Research Article

Production and Maternal Report of 16- and 18-Month-Olds' Vocabulary in Low- and Middle-Income Families

Joan E. Furey
The College of Wooster, Wooster, OH

Purpose: To compare maternal report of children's vocabularies on the MacArthur Communicative Development Inventories Words and Gestures form (CDI:WG; Fenson et al., 1993) with spontaneous production data in both low- and middle-income families.
Method: As part of a longitudinal investigation, language samples were gathered from 23 mother–child dyads based on Stoel-Gammon's (1987) protocol for the Language Production Scale when the children were 16 and 18 months of age. The mothers also completed the CDI:WG at both visits. The words that the children produced were compared with those the mothers reported on the vocabulary checklist, with family income and vocabulary size as grouping factors.
Results: Maternal reporting did not differ as a function of socioeconomic status but did increase from 16 to 18 months.
Conclusions: The vocabulary differences observed on the CDI:WG for children from low-income families do not appear to be a reflection of inaccurate maternal reporting. Further research is needed to determine whether these findings will generalize more broadly.

Key Words: vocabulary, socioeconomic status, toddlers

Several researchers have documented vocabulary differences among children from low- versus middle-income families (e.g., Hart & Risley, 1995; Horton-Ikard & Weismer, 2007; Whitehurst, 1997). Using monthly, hour-long home observations, Hart and Risley (1995) observed significant differences in the expressive vocabularies of children in professional, working-class, and welfare families at 36 months of age. In the study, children in working-class and welfare families produced 400 and 600 fewer words, respectively, than the children in professional families. Similarly, Whitehurst (1997) reported smaller vocabularies in low-income, suburban preschool children. Consistent with these findings, Dollaghan et al. (1999) identified a significant linear trend between maternal education and the number of different words that children produced in spontaneous language samples at age 3. Horton-Ikard and Weismer (2007) observed significant differences in the vocabularies of low- and middle-income African American toddlers on the Peabody Picture Vocabulary Test—III (Dunn & Dunn, 1997) and Expressive Vocabulary Test (Williams, 1997). Given that slow expressive vocabulary development may be clinically significant (e.g., Rescorla, 1989), it is important to ensure that we identify valid methods of measuring the expressive vocabularies of young children from low-income families.

Parent report instruments offer valuable information to both clinicians and researchers. These measures may be practical as screening tools or to supplement information gained from direct observation of a child. The MacArthur–Bates Communicative Development Inventories (CDI; Fenson et al., 2007) are a set of widely employed parent report instruments available in two long forms: Words and Gestures (CDI:WG), with norms for children from 8 to 18 months of age, and Words and Sentences (CDI:WS), with norms for children from 16 to 30 months of age. A significant advantage of employing parent report measures is that they present a cost-effective method of obtaining information (cf. Fenson et al., 1993). However, because the accuracy of the data is inherently a function of the reporter, it is essential to establish the data's validity prior to making clinical judgments based on parental report.

Previous investigations have examined the validity of parent report instruments as screening tools in low-income families. For example, Squires, Potter, Bricker, and Lamorey (1998) compared parent report for low- and middle-income families using the Ages and Stages Questionnaires (ASQ; Bricker, Squires, & Mounts, 1995) with their children's performance on the Bayley Scales of Infant Development (Bayley, 1969). Parents were mailed the ASQ when their children

were 4, 8, 12, 16, 20, 24, 30, and 36 months of age. Children's scores on the ASQ were used to classify the children as either "at risk" or "okay." At 12, 24, and 30 months, a similar dichotomous classification was obtained from administration of the Bayley by a trained child development specialist. They observed no significant differences between the low- and middle-income families with respect to the percentage of accurate classifications (i.e., the parental ASQ classification was consistent with the results of the Bayley). The ASQ examines children's development at a global level rather than focusing on a narrow aspect (e.g., language). Nevertheless, the results are encouraging, as parents from low-income families completed the task as accurately as parents from middle-income families.

Turning to instruments for assessing language development, Rescorla (1989) examined the validity of the Language Development Survey (LDS) as a screening tool in children attending an urban hospital clinic. In Rescorla's sample, 62% of the mothers were unemployed. In addition, 55% of the mothers had no education beyond high school, and 12% of the mothers had not completed high school. Rescorla compared the LDS results with the children's performance on subtests from two expressive language measures that were administered to the children. Significant correlations that exceeded $r = .70$ were found between the LDS results and each of the subtests. These results suggest that mothers from sociodemographically diverse backgrounds are able to report their children's vocabulary.

Dale, Bates, Reznick, and Morisset (1989) examined the validity of the Early Language Inventory (ELI), a parent report instrument that was a precursor to the CDI. The sample included both full- and preterm children as well as a "social risk" group representing children of mothers who "lacked social support and were young, or poor, or had a low level of education" (Dale et al., 1989, p. 242). Significant, moderately strong correlations ($r = .43–.63$) were observed between vocabulary as reported on the ELI at 20 months and an expressive language subscale derived from selected items on the Bayley Scales of Infant Development (Bayley, 1969) administered at 20 or 24 months of age. It is worth noting that the weakest correlation was observed for the social risk group. Furthermore, across all of the full-term children, a significant correlation ($r = .17$) was observed between children's total vocabulary and the family's Hollingshead Index, a measure of socioeconomic status.

An advantage of the CDI is that it can be employed in a longitudinal analysis without the repeated administration affecting the results. Reznick and Schwartz (2001) compared the reported vocabularies of children at 12 months of age whose parents were completing the CDI for the first time with children whose parents had completed the CDI on two previous occasions. Had a practice effect influenced parent report, these researchers would have found larger vocabularies for the children whose parents had had prior experience completing the CDI. However, the authors reported no differences in vocabulary size for these two groups of children.

Despite the fact that the CDI's validity has been established, both the user's manual and the literature (e.g., Arriaga, Fenson, Cronan, & Pethick, 1998; Fenson et al., 1993) have cautioned researchers and clinicians against employing the norms with children from low-income families because previous investigations that employed the CDI in a low-income population have yielded mixed results. Arriaga et al. (1998) analyzed the scores of low- and middle-income children on the CDI:WS. The resulting scores reflected the entire continuum (i.e., 5th to 99th percentile). Relative to the children from middle-income families, a disproportionate number of the children from low-income families had scores at or below the 50th percentile. One of the explanations that the authors explored was the possibility that low-income parents may have underreported their children's vocabulary. Similarly, Fenson et al. (2000) commented that "it is not yet clear whether the lower scores for low-income children reflect a slower pace of language development or underestimation or incomplete reporting by their parents" (p. 327).

Roberts, Burchinal, and Durham (1999) addressed the issue of incomplete reporting in their examination of maternal report of children's vocabularies in African American families who were predominantly low-income (i.e., 69% of the families met the federal poverty thresholds). At 18, 24, and 30 months, maternal report using a short form of the CDI was compared to expressive language scores obtained from administering the Sequenced Inventory of Communication Development, Revised Edition (SICD–R; Hedrick, Prather, & Tobin, 1984). The scores for the CDI were separated by quartiles and compared to chronological age scores on the SICD–R in an effort to identify mismatched scores (i.e., parental report that was much higher or lower than the child's performance on the SICD–R). The proportion of parents who appeared to be underreporting their children's performance increased from 25% at 18 months to 39% at 30 months. Conversely, the percentage of parents who appeared to be overreporting never exceeded 13% (at 24 months). Houston-Price, Mather, and Sakkalou (2007) also found that parents underreported their children's comprehension when they compared parental report on the Oxford CDI (Hamilton, Plunkett, & Schafer, 2000) to the children's performance using a preferential looking paradigm. At 15, 18, and 21 months of age, the children responded to words that the parents had indicated the children did not know.

Addressing the issue of underestimation or incomplete reporting necessitates comparison with an alternative estimate of a child's vocabulary. Vocabulary estimates in young children are available from a variety of sources (e.g., lexical diaries, standardized testing, and vocabulary measured in a conversational sample). It would not be possible to address the issue of accuracy of reporting by examining alternate parental estimates. Comparing a checklist against a diary could yield comparable lexical inventories, but they might both fail to represent the child's actual lexicon. A better method for examining accuracy of parental report would be through comparison against a set of words that the child is known to produce. In children younger than 24 months, obtaining a sample via standardized testing would not be a viable option. There are test items addressing productive vocabulary on several global measures (e.g., Wetherby & Prizant, 2002; Zimmerman, Steiner, & Pond, 2002); however, global assessment tools are inappropriate for obtaining lexical estimates, as the focus of these instruments is to differentiate between children with and without language impairments and not to obtain a

representative sample of the child's lexicon. Although tools have been developed to focus on vocabulary assessment, the normative data begin at 24 months (e.g., Brownell, 2000) or 30 months (e.g., Williams, 2007).

Parent report instruments have also been shown to be valid by comparing them to children's performance in a language sample (e.g., Pan, Rowe, Spier, & Tamis-LeMonda, 2004; Thal, Jackson-Maldonado, & Acosta, 2000). Pan et al. (2004) calculated the number of lexical types produced by children from low-income families in 10-min language samples. Similarly, Thal and colleagues (2000) calculated the number of different words produced in a 35-min language sample. Both groups of researchers reported moderately high correlations between children's vocabulary from language samples and data obtained from parent report instruments (r = .49 and r = .66, respectively). The use of correlational data lends itself well to examining the broad validity of parental report (e.g., Thal et al., 2000). Unfortunately, correlational analyses only address the extent to which the estimates are comparable, and as Houston-Price and colleagues (2007) argued, "correlational studies say nothing about the absolute accuracy of parents' estimates" (p. 704). Although the CDI forms were not developed to yield specific vocabularies for individual children, the question remains, "How do parents select the words that contribute to these estimates of lexicon size?" To examine accuracy, more detailed analyses are required.

Arguably, language sampling represents an ideal choice for obtaining an alternative lexical estimate to address accuracy of reporting. All of the words the child produces in the language sample that appear on the parent report instrument can be examined to determine whether the parent's report is consistent with the child's productions. In addition, the sample is more likely to be representative of the child's actual behaviors because the child is interacting with a familiar adult (e.g., parent or caregiver). Furthermore, a parent is likely to be able to gloss words that may be unfamiliar to an examiner. For the purposes of this investigation, accuracy does not represent the extent to which the child's lexicon is adequately represented by a measure. Rather, accuracy refers to how well mothers were able to report the words that their children produced in a language sample.

To that end, the purpose of this investigation was to compare the accuracy of vocabulary reporting by low- and middle-income mothers on the CDI:WG. In particular, parent report data of socioeconomically diverse children's vocabularies were compared with language sample production data at 16 and 18 months of age.

Three primary research questions were posed. First, are there differences in the accuracy with which mothers from low- and middle-income families report their children's vocabulary on the CDI:WG form? If the mothers in low-income families have significantly lower accuracy levels than the mothers in middle-income families, this would be evidence that they underestimated their children's repertoires. Second, does maternal accuracy differ as a function of the child's age (16 vs. 18 months)? Maternal accuracy is not expected to increase from 16 to 18 months, given Reznick and Schwartz's (2001) finding of no practice effect for the CDI. Rather, accuracy may decrease due to an expansion of the child's lexicon, which could make it more difficult to precisely report the words the child uses. Finally, is the percentage of words produced that are not included on the CDI:WG comparable for low- and middle-income children? If children from low-income families have lexicons that are qualitatively different, one would anticipate that a larger percentage of the words that they produce in the language sample would be unavailable on the CDI form. As a consequence, the CDI results would underestimate the vocabulary of children from low-income families.

Method

Participants

The data were drawn from a longitudinal data set in which 23 children from socioeconomically diverse families had been followed from 12 to 18 months of age (Furey, 2003). Families were recruited by distributing fliers through a variety of agencies that provide services to young children (e.g., day care centers, libraries, public health department, park districts, and early childhood programs). Each mother–child dyad traveled to a campus speech and hearing clinic for a total of four visits when the child was 12, 14, 16, and 18 months old. Only the 16- and 18-month visits (plus or minus 2 weeks) were employed in this analysis to ensure that the children's repertoires were large enough to facilitate meaningful comparisons.

Eleven of the children (five boys and six girls) were from low-income families, and 12 of the children (five boys and seven girls) were from middle-income families. All of the low-income mothers reported that their family income fell below 150% of the federal poverty thresholds that were in effect at the start of the investigation. The ethnicities of these children were reported to be African American (n = 4), European American (n = 4), and multiracial (n = 3). In contrast, the middle-income families all reported income levels that exceeded 200% of the federal poverty thresholds. The ethnicities of these children were reported as African American (n = 1), European American (n = 9), and multiracial (n = 2). At the 12-month visit, all of the children who participated in this investigation obtained a standard score of 85 or greater on the auditory comprehension subscale of the Preschool Language Scale, Third Edition (Zimmerman, Steiner, & Evatt-Pond, 1991).

Materials

Stoel-Gammon's (1987) protocol for the Language Production Scale was followed to obtain the language samples. The four toy sets were vehicles, a farm set, a tea party set, and nurturing toys (i.e., cradle, dolls, blankets, bottles, comb, hat, and phone). An audio recording (Tascam 112 MKII) of the mother–child interaction was made using two omnidirectional microphones.

The CDI:WG was employed at all four visits in the original investigation. Using the CDI:WG ensured that children's vocabularies did not increase simply as a function of the larger set of words on the CDI:WS.

Procedure

The examiner introduced one of the four toy sets every 8 min and left the room. The examiner asked the mothers to

play with their children as they would at home. To facilitate later transcription and analysis, the mothers were also asked to repeat anything that the children produced which they recognized as words.

At the conclusion of the language sample, the mothers were asked to complete the CDI:WG (Fenson et al., 1993). The mothers were not given any directions other than those that appeared on the form. New forms were provided at each visit. After the researcher observed inconsistent reporting for one of the children (i.e., the mother reported that the child produced words at 12 months but reported no words at 14 months), a procedural change was implemented. As a result, if mothers did not indicate that their children "understands and says" any words, the examiner inquired, "So your child is not saying any words at this point?" If the mother's response was "no," the form was accepted. If her response was "yes," she was permitted to make changes to the form. This question was not posed if the mother had identified one or more words that her child "understands and says."

Transcription Reliability

All of the samples were transcribed by the author using broad phonetic transcription. After a delay of several weeks, the author relistened to each session while examining the initial transcription. A total of 186 utterances, representing 1% of the four-visit data corpus, were discarded because the child was whispering or there was difficulty with the transcription. Across all four visits, a total of 15,489 vocalizations were retained. Interrater reliability was calculated using 50 vocalizations from each of 10 randomly selected sessions. Intrarater reliability was calculated for 50 vocalizations from each of seven randomly selected sessions. Interrater reliability was calculated using a point-by-point method for supraglottal consonants disregarding voicing differences; interrater reliability was found to be 77% and 83% for the 16- and 18-month samples, respectively. The average intrarater reliability, calculated in the same manner, was 92%.

Checking of Coding

Six criteria were employed to differentiate babble from lexical items. An item was considered to be lexical if it met any one of the following six criteria: (a) it was an exact phonetic match of the adult target, (b) it contained "an adjacent consonant vowel or vowel consonant pair from the adult word target" (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991, p. 239), (c) it was glossed by the mother as a word, (d) it was treated as meaningful by the mother, (e) it was judged to be an imitation or attempted imitation of the mother's previous utterance, or (f) its identification as a word was facilitated by the conversational context. Interrater agreement for the identification of lexical items averaged 89%. The transcriber who checked the coding was also asked to listen to tokens where the author was uncertain of the lexical status. All disagreements were resolved through discussion. Across all four visits, a total of 51 vocalizations were discarded, as it was not possible to resolve their lexical status.

Analysis

The researcher completed the accuracy calculations in the following manner. First, after each child's productive inventory from the language samples was compiled, each inventory was reduced to unique forms. Second, the words that the children produced which also appeared on the CDI:WG were used as the denominator in all calculations. Words the children produced that were glossed as inflected (e.g., "what's") were compared with uninflected forms on the CDI:WG. Similarly, words produced as diminutives (e.g., "horsie) were compared with the nondiminutive counterparts. In addition, reduplicated words (e.g., "bite bite") were compared with nonreduplicated forms (i.e., "bite"). Third, each word in the child's productive inventory was placed into one of the following three categories: (a) the mother indicated that her child "understands and says" the word, (b) the mother indicated that her child "understands" the word, or (c) the mother made no marks indicating that she believed her child understood or said the word. Two different calculations were obtained for each visit: production reporting and comprehension reporting. Production reporting values were tallied from the words that the mother reported the child understood and said. Comprehension reporting values were tallied from all of the words that the mother checked on the CDI:WG as either "understands" or "understands and says." Finally, accuracy values were obtained by dividing the number of words reported by the number of unique words in the child's productive inventory. For example, if a child produced 17 words and the mother reported that her child produced 11 of those words, the production reporting accuracy was 65%. If the mother reported that her child understood 12 of the 17 words, the comprehension reporting accuracy was 71%.

To accommodate the vocabulary differences among the children, the researcher used percentages of words accurately identified by the mothers rather than raw numbers. One limitation of using percentages is that they fail to account for size differences of children's vocabularies. For example, if a parent identified either two of two words or 33 of 33 words, the accuracy would be 100%, even though these are not equivalent tasks. A mother's failure to identify one word when the child had a three-word repertoire (i.e., 2/3 = 67%) would have a far more detrimental effect on accuracy than when a child had a 30-word repertoire (i.e., 29/30 = 97%).

According to Zar (1999), percentage data "form a binomial, rather than a normal, distribution" (p. 278). To adjust the distribution, Zar and others (e.g., Reinard, 2006) have recommended the application of an arcsine transformation. Therefore, in this study, the percentages were converted using an arcsine transformation, and the arcsine values were entered into SPSS Version 11.5 for the first two research questions. Given that arcsine-transformed values are not readily interpreted (e.g., the arcsine of 0.88 is 1.23), the values reported in the results section reflect the percentages prior to the transformation (e.g., 88%).

The words that the children produced which did not appear on the CDI:WG were tallied for both visits by group (e.g., low-income at 16 months) to address the final research question. These words were tallied as unique forms so that they would most accurately represent the words that were not

available on the instrument. A tally by compiling all of the words reported would have resulted in an overestimate of the words that did not appear on the CDI:WG. For example, a stuffed Ernie was included in two of the toy sets, and several children produced his name (e.g., three different children in middle-income families at 16 months). Each of the master lists of all words the children produced was sorted alphabetically, and the unique words were identified. The percentage of words not on the CDI:WG was obtained by dividing the number of unique words not on the CDI:WG by the number of unique words produced by the group at each visit.

Results

At 16 and 18 months, the number of words included in the accuracy analyses of words on the CDI:WG was 294 and 417, respectively. At 16 months, the children from low- and middle-income families produced 120 and 174 words, respectively. The children from low-income families also had smaller inventories at 18 months (135 vs. 282).

The number of unique words was tallied to be used in the calculation of the percentage of words that were not on the CDI:WG. The children in low-income families produced 83 and 115 unique words at the 16- and 18-month visits, respectively. The children in middle-income families produced 115 unique words at 16 months and 179 unique words at 18 months.

The number of words that were reported for each participant by visit is reported in Table 1. The denominator used to calculate each of the accuracy percentages can be identified by subtracting the number of words that were not available on the CDI from the number of words produced.

Research Questions One and Two

Production reporting. A mixed analysis of variance (ANOVA) was calculated to determine whether production reporting accuracy varied as a function of either income (between-subject factor) or visit (repeated measure). Neither the interaction between income and visit, $F(1, 21) = 0.03$, $p = .86$, nor the effect of income were significant, $F(1, 21) = 1.13$, $p = .30$. The increase in maternal accuracy from the 16-month visit ($M = 0.56$, $SD = 0.25$) to the 18-month visit ($M = 0.75$, $SD = 0.25$) was significant, $F(1, 21) = 7.00$, $p = .015$, partial $\eta^2 = 0.25$. The partial eta squared value indicates that 25% of the variance in maternal accuracy is explained by the visit. Based on Cohen's (1988) guidelines, this represents a large effect.

Comprehension reporting. A second mixed ANOVA was run to examine whether comprehension reporting was different as a function of income or visit. The interaction between income and visit, $F(1, 21) = 1.82$, $p = .19$, and the effect of income were not significant, $F(1, 21) = 2.20$, $p = .15$. The increase in accuracy from the 16-month visit ($M = 0.80$, $SD = 0.21$) to the 18-month visit ($M = 0.96$, $SD = 0.08$) was statistically significant, $F(1, 21) = 13.21$, $p = .002$, partial $\eta^2 = 0.39$. Thirty-nine percent of the variance in maternal accuracy was explained by the visit, which also represents a

TABLE 1. Words produced that were reported on the MacArthur Communicative Development Inventories Words and Gestures (CDI:WG) form for individual participants by visit.

	16 months				18 months			
Child	Produced	Not on CDI:WG	"Says"[a]	"Understands"[b]	Produced	Not on CDI:WG	"Says"[a]	"Understands"[b]
L1	3	2	0 (0%)	0 (0%)	3	2	1 (100%)	1 (100%)
L2	13	2	7 (64%)	8 (73%)	7	2	5 (100%)	5 (100%)
L3	14	4	7 (70%)	8 (80%)	31	10	11 (52%)	15 (71%)
L4	6	0	2 (33%)	3 (50%)	10	3	2 (29%)	7 (100%)
L5	7	1	4 (67%)	5 (83%)	4	2	0 (0%)	2 (100%)
L6	21	3	9 (50%)	15 (83%)	29	4	22 (88%)	24 (96%)
L7	11	2	7 (78%)	7 (78%)	4	1	3 (100%)	3 (100%)
L8	9	2	5 (71%)	6 (86%)	16	6	5 (50%)	8 (80%)
L9	24	7	11 (65%)	12 (71%)	32	10	22 (100%)	22 (100%)
L10	5	3	0 (0%)	2 (100%)	8	2	3 (50%)	6 (100%)
L11	40	7	26 (79%)	32 (97%)	48	15	27 (82%)	32 (97%)
M1	6	0	3 (50%)	5 (83%)	26	4	21 (95%)	22 (100%)
M2	17	6	8 (73%)	9 (82%)	54	19	28 (80%)	31 (89%)
M3	48	11	21 (57%)	29 (78%)	70	18	37 (71%)	46 (88%)
M4	13	3	9 (90%)	10 (100%)	36	7	25 (86%)	27 (93%)
M5	35	11	16 (67%)	22 (92%)	41	8	31 (94%)	33 (100%)
M6	9	1	1 (13%)	8 (100%)	20	1	14 (74%)	18 (95%)
M7	30	5	11 (44%)	18 (72%)	41	9	17 (53%)	28 (88%)
M8	13	1	8 (67%)	11 (92%)	17	5	11 (92%)	12 (100%)
M9	4	1	1 (33%)	2 (67%)	17	5	10 (83%)	12 (100%)
M10	13	4	8 (89%)	9 (100%)	16	5	9 (82%)	11 (100%)
M11	23	6	12 (71%)	16 (94%)	16	5	10 (91%)	11 (100%)
M12	15	3	8 (67%)	10 (83%)	20	6	11 (79%)	14 (100%)

Note. L = child from a low-income family; M = child from a middle-income family.

[a]Includes words that were checked as "understands and says."
[b]Includes words checked as either "understands" or "understands and says."

large effect. The accuracy percentages for production and comprehension reporting by visit are graphed in Figure 1.

Research Question Three

Two tests of independent proportions were run to examine the percentage of words that did not appear on the CDI:WG at 16 and 18 months for the low- and middle-income families. At 16 months, 29% of the unique words produced by the children in low-income families were not listed on the CDI:WG. For children in middle-income families, 28% of the unique words that they produced were similarly unavailable. An increase for children in both low- and middle-income families was noted at 18 months when these values rose to 35% and 36%, respectively. No significant differences were observed for low- versus middle-income families for the percentage of words that were not available on the CDI:WG at either 16 months ($z = 0.1679$, $p = .75$) or 18 months ($z = –0.2672$, $p = .71$). That is, it was not the case that the children in low-income families produced a disproportionate number of words that did not appear on the CDI:WG. A complete list of the unique words that were not listed on the CDI:WG is available in the Appendix.

Discussion

Three major findings were revealed. First, there were no differences in maternal reporting accuracy between low- and middle-income mothers at either 16 or 18 months; that is, mothers in low-income families reported their children's vocabularies on the CDI:WG with the same degree of accuracy as mothers in middle-income families. Second, for both low- and middle-income mothers, maternal accuracy was greater at 18 months than at 16 months. Third, the percentage of words that the children produced that were not listed on the CDI:WG was comparable for low- and middle-income children. These results suggest that using the CDI:WG is a viable way to estimate the vocabulary of 16- and 18-month old children from low-income families.

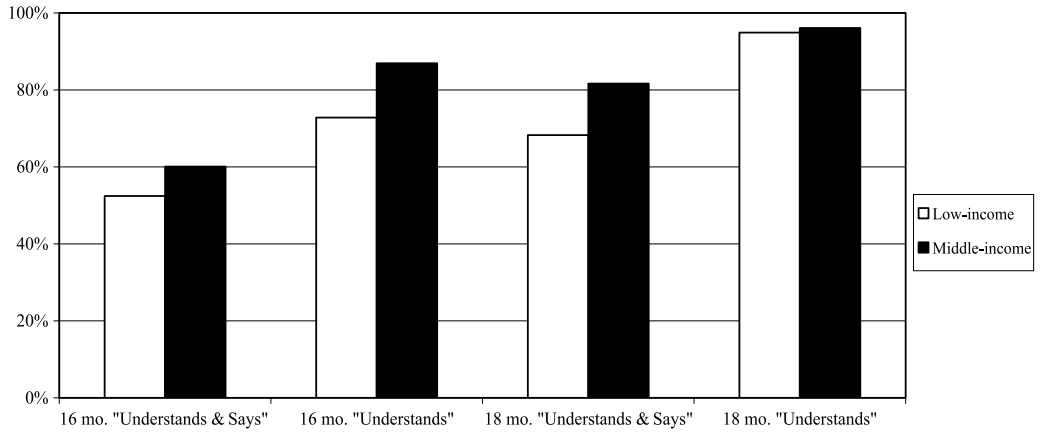
The absence of accuracy differences as a function of income suggests that when vocabulary differences are observed for children from low-income families on the CDI, they are not the result of "underestimation or incomplete reporting by their parents" as Fenson et al. (2000, p. 327) suggested. Overall, the children from low-income families were reported to have produced and understood collectively fewer words than their peers in middle-income families on the CDI, which is consistent with Feldman et al.'s (2000) 24-month data. Although the mothers in low-income families underestimated their children's vocabularies, the same pattern was observed for mothers in middle-income families. Therefore, it is likely that when smaller lexicons are seen for children from low-income families on the CDI, they reflect the discrepancies that researchers have observed when they have employed other methodologies (e.g., Hart & Risley, 1995).

Although the present data are not consistent with Feldman et al.'s (2000) 12-month data, this investigation yielded no evidence that the mothers in low-income families employed a "more liberal criterion" (p. 319) for identifying words. If the low-income mothers in the present study had selected more words than the middle-income mothers, there would have been more opportunities for the reported words to match those produced in the language samples. Were this the case, one might have expected greater accuracy among low-income mothers, a result that did not occur.

In the present investigation, imperfect reporting accuracy was observed at both visits. At 16 and 18 months, the mothers underreported their children's productive vocabularies by approximately 44% and 25%, respectively (see Figure 1). Previous research has documented a similar pattern of underreporting. For example, Roberts and colleagues (1999) found that the number of parents who underreported their children's vocabularies increased from 25% at 18 months to 39% at 30 months.

For mothers in both low- and middle-income families, accuracy was higher when the words the children produced were compared with maternal report of comprehension (i.e.,

FIGURE 1. Percentage of words produced that were reported on the MacArthur Communicative Development Inventories Words and Gestures form by visit.



"understands" and "understands and says"). This result is somewhat surprising, as the production data were obtained from language samples of mother-child dyads, and the mothers completed the forms right after having heard their children produce the words. One possible interpretation is that this increase is a reflection of the mother's partial knowledge. That is, a mother may be aware that she uses a form and her child is responsive to it; however, she may not have consciously recognized that her child has begun to produce the word. The higher accuracy levels in comprehension reporting may reflect the acquisition process. One would anticipate that much of the time children understand words before they produce them. Therefore, a mother would have had greater experience with the child responding to (i.e., comprehending) rather than producing the form. Alternatively, mothers may have used a more restrictive set of criteria to identify the child's words; that is, the mothers may have been reluctant to give their children credit for producing a particular word until they heard it produced on several different occasions. Although a maternal gloss was one of the criteria used to differentiate words from babble, there were five additional criteria. For example, the researcher classified as lexical all vocalizations that contained "an adjacent consonant vowel or vowel consonant pair from the adult word target" (Huttenlocher et al., 1991, p. 239). It is possible that these additional criteria were less stringent than what the mothers employed.

The increases in accuracy from 16 to 18 months were significant for both production and comprehension reporting. At 16 months, the mothers accurately reported approximately 56% of the words that their children produced in a language sample. By the 18-month visit, this value increased to roughly 75% of the words that their children had produced. If the words produced were compared to comprehension reporting, these values increased to 80% at 16 months and 96% at 18 months. There are several possible explanations for the observed increase in the percentage of words reported from 16 to 18 months. A practice effect is one possibility; however, other findings contradict this explanation (e.g., Reznick & Schwartz, 2001). Alternatively, it is possible that as mothers hear children's words more frequently, they can identify them more accurately. Finally, it is important to recall that although the children were producing a greater variety of words at the later visit, mothers used the same form. Therefore, by virtue of selecting a larger set of words on the same form, one would anticipate a corresponding increase in accuracy.

No significant differences were observed between low- and middle-income families in the percentage of words the children produced that were not available on the CDI:WG at either visit. As might be anticipated given an expanding lexicon and a constant form, the percentage of words that did not appear on the CDI:WG increased from the 16- to 18-month visits for both groups. The fact that there was no significant difference between the groups suggests that the pattern of smaller vocabularies seen for children in low-income families is not a function of the CDI form.

Although the present results are encouraging, they should be viewed as preliminary. As there were vocabulary size differences in the language samples between the low- and middle-income families, the mothers' task may not have been equivalent. The extent to which missing or correctly reporting a word resulted in large changes in the accuracy percentage was a function of the number of words the child produced. In the future, increasing the number of participants may enable comparisons to be made between pairs of mothers from low- and middle-income families matched by the size of their children's vocabulary.

There are several methodological issues that affect the interpretation of the results. First, the CDI forms were completed after the language sample was obtained. This was done to ensure that the language samples were gathered while the child was most alert. One alternative would have been to mail the forms to the parents and request that they return them when the child came for the language sample. However, this would have introduced variability in the interval between completing the form and recording the language sample, and it is likely that some parents would have forgotten to return the completed form. Second, replication with a larger number of participants would provide stronger evidence for the pattern that was observed in this relatively small sample of convenience. Finally, it would be important to examine maternal accuracy in low-income families on the CDI:WS as well.

To date, all published investigations examining the validity of maternal report in low-income families have been addressed through the application of correlational data; that is, parental report was deemed valid when vocabulary size based upon parental report and performance on some behavioral measure were positively correlated. The current investigation serves to supplement the extant literature by examining accuracy as a function of the words that the parents select on the CDI:WG. When their children were 16 and 18 months of age, mothers in low- and middle-income families demonstrated comparable accuracy in reporting their children's vocabularies using the CDI:WG. At these ages, the percentage of words that the children produced that were not listed on the CDI:WG was comparable for the low- and middle-income families. The results of this investigation suggest that the CDI:WG is an appropriate measure for clinicians and researchers who require estimates of the emerging lexicons of children from low-income families.

Acknowledgments

A previous version of this article was presented at the Annual Convention of the American Speech-Language-Hearing Association in San Diego, CA, in November 2005. I am grateful to the families who participated in the longitudinal investigation that was completed at the University of Illinois at Urbana-Champaign. Support for the data collection was provided by a Student Research Grant in Early Childhood Language from the ASHFoundation, the Bamford-Lahey Children's Foundation, and a Robert L. Sprague Thesis Award. Partial support for this project was provided by an ASHA Advancing Academic–Research Careers Award. I am grateful to Chris Dollaghan and Margaret Wick for their valuable comments on earlier drafts of this article. In addition, I want to thank Gary Gillund for his advice regarding the statistical analyses. I want to acknowledge Brookelyn Anhalt, Catherine Capo, Kristen Conklin, Naomi Dunn, Jamie Fox, Laura Geller, Brianne Kosch, Angela

Locantore, Stephanie Mancini, Kaitlan O'Malley, Abby Prospero, Carrie Smith, and Natalie Unger for their assistance transcribing and checking the data.

References

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low- and middle-income families. *Applied Psycholinguistics, 19,* 209–223.

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York, NY: The Psychological Corporation.

Bricker, D., Squires, J., & Mounts, L. (1995). *Ages and Stages Questionnaires: A parent-completed, child monitoring system*. Baltimore, MD: Brookes.

Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test—Third Edition*. Novato, CA: Academic Therapy.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dale, P. S., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at 20 months. *Journal of Child Language, 16,* 239–249.

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research, 42,* 1432–1443.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test—III*. Circle Pines, MN: AGS.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development, 71,* 310–322.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development, 71,* 323–328.

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

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

Furey, J. E. (2003). *The growth of early phonological and lexical development: A longitudinal investigation* (Unpublished doctoral dissertation). University of Illinois, Urbana.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory. *Journal of Child Language, 27,* 689–705.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Hedrick, D. L., Prather, E. M., & Tobin, A. R. (1984). *Sequenced Inventory of Communication Development, Revised Edition*. Seattle: University of Washington Press.

Horton-Ikard, R., & Weismer, S. E. (2007). A preliminary examination of vocabulary and word learning in African American toddlers from middle and low socioeconomic status homes. *American Journal of Speech-Language Pathology, 16,* 381–392.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behavior in a preferential looking task. *Journal of Child Language, 34,* 701–724.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236–248.

Pan, B. A., Rowe, M. L., Spier, E., & Tamis-LeMonda, C. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language, 31,* 587–608.

Reinard, J. C. (2006). *Communication research statistics*. Thousand Oaks, CA: Sage.

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54,* 587–599.

Reznick, J. S., & Schwartz, B. B. (2001). When is an assessment an intervention? Parent perception of infant intentionality and language. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 11–17.

Roberts, J. E., Burchinal, M., & Durham, M. (1999). Parents' report of vocabulary and grammatical development of African American preschoolers: Child and environmental associations. *Child Development, 70,* 92–106.

Squires, J. K., Potter, L., Bricker, D. D., & Lamorey, S. (1998). Parent-completed developmental questionnaires: Effectiveness with low and middle income parents. *Early Childhood Research Quarterly, 13,* 345–354.

Stoel-Gammon, C. (1987). Language Production Scale. In L. B. Olswang, C. Stoel-Gammon, T. E. Coggins, & R. L. Carpenter (Eds.), *Assessing prelinguistic and early linguistic behaviors in developmentally young children* (pp. 118–150). Seattle: University of Washington Press.

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

Wetherby, A. M., & Prizant, B. M. (2002). *Communication and Symbolic Behavior Scales Developmental Profile*. Baltimore, MD: Brookes.

Whitehurst, G. J. (1997). Language processes in context: Language learning in children reared in poverty. In L. B. Adamson & M. A. Romski (Eds.), *Communication and language acquisition: Discoveries from atypical development* (pp. 233–265). Baltimore, MD: Brookes.

Williams, K. T. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: AGS.

Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition*. Circle Pines, MN: AGS.

Zar, J. H. (1999). *Biostatistical analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Zimmerman, I., Steiner, V., & Evatt-Pond, R. (1991). *Preschool Language Scale, Third Edition*. San Antonio, TX: The Psychological Corporation.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale, Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Received August 13, 2009
Revision received March 8, 2010
Accepted September 30, 2010
DOI: 10.1044/1058-0360(2010/09-0073)

Contact author: Joan E. Furey, Department of Communication, The College of Wooster, Wishart Hall, Wooster, OH 44691. E-mail: jfurey@wooster.edu.

Appendix

Words That Do Not Appear on the CDI:WG

16 months, low-income

because	lid	
bok bok	like*	
can't*	*mmm*	
did*	*oh*	
eieio	*ooh*	
Elmo	ring	
hay	roof*	
here's*	taste*	
hey	two	
huh	*uhhuh*	
is*	*uhuh*	
Laa Laa	world	

16 months, middle-income

bag	honk honk	snack*
barn	*huh*	three
beep	*mmm*	*uhhuh*
boat*	neigh	*uhuh*
bok bok	*oh*	want*
bonk	oink oink	whee
booger	okay	whoa
boom	*ooh*	wow
broccoli	pasta*	
cozy	pin	
Elmo	scarecrow	
Ernie	sit*	

18 months, low-income

ambulance	cowboy*	sorry
and	eieio	splash*
ashes	*Elmo*	*stuck**
at*	*Ernie*	*tea*
barn	hay	*too**
*bath**	*here**	uhuh
because	huh	*whee*
belly*	la la la	*whoa*
*boat**	laying	*wow*
bok bok	*mmm*	zoom
bonk	*oh*	
boom	Old McDonald	
burp	ooh	
can*	pick*	
cook*	roar	

18 months, middle-income

again	howdy	shut
all done	huh	side
all gone*	king	sit*
alright	lady*	snack
barn	microphone	squeal
Barney	*mmm*	standing*
*bath**	neigh	sticker
beep beep	*oh*	stinky
*boat**	oink oink	*stuck**
boo	okay	sucker*
boom	oops	sugar
call	pasta*	*tea*
chalk	pee pee	the*
clothes	phone*	*too**
Elmo	pink	top
Ernie	pot	two
full*	PU	uhhuh
guy	ring ring	wagon
hay	rooster*	*whee*
helicopter	sandals	*whoa*
*here**	scarecrow	*wow*
hmm	seat	

Note. Words in italics are present on lists of both low- and middle-income families. Words with an asterisk are on the CDI:WS.