc., Minneapolis, MN) and a tabletop microphone (Radio Shack PZM, Radio Shack, Division of Tandy Corp., Fort Worth, TX). Transcribers used the tape with the better sound quality. Transcribers were trained research assistants, blind to the children's otitis media histories and sociodemographic characteristics. Samples were transcribed orthographically. Analysis was performed using the Systematic Analysis of Language Transcripts (SALT) computer program together with a set of transcription conventions to assure consistency. One study team investigator, Christine A. Dollaghan, reviewed all transcripts for consistency. After errors were corrected, SALT analyses were used to generate two measures: number of different words, a measure of word diversity, and mean length of utterance in morphemes (MLUm), a measure of sentence length and grammatical complexity.

Analyses

Descriptive statistics were obtained for the scales of the CDI–WS and the CDI–III in the present sample as a whole and in subgroups defined by sex and maternal education. Differences

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

between subgroups were assessed using t tests and analysis of variance (ANOVA). Statistical significance was set at $p<.05$.

To assess concurrent validity, we calculated the degree of association between the continuous scale scores from the CDI–III and the scores on the age 3 assessment measures (McCarthy, PPVT–R, and conversational language sample), using Spearman correlation coefficients. Here, statistical significance was set at $p<.01$ to correct in part for multiple comparisons.

To assess predictive validity, we calculated the degree of association between the continuous scale scores from the CDI–WS and scores on the various age 3 assessment measures, again using Spearman correlation coefficients.

For calculation of the test characteristics of the CDI–WS in relation to the age 3 assessment measures, we first required definitions of the presence or absence of language delay at age 3. Because a consensus on the most useful measure of language delay has not been established, we used two definitions. The parent-report definition was based on the Vocabulary score on the CDI–III, and the conversation-derived definition was based on the number of different words from the conversational sample. There is also no consensus on what constitutes a significant level of delay on these measures. Therefore, we used three thresholds for each of the two definitions: the lowest 10th percentile, 1 *SD* below the mean (approximately the lowest 16th percentile), and the lowest 20th percentile.

For reasons described earlier, we used the Vocabulary Production scale scores on the CDI–WS as the screening test measure of interest. We generated an ROC curve for these CDI–WS scores in relation to each of the two measures of language delay at each of the three levels of performance and calculated the areas under the six respective curves. We then considered in each instance the sensitivity, specificity, positive predictive value, and negative predictive value that would result from each of several CDI–WS pass–fail cutoff scores. We chose the scores based on the norms for the Vocabulary Production scale reported in the test manual (Fenson et al., 1993). The manual calculated the percentile scores for boys and girls separately. However, in the interest of simplicity we chose to use a single cutoff score regardless of gender. Accordingly, for the present analysis we used the arithmetic mean of values for boys and girls at 24 months of age for the 5th, 10th, 15th, 20th, and 25th percentiles, respectively.

We repeated the same procedures for cognitive delay using the McCarthy General Cognitive Index (GCI) (McCarthy, 1972) as the measure of cognitive delay at age 3 years. Because the McCarthy GCI is a standardized instrument, we defined cognitive delay as a score of less than 85 (1 *SD* below the population mean). Only 8 of the 113 children scored less than 85. We did not consider the 10th percentile threshold because the number of children would have been too small for making predictions.

Results

Scores on Predictor and Outcome Measures

In our sample, the mean for Vocabulary Production for girls at age 24 months was 324, which was approximately the median for girls from the norming sample at age 23 months and the 40th percentile for girls at age 24 months (Fenson et al., 1994). The mean in our sample for boys was 265, which fell close to the median of 276 for boys in the norming sample.

Table 2 shows mean scores and standard deviations on the scales of the CDI–WS at age 2 years and of the CDI–III at age 3 years in relation to children's sex and to maternal education as a measure of SES. Girls had significantly higher scores than boys on all scales except the Sentence Complexity scale of the CDI–WS. The differences across subgroups according to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

maternal education were statistically significant for the Vocabulary Production and Three Longest Sentences scales of the CDI–WS and for the Sentences and Using Language scales of the CDI–III.

The mean scores of the children on the standardized tests and on the parent–child conversational measures at age 3 years did not differ appreciably from the mean scores of the entire group of 402 tested randomized children (Paradise et al., 2001). Mean scores on the McCarthy GCI (*M* = 100.0, *SD* = 14.6) and the McCarthy Verbal scale (*M* = 49.1, *SD* = 9.1) approached the respective population mean scores and standard deviations. However, the mean score on the PPVT (*M* = 93.1, *SD* = 14.3) was below the population mean score, presumably as a reflection of the relatively low SES of the study sample. The mean number of different words for the present sample at age 3 was 127.2 (*SD* = 33.6), and the mean MLUm was 2.8 (*SD* = 0.7).

Correlations Between CDI Scores and Scores on Measures Obtained at Age 3 Years

Table 3 shows the correlations between scores on the CDI–WS at age 2 years and scores on the corresponding scales of the CDI–III at age 3 years. All of the correlations were statistically significant.

Table 4 summarizes correlations between scores on both the CDI–III and the CDI–WS, and scores obtained at age 3 years on the three standardized tests (McCarthy GCI, McCarthy Verbal scale, and PPVT–R) and on the conversational language measures (number of different words and MLUm). Correlations between scores on all three scales of the CDI–III and scores on each of the age 3 standardized tests and conversational language measures were statistically significant. Correlations between scores on the CDI–WS Vocabulary Production and Three Longest Sentences scales and scores on all three age 3 standardized tests and conversational language measures were also statistically significant. However, correlations between scores on the CDI–WS Sentence Complexity scale and scores on all age 3 measures were not significant. It should be noted that scores on the latter scale were available for only 80 of the 113 children because of difficulty experienced by some parents in completing all of the scale's items.

ROC Curves

The areas under the ROC curves for CDI–WS Vocabulary Production scores in relation to parent-reported language delay—defined, respectively, as Vocabulary score on the CDI–III below the 10th percentile, or 1 *SD* below the mean, or below the 20th percentile—were 0.79, 0.77, and 0.78, respectively. The areas under the ROC curves for CDI–WS Vocabulary Production scores in relation to the conversation-derived measure of language delay—defined, respectively, as the number of different words in the age 3 conversational sample below the 10th percentile, or 1 *SD* below the mean, or below the 20th percentile—were 0.76, 0.77, and 0.72, respectively. The area under the ROC curve for CDI–WS Vocabulary Production scores in relation to cognitive delay at age 3—defined as 1 *SD* below the mean on the McCarthy GCI—was 0.83. Based on these analyses and in the interest of uniformity, we chose to define language delay, for both the parent-reported and conversation-derived measures, as 1 *SD* below the mean.

Test Characteristics of the CDI–WS Vocabulary Production Score at Age 2 Years

Language delay at age 3 years. Table 5 shows the specificity, sensitivity, positive predictive value, and negative predictive value of different, arbitrarily chosen cutoff scores on the Vocabulary Production scale of the CDI–WS in relation to the two definitions of language delay at age 3 years. At each cutoff level shown, specificity was higher than sensitivity, and negative predictive value was higher than positive predictive value. In addition, negative predictive value remained relatively high at each of the levels.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Cognitive delay at age 3 years. Table 6 shows the sensitivity, specificity, positive predictive value, and negative predictive value of the same arbitrarily chosen cutoff scores in relation to cognitive delay at age 3 years. At all cutoff scores, positive predictive value was low and negative predictive value was high.

Discussion

Because of the increasing popularity of the CDI in large-scale research studies on child language and the persistent controversies regarding the accuracy of parent reports of language development, we assessed the concurrent validity of the CDI–III at age 3 years and the predictive validity of the CDI–WS at age 2 years by evaluating the degree of association between scores on these measures and scores at age 3 years on five measures: the McCarthy GCI, the McCarthy Verbal scale, and the PPVT–R, and the number of different words and MLUm derived from analyses of parent–child conversations. Regarding concurrent validity, we found that correlations between scores on all three scales of the CDI–III and scores on the other measures obtained at the same age were positive and statistically significant. These results suggest that the CDI–III has reasonable concurrent validity. Regarding predictive validity, correlations between scores on two of three scales of the CDI–WS and scores on the five measures obtained at age 3 years were positive and statistically significant. ROC curves for scores on the Vocabulary Production scale of the CDI–WS in relation to two definitions for language delay at age 3 years and in relation to cognitive delay at age 3 years indicate that this CDI–WS scale provided fair to good diagnostic accuracy overall. However, whereas the negative predictive values of scores on this scale were relatively high, the positive predictive values were low.

The results of this study are consistent with those in previous reports. The predominance of the evidence suggests that parents are reasonably good informants about their child's expressive language development, at least beginning from ages 18 to 30 months, when contemporaneous validity is established by diary study (Robinson & Mervis, 1999) or by formal testing (Rescorla & Alley, 2001; Rescorla, Roberts, & Dahlsgaard, 1997). Previous studies have found that the accuracy of parent reports about language comprehension skills at several ages is poorer than that about language production skills (Feldman et al., 2000; Ireton & Glascoe, 1995; Westerlund & Sundelin, 2000). The CDI–WS does not include an assessment of comprehension, precisely because of the poor validity of parent reports in this subdomain at this age.

Concerns have been raised about the accuracy of reports by parents of low SES in particular. In one study, African American parents of low SES appeared to underestimate the size of their children's vocabularies (Roberts, Burchinal, & Durham, 1999). In another study, parents of low SES appeared to overestimate production of irregular verbs and overregularization—measures therefore not included in the present study—but not to overestimate productive vocabulary or sentence complexity at the same ages (Feldman et al., 2000). On the other hand, in a New Zealand study, parents of low SES provided results of high reliability and validity (Reese & Read, 2000). In the present study we found that at both ages 2 and 3 years, scores on the various scales were in the predicted direction, in that girls performed better than boys and children of upper SES performed better than children of lower SES.

The correlation coefficients for concurrent validity were very similar to the correlations between scores on the CDI–III Vocabulary scale and scores on the McCarthy Verbal scale reported by Oliver et al. (2002) in a sample of 85 British 3-year-olds. These results are encouraging for the appropriateness of using the CDI–III in large-sample studies.

The predictive validity of the CDI–WS at age 2 years, as measured by correlation coefficients in relation to laboratory measures at age 3 years, was moderate, although the correlations with

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

those measures were lower than those of the CDI–III with the same measures. As part of the assessment of predictive validity, we evaluated the Vocabulary Production score of the CDI–WS at age 2 years in relation to different definitions of language delay and in relation to cognitive delay at age 3 years. In these analyses, we divided children into two categories on the basis of positive or negative outcomes, that is, the presence or absence of arbitrarily defined language or cognitive delay. The estimated prevalence of language delay at age 3 differed as a function of the definition used. For example, using the threshold of 1 *SD* below the mean on the CDI–III, 26.4% of the children were categorized as having language delay, whereas using the threshold of 1 *SD* below the mean on the number of different words from the conversational sample, 16.0% were categorized as having language delay. We generated ROC curves for the Vocabulary Production score at age 2 as a screening test in relation to different definitions and different thresholds of language and cognitive delay at age 3. In each case, the area under the ROC curve ranged from 0.72 to 0.83. These values indicate that the screening test has fair to good diagnostic accuracy.

Irrespective of the thresholds chosen to define language and cognitive delay, we found that the negative predictive value of the CDI–WS Vocabulary Production scale was consistently higher than the positive predictive value. For example, regarding cognitive delay at age 3 years, 90% of the children with scores above the 5th percentile at age 2 years scored more than 1 *SD* below the mean at age 3 years, but so did 37% of children in the lowest 5th percentile at age 2 years. The findings in the present study are thus similar to those reported in a large sample of twins in the United Kingdom by Dale et al. (2003). In that sample, whereas 91.5% of a group of children who seemed to be developing normally at age 2 years had normal language development at age 4 years, 56% of a group of children with language delay at age 2 years had normal language development at age 3 years and 60% had normal language development at age 4 years. As noted earlier in the discussion of correlational measures of predictive validity, the present results cannot distinguish between the possibility that some screening tests, such as the CDI–WS, are less discriminating than other measures and the possibility that, regardless of assessment method, early performance is only loosely associated with later abilities. The high specificity and negative predictive value of the CDI–WS suggest that success in the early stages of language development tends to bode well, in that most children developing at normal rates early in life continue to do so. On the other hand, the limited sensitivity and positive predictive value of the test suggest that some children with seemingly normal early development show delays at later ages, whereas many children whose initial developmental progress is slow increase their rate of development to catch up with previously typically developing peers. The rate of early language development thus appears to be an imperfect predictor of the rate of later language development and of the ultimate level of language skills in the children.

In summary, the concurrent validity of the CDI–III and the predictive validity of two scales of the CDI–WS were, in general, fair to good. Parents, irrespective of SES, were generally able to provide accurate information about their children's language by age 2 years. The Vocabulary Production scales of the CDI–WS at age 2 years had moderately high specificity and negative predictive value for language delay and cognitive delay at age 3 years, but limited sensitivity and limited positive predictive value. We thus concur with Dale et al. (2003) that prediction of outcomes at ages 3 or 4 years on the basis of only the CDI–WS Vocabulary Production data collected when the children are age 2 years is too uncertain to be used for many clinical or research purposes.

Future research should test whether measures of early child language other than the size of the vocabulary or the mean length of sentences provide more accurate predictions of which children will have persistent delays. In addition, future studies should consider the contribution of other factors, such as a positive family history of language disorders or delays, to the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

prediction of language delay. The findings of such research might contribute to theories of language development as well as assist clinicians and public planners in designing accurate screening procedures.

References

Altman DG, Bland JM. Statistics notes: Diagnostic tests 1: Sensitivity and specificity. British Medical Journal 1994a;308:1552. [PubMed: 8019315]

Altman DG, Bland JM. Statistics notes: Diagnostic tests 2: Predictive values. British Medical Journal 1994b;309:102. [PubMed: 8038641]

Altman DG, Bland JM. Statistics notes: Diagnostic tests 3: Receiver operating characteristic plots. British Medical Journal 1994c;309:188. [PubMed: 8044101]

Bornstein MH, Haynes OM. Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. Child Development 1998;69:654–671. [PubMed: 9680678]

Caselli MC, Vicari S, Longobardi E, Lami L, Pizzoli C, Stella G. Gestures and words in early development of children with down syndrome. Journal of Speech, Language, & Hearing Research 1998;41:1125–1135.

Clark JG, Jorgensen SK, Blondeau R. Investigating the validity of the clinical linguistic auditory milestone scale. International Journal of Pediatric Otorhinolaryngology 1995;31:63–75. [PubMed: 7537257]

Dale PS. The validity of a parent report measure of vocabulary and syntax at 24 months. Journal of Speech, Language, & Hearing Research 1991;34:565–571.

Dale, PS. Parent report assessment of language and communication. In: Cole, KN.; Dale, PS.; Thal, DJ., editors. Assessment of communication and language. Brookes; Baltimore: 1996. p. 161-182.

Dale PS, Bates E, Reznick JS, Morisset C. The validity of a parent report instrument of child language at twenty months. Journal of Child Language 1989;16:239–249. [PubMed: 2760125]

Dale PS, Price TS, Bishop DVM, Plomin R. Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. Journal of Speech, Language, & Hearing Research 2003;46:544–560.

DalePSReznickJSThalDMarchmanVAA parent report measure of language development for three-year-olds. 2001Unpublished manuscript, University of Missouri–Columbia

Daniels JL, Longnecker MP, Rowland AS, Golding J. Fish intake during pregnancy and early cognitive development of offspring. Epidemiology 2004;15:394–402. [PubMed: 15232398]

Dionne G, Dale PS, Boivin M, Plomin R. Genetic evidence for bidirectional effects of early lexical and grammatical development. Child Development 2003;74:394–412. [PubMed: 12705562]

Dollaghan CA, Campbell TF, Paradise JL, Feldman HM, Janosky JE, Pitcairn DN, et al. Maternal education and measures of early speech and language. Journal of Speech, Language, & Hearing Research 1999;42:1432–1443.

Dunn, LM.; Dunn, LM. Peabody Picture Vocabulary Test–Revised. American Guidance Service; Circle Pines, MN: 1981.

Feldman HM, Dollaghan CA, Campbell TF, Colborn DK, Janosky J, Kurs-Lasky M, et al. Parent-reported language skills in relation to otitis media during the first 3 years of life. Journal of Speech Language, & Hearing Research 2003;46:273–287.

Feldman HM, Dollaghan CA, Campbell TF, Colborn DK, Kurs-Lasky M, Janosky JE, et al. Parent-reported language and communication skills at ages 1 and 2 years in relation to otitis media in the first two years of life. Pediatrics 1999;104:e52. [PubMed: 10506277]

Feldman HM, Dollaghan CA, Campbell TF, Kurs-Lasky M, Janosky JE, Paradise JL. Measurement properties of the MacArthur Communicative Development Inventories at ages 1 and 2 years. Child Development 2000;71:310–322. [PubMed: 10834466]

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ. Variability in early communicative development. Monographs of the Society for Research in Child Development 1994;59:1–185. [PubMed: 7845413]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Fenson, L.; Dale, PS.; Reznick, JS.; Thal, D.; Bates, E.; Hartung, JP., et al. The MacArthur Communicative Development Inventories: User's guide and technical manual. Singular; San Diego, CA: 1993.

Hoff E. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. Child Development 2003;74:1368–1378. [PubMed: 14552403]

Horwitz SM, Irwin JR, Briggs-Gowan MJ, Bosson Heenan JM, Mendoza J, Carter AS. Language delay in a community cohort of young children. Journal of the American Academy of Child & Adolescent Psychiatry 2003;42:932–940. [PubMed: 12874495]

Ireton H, Glascoe FP. Assessing children's development using parents' reports. The Child Development Inventory. Clinical Pediatrics 1995;34:248–255. [PubMed: 7543037]

Landry SH, Smith KE, Swank PR. Environmental effects on language development in normal and high-risk child populations. Seminars in Pediatric Neurology 2002;9:192–200. [PubMed: 12350040]

McCarthy, D. McCarthy Scales of Children's Abilities. Psychological Corp; San Antonio, TX: 1972.

Mervis CB, Robinson BF. Expressive vocabulary ability of toddlers with Williams syndrome or Down syndrome: A comparison. Developmental Neuropsychology 2000;17:111–126. [PubMed: 10916578]

Miller JF, Sedey AL, Miolo G. Validity of parent report measures of vocabulary development for children with Down syndrome. Journal of Speech, Language, & Hearing Research 1995;38:1037–1044.

Oliver B, Dale PS, Saudino KJ, Petrill SA, Pike A, Plomin R. The validity of a parent-based assessment of the nonverbal cognitive abilities of three-year-olds. Early Child Development and Care 2002;17:337–348.

Paradise JL, Dollaghan CA, Campbell TF, Feldman HM, Bernard BS, Colborn DK, et al. Language, speech sound production, and cognition in three-year-old children in relation to otitis media in their first three years of life. Pediatrics 2000;105:1119–1130. [PubMed: 10790473]

Paradise JL, Dollaghan CA, Campbell TF, Feldman HM, Bernard BS, Colborn DK, et al. Otitis media and tympanostomy tube insertion during the first three years of life: Developmental outcomes at the age of four years. Pediatrics 2003;112:265–277. [PubMed: 12897272]

Paradise JL, Feldman HM, Campbell TF, Dollaghan CA, Colborn DK, Bernard BS, et al. Effect of early or delayed insertion of tympanostomy tubes for persistent otitis media on developmental outcomes at the age of three years. New England Journal of Medicine 2001;344:1179–1187. [PubMed: 11309632]

Paradise JL, Feldman HM, Campbell TF, Dollaghan CA, Colborn DK, Bernard BS, et al. Early versus delayed insertion of tympanostomy tubes for persistent otitis media: Developmental outcomes at the age of three years in relation to prerandomization illness patterns and hearing levels. Pediatric Infectious Disease Journal 2003;22:309–314. [PubMed: 12690269]

Paradise JL, Feldman HM, Colborn DK, Campbell TF, Dollaghan CA, Rockette HE, et al. Parental stress and parent-rated child behavior in relation to otitis media in the first three years of life. Pediatrics 1999;104:1264–1273. [PubMed: 10585976]

Paradise JL, Rockette HE, Colborn DK, Bernard BS, Smith CS, Kurs-Lasky M, et al. Otitis media in 2253 Pittsburgh-area infants: Prevalence and risk factors during the first two years of life. Pediatrics 1997;99:318–333. [PubMed: 9041282]

Price TS, Eley TC, Dale PS, Stevenson J, Saudino K, Plomin R. Genetic and environmental covariation between verbal and nonverbal cognitive development in infancy. Child Development 2000;71:948–959. [PubMed: 11016558]

Ratner NB, Silverman S. Parental perceptions of children's communicative development at stuttering onset. Journal of Speech, Language, & Hearing Research 2000;43:1252–1263.

Reese E, Read S. Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. Journal of Child Language 2000;27:255–266. [PubMed: 10967887]

Rescorla L. The Language Development Survey: A screening tool for delayed language in toddlers. Journal of Speech & Hearing Disorders 1989;54:587–599. [PubMed: 2811339]

Rescorla L, Alley A. Validation of the language development survey (LDS): A parent report tool for identifying language delay in toddlers. Journal of Speech, Language, & Hearing Research 2001;44:434–445.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Rescorla L, Roberts J, Dahlsgaard K. Late talkers at 2: Outcome at age 3. Journal of Speech, Language, & Hearing Research 1997;40:556–566.

Roberts JE, Burchinal M, Durham M. Parents' report of vocabulary and grammatical development of African American preschoolers: Child and environmental associations. Child Development 1999;70:92–106. [PubMed: 10191517]

Roberts JE, Burchinal MR, Zeisel SA, Neebe EC, Hooper SR, Roush J, et al. Otitis media, the caregiving environment, and language and cognitive outcomes at 2 years. Pediatrics 1998;102:346–354. [PubMed: 9685437]

Robinson BF, Mervis CB. Comparing productive vocabulary measures from the CDI and a systematic diary study. Journal of Child Language 1999;26:177–185. [PubMed: 10217894]

Thal D, Jackson-Maldonado D, Acosta D. Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. Journal of Speech, Language, & Hearing Research 2000;43:1087–1100.

Thal DJ, O'Hanlon L, Clemmons M, Fralin L. Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. Journal of Speech, Language, & Hearing Research 1999;42:482–496.

Thal D, Tobias S, Morrison D. Language and gesture in late talkers: A 1-year follow-up. Journal of Speech, Language, & Hearing Research 1991;34:604–612.

Tsao FM, Liu HM, Kuhl PK. Speech perception in infancy predicts language development in the second year of life: A longitudinal study. Child Development 2004;75:1067–1084. [PubMed: 15260865]

Westerlund M, Sundelin C. Screening for developmental language disability in 3-year-old children. Experiences from a field study in a Swedish municipality. Child: Care, Health & Development 2000;26:91–110.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Distribution of Selected Sociodemographic Characteristics of Study Participants

Characteristic	Boys (*n* = 65)	Girls (*n* = 48)	Total (*N* = 113)
			Number (%) of participants
Maternal education			
<High school graduate	6 (9.2)	8 (16.7)	14 (12.4)
High school graduate	17 (26.2)	11 (22.9)	28 (24.8)
High school graduate+	34 (52.3)	24 (50.0)	58 (51.3)
College graduate	8 (12.3)	5 (10.4)	13 (11.5)
Health insurance			
Medicaid	36 (55.4)	27 (56.2)	63 (55.8)
Private insurance	29 (44.6)	20 (41.7)	49 (43.4)
None	0 (0)	1 (2.1)	1 (0.9)
Race			
Black	20 (30.8)	15 (31.3)	35 (31.0)
White	43 (66.2)	31 (64.6)	74 (65.5)
Other	2 (3.1)	2 (4.2)	4 (3.5)

Note. **"High school graduate +" indicates some college or technical school beyond high school or an associate degree. "College graduate" indicates completion of baccalaureate with or without completion of postbaccalaureate study.**

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Mean Scores (and Standard Deviations) on the Scales of the CDI–WS at Age 2 Years and the CDI–III at Age 3 Years

		CDI–WS			CDI–III	
Group	**Subgroup**	**Vocabulary production**	**Three Longest Sentences**	**Sentence Complexity**	**Vocabulary**	**Sentence**
Total (N = 80–113)[a]		260.6 (173.3)	3.2 (1.7)	8.3 (7.0)	59.3 (26.1)	6.5 (3.7)
Sex	Boys (n = 40–65)[a]	216.5 (165.6)	2.7 (1.6)	6.9 (5.6)	52.4 (27.0)	5.5 (3.9)
	Girls (n = 40–48)[a]	320.2 (167.2)**	3.8 (1.7)***	9.6 (8.0)	68.6 (21.9)***	8.0 (3.0)***
Maternal education	<High school (n = 7–14)[a]	55.1 (111.2)	2.1 (1.1)	7.9 (8.1)	44.6 (23.0)	3.7 (3.3)
	High school (n = 61–86)[a]	279.8 (179.3)	3.3 (1.8)	8.8 (7.0)	60.6 (27.0)	6.8 (3.7)
	College graduate (n = 12–13)[a]	257.0 (158.4)*	3.4 (1.4)*	5.8 (6.5)	66.0 (18.5)	7.7 (2.9)**

Note. **CDI–WS = MacArthur–Bates Communicative Development Inventories–Words and Sentences; CDI–III = CDI designed for children ages 30 to 42 months of age.**

[a] Variable number of respondents for different scales.

* p<.05.

** p<.01.

*** p<.001.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Correlations Between Scores on the CDI–WS at Age 2 Years and the CDI–III at Age 3 Years

	Scores on CDI–III		
	Vocabulary score	**Sentence score**	**Language score**
CDI–WS			
Vocabulary Production	.58***	.52***	.70***
Three Longest Sentences	.47***	.54***	.53***
Sentence Complexity	.39***	.37***	.51***

Note. **CDI–WS = MacArthur–Bates Communicative Development Inventories–Words and Sentences; CDI–III = CDI designed for children ages 30 to 42 months.**

*** $p<.001$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Correlations Between Scores on the CDI–WS at Age 2 Years and the CDI–III at Age 3 Years, and Scores on the Standardized Tests and Measures from Conversation at Age 3 Years

	Scores on standardized tests			Measures from conversation	
	McCarthy–General Cognitive Index	McCarthy–Verbal	PPVT	Number of different words	Mean length of utterance–morphemes
CDI–III					
Vocabulary	.53***	.49***	.41***	.35***	.33***
Sentences	.56***	.48***	.49***	.41***	.42***
Using Language	.52***	.47***	.49***	.26**	.31***
CDI–WS					
Vocabulary Production	.38***	.37***	.32***	.31**	.28**
Three Longest Sentences	.39***	.36***	.38***	.38***	.37***
Sentence Complexity	.14	.14	.15	.04	–.01

Note. **PPVT = Peabody Picture Vocabulary Test; CDI–WS = MacArthur–Bates Communicative Development Inventories–Words and Sentences; CDI–III = CDI designed for children ages 30 to 42 months.**

** $p<.01$.

*** $p<.001$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) for Language Delay in Relation to CDI–WS Vocabulary Production Scores at Age 2 Years

Score on vocabulary production at 2 years	Sensitivity	Specificity	PPV	NPV
Parent-reported language delay (vocabulary at 3 years <*M* – 1.0 *SD*)				
<59 (5th percentile)	0.39	0.95	0.73	0.81
<100 (10th percentile)	0.50	0.90	0.64	0.83
<130 (15th percentile)	0.57	0.85	0.57	0.85
<156 (20th percentile)	0.61	0.79	0.52	0.85
<187 (25th percentile)	0.64	0.74	0.47	0.85
Conversation-derived language delay (number of different words at 3 years <*M* – 1.0 *SD*)				
<59 (5th percentile)	0.41	0.91	0.47	0.89
<100 (10th percentile)	0.53	0.85	0.41	0.90
<130 (15th percentile)	0.59	0.80	0.36	0.91
<156 (20th percentile)	0.59	0.74	0.30	0.90
<187 (25th percentile)	0.71	0.71	0.32	0.93

Note. **CDI–WS = MacArthur–Bates Communicative Development Inventories–Words and Sentences.**

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 6

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value(NPV) for Cognitive Delay in Relation to CDI – WS Vocabulary Production Scores at Age 2 Years

	Cognitive delay (McCarthy GCI at 3 years < 85)			
Score on vocabulary production at 2 years	**Sensitivity**	**Specificity**	**PPV**	**NPV**
< 59 (5th percentile)	0.63	0.90	0.33	0.97
< 100 (10th percentile)	0.75	0.84	0.27	0.98
< 130 (15th percentile)	0.75	0.78	0.21	0.97
< 156 (20th percentile)	0.75	0.72	0.18	0.97
< 187 (25th percentile)	0.88	0.68	0.18	0.99

Note. **CDI – WS = MacArthur – Bates Communicative Development Inventories – Words and Sentences; GCI = General Cognitive Index.**