VOCABULARY LEARNING IN CHILDREN FROM 8 TO 30 MONTHS OF AGE: A COMPARISON OF PARENTAL REPORTS AND LABORATORY MEASURES

Jennifer Jahn-Samilo
University of California, San Diego

Judith Goodman
University of Missouri

Elizabeth Bates and Monica Sweet
University of California, San Diego

Technical Report CND-0006
2000

Project in Cognitive and Neural Development
Center for Research in Language
University of California, San Diego
La Jolla, CA 92093-0526

SUBMITTED FOR PUBLICATION -- DO NOT CITE WITHOUT PERMISSION

VOCABULARY LEARNING IN CHILDREN FROM 8 TO 30 MONTHS OF AGE: A COMPARISON OF PARENTAL REPORT AND LABORATORY MEASURES

Abstract

The present study examined the validity of parental report of language production by concurrently examining vocabulary development using a parental report instrument (i.e., the MacArthur Communicative Development Inventories (CDIs)) and a laboratory measure (i.e., an elicited production task) for children 12 to 30 months of age, a period of dramatic language growth. Analyses revealed that a parental report and a laboratory measure of vocabulary yield similar patterns of growth although parental report may provide the earliest indicators of vocabulary development. In addition, both measures provided evidence for immense individual variation in vocabulary size and growth during this period of development.

INTRODUCTION

Parents are a rich source of information about their children's development. In fact, almost all parents have in some way compiled a report of their children's behavior, whether it be filling in dates and events in a baby book or a meticulous journal constructed by some parent-scientist with hours of training (e.g., Darwin, 1877; Leopold, 1949; Piaget, 1952, 1954). As these latter cases make clear, parental report has a long-standing history as a methodology in psychological studies. When parents are sensitive and accurate observers, this type of data collection should be a boon for the scientific study of language development because of the inherent advantages of parental report measures. By using a parental report measure, researchers can collect large amounts of data with relatively little expense, time and trained personnel. In addition, parental report has the potential to be very representative of a child's true behavior. Parents spend a lot of time with their children in multiple contexts, and thus have the opportunity to observe infrequent events. For instance, parents' records of their children's vocabulary can sample the entire vocabulary range, not just the frequent items. Further, since any child's behavior can run the gamut from "angel" to "devil", parents also have the opportunity to wait for instances of compliance and evidence of particular behaviors.

Despite these advantages, researchers are reluctant to use parental report measures, because they question whether this technique provides a valid measure of children's behavior (e.g., Feldman et al., 1999; Stiles, 1994; Tomasello & Mervis, 1994). This is important because clinicians and researchers must be certain they are measuring the skills or attributes of interest, rather than something particular to a specific parent. Reluctance to use parental reports and concerns about their validity stem from a number of factors. A natural pride may lead parents to embellish their children's behavior, thus inflating their children's performance above and beyond their actual behavior. On the other hand, modesty may lead parents to downplay their children's accomplishments, thus understating the actual behavior in question. Whether a child's behavior is inflated or deflated, such misreporting would reduce the validity of the measure.

Another factor important to the validity of parental report is training. Few of the parents who participate in scientific research with their children are parent-scientists who have specialized training. Without proper training, one could question whether parents can accurately detect and/or remember the behaviors in question (Stiles, 1994). For example, parents might not differentiate whether their child really understands a word or simply recognizes a certain context. Suppose a child retrieves a ball as it rolls across the floor in front of her while her parent is saying "See the ball. Get the ball." Parents may credit the child with understanding the word 'ball', but the child may simply want to play with the object. Indeed, a parent might infer that a child comprehends a word simply because the corresponding object is familiar to the child or solely because he or she recognizes the sound pattern, rather than because the child has made the connection between the word and its referent.

In addition to these concerns about validity, some researchers are reluctant to use parental report measures due to concerns about reliability. The reliability issue is of course important for any measure: as Spearman pointed out long ago, no measure can correlate higher with another measure than it correlates with itself. Therefore, a parental report measure as well as any criterion measures (e.g. language samples, language tasks) must be reliable. Specifically, whatever the parental report is designed to measure, it must measure it the same way each time it is used. Likewise, a criterion measure must be reliable. For example, suppose we are able to show that the reliability of our parental report measures exceeds +.80 (by test-retest, or split-half correlations). Suppose, however, that the reliability of a criterion measure used to validate parental report stands at +.50. Under these conditions, the highest correlation possible between parental report and the criterion measure is +.50, even if the validity of our parental report measure is perfect (that is, parental report and the criterion measure used for validation are measuring exactly the same skill).

Many recent studies of early language development have relied on a relatively new parental report instrument, the MacArthur Communicative Development Inventories (CDIs). The CDIs were developed across a 20-year period, and versions are now available in more than a dozen languages. They have become increasingly popular as research and diagnostic tools in the U.S. and elsewhere. The Words and Gestures form of the CDI, a 396-item checklist organized into 19 semantic categories, is designed to be used with children from 8 to 16 months of age. This first form requests information on both the child's comprehension and production vocabulary as well as the child's early use of communicative and symbolic gestures. The second form, Words and Sentences, is designed to be used with children from 16 to 30 months of age. This form requests information on the child's vocabulary production and morphological and syntactic development and contains 680 items organized into 22 semantic categories. A large normative, cross-sectional study including over 1800 children was carried out using the CDIs, providing data on language development between 8 and 30 months of age (Fenson et al., 1993). These inventories are a valuable resource for clinical purposes (e.g., screening for language delay) and basic research questions (e.g., investigation of the relationship between cognition and language). Given the impact that the CDI could have on both theory and treatment, it is critical to confirm its validity.

Using laboratory tests as the "gold standard", the results of a number of studies suggest that the CDIs are valid measures of several aspects of language acquisition (e.g., Dale, Bates, Reznick & Morriset, 1989; Ring & Fenson, in press). For instance, vocabulary and syntactic development in two-year-old children measured by a recent version of the CDI, were substantially and significantly correlated with laboratory tests (Dale, 1991). Miller, Sedey and Miolo (1995) compared scores from the vocabulary checklist portion of the CDI with both a spontaneous language sample (number of different words) and items from the Bayley Scales of Infant Development Mental Scale (BSID). They found significant correlations between the CDI and both the language sample ($r=0.75$, $p<.01$) and the BSID ($r=0.70$, $p<.01$).

The CDIs also have been abbreviated into various short forms. The scores obtained are significantly related to language samples (observed lexical types) at 18 and 24-months of age (Corkum & Dunham, 1996) and to Bayley MDI scores in the same period (Saudino et al., 1998). Short-form scores are also related to parent diaries, a more traditional form of report (Reznick & Goldfield, 1994).

The CDIs have been used successfully with populations other than normally developing English-speaking children. For instance, researchers using Spanish adaptations document a high correlation ($r=0.84$, $p<.0001$) between the number of different words in a spontaneous language sample and the number of words parents report that their children produce (Jackson-Maldonado, Thal, Marchman, Bates, & Gutierrez-Clellen, 1993). The CDI is also a valid measure of the language development of children whose language may be somewhat delayed. O'Hanlon and Thal (1991) found that scores from language-impaired children between 39 and 49 months of age on the Words and Gestures form of the CDI correlated significantly with scores on the Expressive One Word Picture Vocabulary Test and with observed vocabulary. In another study, researchers found that the CDI is also a valid measure of the language in a Down syndrome sample (Miller et al., 1995). Spontaneous language (0.82, $p<.01$) and items from the BSID (0.77, $p<.01$) were both correlated with vocabulary checklist scores on the CDI.

Finally, the CDI appears to be a valid measure of components of language other than simple vocabulary totals. Smith (1999) had parents complete the CDI in order to assess the number of count nouns in toddler’s productive vocabulary. The toddlers were then grouped according to their vocabularies and tested in the laboratory on either a novel word generalization task or a nonnaming task. When children were divided according to their performance on the CDI, these divisions also reflected developmental differences in the children’s tendency to generalize novel items by shape. Smith (1999) also found that when children were grouped according to the number of adjectives in their CDI productive vocabulary, these groups also reflected developmental differences in performance on a novel count noun generalization task and a novel adjective generalization task. Another study demonstrated that parents are able to discriminate between verbal (CDI) and non-verbal cognitive abilities, further attesting to their ability to provide valid estimates of their children's behavior (Saudino et al., 1998).

In addition to its impressive validity, the CDI appears to be reliable. Fenson, Dale, Reznick, Bates, Thal, & Pethick (1994) found the forms to be reliable when parents completed a second questionnaire about 1.5 months after completing their first. Jackson-Maldonado et al., (1993) also found impressive test-retest reliability in their study which used Spanish adaptations of the CDI. On the Spanish adaptation of the Words and Gestures form of the CDI, the correlations were .97 for comprehension and .81 for production. On the Words and Sentences adaptation, the correlation was .70 for production.

While the above studies are encouraging as a whole, they do not provide information about the validity of parental report over time. Many of the above studies sampled only part of the testing range covered by the CDI. Studies that collected data at two time points provide only an indication of validity over time. Concurrent validity of the CDI across the entire range of ages tested is necessary given the diagnostic and research applications for which it is currently being used. A longitudinal design that collects monthly data would enable one to more closely examine relationships between performance measured on laboratory tasks and parental report measures.

Prior work indicates that parental report measures are reliable, but corresponding measures are lacking for laboratory tasks. This is important since the validity of any measure may be limited by the reliability of that measure (cf. Fenson et al., 1994). If laboratory tests that claim to accurately assess children's language comprehension and production abilities across a range of ages bring into question the validity of the parental report measures, then one needs to establish whether those tests are stable over time. Hence we will pay special attention to the internal reliability of both the CDI and the laboratory measures used for validation.

To summarize, the present study represents an effort to assess the validity and reliability of a popular parental report instrument by concurrently examining vocabulary development using parental report and laboratory measures in a longitudinal design. Vocabulary production was assessed using the CDI (from 8 to 30 months) and both spontaneous and elicited production in the laboratory (from 12 to 30 months). Hence we can assess the utility of these instruments across a period of dramatic growth in many aspects of language and communication.

METHOD

Participants

Twenty-eight children (17 males and 11 females) participated in the present study. Children participated monthly beginning at eight (n=20) or nine (n=8) months of age and continued until they were 30 months old. Eighteen (64%) of the children were first born or had no siblings in the same household. Nine (32%) of the remaining children were second born and one child was fourth born. Two (7%) of the children were African-American, 6 (21%) were of various racial and ethnic combinations, and 20 (71%) were Caucasian. The socioeconomic status and educational levels of the parents of the children in this sample varied widely, and three of the children came from single-parent households most of the time they were participating in the study.

Participants were recruited through a subject pool and through personal referrals. The subject pool contained names, phone numbers, and birth dates of children whose parents responded to advertisements in parent magazines available in the county for a wide variety of studies in cognitive, social, and language development. Parents were contacted for this study when their children were between six and eight months of age, that is, prior to a time when any language delays might be noticed. Children who were regularly exposed to a second language were not enrolled in the study although eventually three children were exposed to a second language on a regular basis due to childcare arrangements.

Materials

Parental Report Measures. Parental report of the participants' language ability was gathered using the MacArthur Communicative Development Inventories (CDIs). As noted earlier, there are two forms of this measure, Words and Gestures and Words and Sentences. (For a detailed description of the inventories see Fenson, et. al., 1993.) The present study includes data from the vocabulary checklist portion of the inventories.

Laboratory Measure. Vocabulary production was assessed using an elicited production task called the grab bag. Three large cloth bags contained different objects or toys. The items chosen were based on data from the cross-sectional norming study (Fenson, et al., 1993). The norming study provided item-by-item analyses for the inventories indicating the percentage of children who produced and comprehended each individual item. In addition, we considered whether a word's referent could be clearly represented with an object or a toy that could fit into the bags. The three bags represented three levels of difficulty. The first bag contained a total of 15 items; 10 items which an average of 55% of 12-month-old children typically comprehend and five items which an average of 50% of 18-month-old children produce. Comprehension norms were used for the first bag because the limited productive vocabulary of the youngest children in the study severely limited our choices for inclusion. The second bag contained 10 items which an average of 75% of 28-month-old children typically produce. The third bag contained 11 items which an average of 46.5% of 30-month-old children typically produce.

Design and Procedure

Parental Report Procedure. The present study uses data from questionnaires collected between 12 and 30 months of age. From 8 to 16 months, parents completed the Words and Gestures form of the CDI. After 16 months, the child was switched to the Words and Sentences form, unless s/he did not yet comprehend 50 words. Eight of the children were delayed in switching to the Words and Sentences form. Table 1 lists all participants by ID number and sex. The table is separated into months, indicating the form (i.e., Words and Gestures or Words and Sentences) of the CDI each participant received during the course of the study. In addition, Table 1 indicates whether there was also a lab visit for a particular month.

The parents were mailed questionnaires monthly beginning when their children were either eight or nine months of age. Questionnaires were mailed one week prior to the child's monthly birthday. Parents received and returned the inventory by mail until their children were 12 months of age. Once their children were 12 months old, they began their monthly visits to the Child Development Laboratory and brought the inventories to their appointment. The items checked off by the parents each month were transferred to the next month's form prior to mailing. This transfer of items meant that the data were cumulative although parents were instructed that they could change their minds and remove items if their child did not appear to understand or produce those items during the current month. Parents did exercise this option on several occasions.

Laboratory Procedure. Parents brought their children to the Child Development Laboratory monthly to participate in a series of language tasks. The present study is part of a larger study, and parents and children participated in a number of additional tasks not considered here. Each session lasted between 45 minutes and 1-1/4 hours. Sessions began with a free-play period. The production task occurred approximately mid-way through the session. All interactions were recorded by wall-mounted video cameras and by an audio tape recorder.

For the laboratory production measure, the children were seated on the floor in front of the experimenter. The experimenter pulled each item out of the bag one at a time. If the child did not spontaneously name the item within approximately 10 seconds, s/he was asked to name the item and was rewarded (e.g., clapping) when s/he attempted to name the item. Once the child correctly named 80% of the items in a particular bag, he or she was given a more difficult bag on his or her subsequent visit. Correct naming was defined as an utterance consisting of at least one recognizable part of the target word accompanied by the child's attention. For example, "ba" was an acceptable approximation for the target word 'banana' as long as the child was attending to the banana. For the first bag, once the child correctly named 8 of the initial 10 items, the additional 5 were added and the child was tested on all 15 the subsequent visit. The average age of the children when the additional 5 items were added to the first bag was 21.92 months. The average age of the children when they were switched to the second bag was 24.62 months and switched from the second to the third bag was 26.6 months. Twenty-six children reached the level of adding the additional five items to the first bag and of switching to the second bag. Of these children, 15 reached the level of switching to the third bag.

RESULTS

Our primary goal was to assess the concurrent validity of the CDI parental report measures longitudinally over a period of dramatic language development. In order to obtain a more accurate assessment of the validity of the CDI, two related sets of scores were used. One set was based on all items on the CDI. Production of these items as reported by parents was compared to the items used on the laboratory measure. The second directly compared those items included on the laboratory measure and those same items on the CDI. Thus the second set utilized only a subset of the entire CDI. The total scores for all measures were converted into percentages so that all scores would be on the same scale.

We conducted six types of analyses. First, we will present descriptive statistics and characterize the overall patterns of vocabulary growth. In this section, we will also compare the shape and variance of expressive vocabulary growth for the respective parent and laboratory measures, including the monthly means, medians and percentiles (based on the Fenson et al. cross-sectional normals) of vocabulary production for each measure across the 8--30-month age range (12–30-month only for the laboratory measure). By comparing the curves that result from the CDI data with those observed in the laboratory data, we can take a qualitative look at the similarity between measures.

Second, we will examine the internal reliability and stability of both the CDI and the laboratory measure using test-retest and split-half analyses, respectively. The emphasis here is on developmental changes in the reliability of each instrument across this period of development.

Third, the scores from the laboratory and both sets of scores from the CDI will be entered into Pearson product-moment correlations with two-tailed tests of significance, at each month of the study in which both measures were available (i.e., 12-30 months). This will yield information about developmental changes in the validity of parental report across the period of study.

Fourth, we will calculate growth functions across this age range for both the CDI data and the laboratory data, using hierarchical linear modeling. The model will provide us with four parameter estimates for each outcome measure: an intercept, a slope (linear component), a quadratic component and a mean squared error. We will then compare the components contributing to the laboratory data growth curve to the components contributing to the CDI data growth curve. In this fashion, we can quantify the informal (visual) growth curve comparisons conducted in the first section, and determine whether individual differences in the shape of change are comparable across parental report and laboratory measures.

Fifth, using these fitted growth curves we will then calculate two separate LD50s[1] which we will define as the points (or ages) at which each child reached 50% of the total on each of the measures. The two LD50 statistics will be correlated, and a CDI-Lab lag score will be computed by calculating the difference between the LD50 from the CDI curve and the LD50 from the laboratory task. In this fashion, we can quantify differences between parental report and laboratory observations in the timing of vocabulary development.

Finally, we were interested in determining the extent to which similarities between parental report and the laboratory measure are dependent upon the actual content of the 36 items selected for the laboratory task. Towards this end, we selected a second set of 36 items from the CDIs that were matched to the original set for relative difficulty (i.e. age of acquisition according to the Fenson et al. cross-sectional norms). For this “virtual set”, we will calculate mean, median and percentile scores as well as reliability coefficients. If the original set and the virtual set show similar patterns of growth, reliability and validity, then we may conclude that they are equivalent forms, and that the correlations between laboratory and parental report reflect lexical ability independent of item-specific content. Alternatively, if we find differences between the original 36 CDI items (which overlap in content with the laboratory test) and the second set of 36 CDI items (which do not overlap in content with the laboratory test), then we may conclude that our validation depends not only on general lexical abilities but on item-specific content.

Vocabulary Growth Curves

Parental Report. We first examined mean scores and standard deviations based on all items from the checklist portions of the two CDI forms. All parental report scores were converted to percentiles (corresponding raw scores are presented in parentheses). In computing the percentiles, we divided all scores, including those from 8 to 16 months of age, by 680, the total possible on the Words and Sentences form of the CDI. Figure 1 presents the gradual development of productive language in our sample. By 15 months of age, the average child produces just over 5% (or 34.44 words out of 680) of all items on the CDI. The mean score rises to 81.6% (554.9 words) by 30 months. As vocabulary develops, we see that the standard deviations are larger than the mean until 17 months of age, reflecting wide variability in language growth (Table 2). As we will see below, this wide variability is observed in both the CDI and the laboratory measure, and thus is not an artifact of parental report per se (cf. Feldman et al., 1999).

The mean curve for the set of items on the CDI that overlapped with our laboratory measure (Figure 1) depicts the same gradual onset of vocabulary. Average scores for the overlapping set of items are slightly higher than those based on the entire CDI. This is not surprising, because the items were chosen specifically to contain words likely to appear early in development. Even with this subset of early words, there is much variability (Table 2).

The reported median scores for all items on the CDIs indicate again that productive vocabulary development is slow and gradual at first (Figure 2). In fact, it is not until 13 months of age that children at the median produce 1% (6.8 words) of the total possible words. By 17 months of age they produce 7.7% (52.4 words) of the total possible and at the completion of testing (at 30 months), they produce 90% (612 words) of the words. This demonstrates that the CDI captures the onset and asymptote of vocabulary development for these items. At the extremes, the lowest 10th percentile first show stirrings of productive vocabulary at 15 months of age, gradually developing to produce just over 1.5% (10.7 words) of the total 4 months later and about 41% (279.5 words) of the total at 30 months of age. Parents of children at the 90th percentile report that their children produce words very early in our study. The vocabulary development of this upper 10% takes off rapidly, surpassing 10% (75.2 words) of the total possible by 14 months of age and 50% (361 words) of the total by 20 months of age. This group has nearly reached the ceiling of the CDI. They produce 99.3% of the total vocabulary on the CDI (675.5 of 680 words) by 25 months of age.

When we turn to the subset of items on the CDI that overlapped with the grab bag, we see a

[1]The term LD50 is not an abbreviation for any variable in this study, but a statistical term used in this type of analysis. In this case, it refers to the age at which we would expect a child to reach before half of the children exposed to input would acquire that particular word.

very similar picture (Figure 3). Median scores indicate that this subset of also appears gradually, with rapid acceleration after 16 months of age. Parents report that the first signs of production appear at 14 months of age (1 word), with slow and steady increases for the next few months. However, by the end of the study (30 months), children at the median were producing 90.3% (32.5 words) of the CDI items shared with the laboratory task. The variance captured by this subset of 36 CDI items was also very large throughout the study, similar to the total CDI scores. As Figure 3 indicates, children scoring at the 10th percentile lie close to the floor until 20 months of age when they produce 7.8% (2.8 words) of items. Those at the 90th percentile reach this level of production 9 months earlier. The difference between these two percentiles is even more striking at 27 months of age when the 90th percentile have reached the ceiling of the measure, but the 10th percentile produce only 13.4% (4.8 words) of the items. Five children produced all of the overlapping CDI items by the end of the study.

Laboratory Measure. Figure 1 presents the means for our laboratory production measure, the grab bag task, between 12 and 30 months of age. Recall that the obtained scores were converted into percentages by dividing by the total number of words possible on this measure (36); the corresponding raw scores are presented in parentheses. As is typically seen with most measures of language development, productive vocabulary is just getting off the ground by 12 months of age. By 16 months of age, the average child produces 5.4% of the total (or 1.9 words out of 36). By 30 months of age the average child in our study is producing more than 70% (25.56 words) of the possible words. As with the set of CDI items that overlapped with the laboratory measure, the standard deviations are no longer larger than the means once the children reach 17 months of age (Table 2).

As we also saw with the parental report data, the curves presented in Figure 4 indicate that a tremendous amount of individual variance is observable in laboratory performance. The median scores closely resemble the mean scores, passing 5% (2.49 words) of the total at 17 months of age and passing 70% (26 words) of the total at 30 months of age. At the extremes, the lowest 10th percentile children do not show signs of expressive vocabulary in the laboratory until 21 months of age whereas the upper 90th percentile children “hit the ground running” at the outset of laboratory testing. At 30 months of age, children at the 10th percentile produce over half (19 words) of the laboratory items, and children at the 90th percentile produce almost 90% (31.39 words) of the total. Hence the laboratory measure mirrors the CDI measures (both total score and the 36-item subset) in the magnitude and nature of individual differences in vocabulary development.

With regard to the shape of change in early vocabulary development, Figure 1 shows that the growth curve constructed from the laboratory production scores is extremely similar to that of overall CDI production. Both measures reflect slow growth at the outset, with a rapid acceleration after 16 months of age. Once production gets off the ground for the group as a whole (around 18 months) the curve constructed from CDI production scores is consistently ahead of the laboratory curve throughout the study. Nonetheless, qualitative similarities in the shape of the curves are retained when the laboratory test is compared with the same 36 items on the CDI, although in this case the offset in timing between parental report and laboratory performance is even greater than it was with the CDI totals. Thus, on average, parents report that their children produce more words than those children actually produce during elicited production tasks, but the rate or shape of vocabulary growth is qualitatively similar for the two measures (a quantitative assessment of this qualitative similarity is provided later by comparing components of growth).

Internal Reliability

The internal reliabilities of our measures were assessed using split-half analyses for each month. Because the full version of the CDI contains a very large number of words and its internal consistency has been established elsewhere (Fenson et al., 1994), we decided to compute reliability coefficients only for the subset of 36 CDI items, for comparison with the same 36 items in the laboratory production measure. The two halves used in the split-half analyses were matched according to age of acquisition (or difficulty) based on the CDI norms. We used the same corresponding halves for both the CDI and the laboratory measures. In other words, we split the two measures similarly.

Parental report. In split-half analyses, the CDI displayed high internal consistency from the outset of data collection through the end of the study, with values ranging from + 0.71 to +0.96. The top line in Figure 5 illustrates the reliability coefficients for these split-half analyses at each age.

Laboratory measure. The internal consistency of our laboratory measure of vocabulary production was also assessed using split-half analyses for each month. Like the CDI items, the laboratory items were divided into two halves matched for difficulty (age of acquisition on the cross-sectional CDI norms). With the exception of 12 months ($r = +0.81$) at which point the children are at the floor of our measure, internal reliabilities for the laboratory measure tended to be substantially lower than those observed for the CDI until the age of 18 months ($r = +0.74$). After that point, the laboratory measure sustained high internal reliability through 30 months of age (see Figure 5 for correlation coefficients at each month). The low internal reliability of the laboratory measures prior to 18 months places a statistical ceiling on our tests of validation of parental report prior to this age.

Stability

We assessed the month-to-month stability of the CDI subset and the laboratory production set of items by entering the scores from adjacent months into Pearson product-moment correlations.

Parental Report. We first computed Pearson month-to-month correlations for scores based on the CDI subset. The correlations are very high and significant from the outset of data collection ($r = +1.00$, $p < .001$) through the end of the study ($r = +.99$, $p < .001$). The lowest level of reliability was observed between 11 to 12 months of age, and even that was quite high ($r = +.83$). The top line in Figure 6 includes all month-to-month correlations for the CDI.

Laboratory measure. In the early months of data collection, month-to-month correlations for our laboratory measure were not as reliable as the CDI. The reliability of our laboratory measure reached levels comparable to the CDI only beginning at 16 to 17 months ($r = +.77$, $p < .001$). However, once this level was reached, correlations remained high until the end of the study (29 - 30 months, $r = +.97$, $p < .001$). These figures are plotted together with those of the CDI measure in the correlogram depicted in Figure 6.

Validity

We assessed the concurrent validity of our production measures by comparing scores from both the parent and laboratory measures for the same months. Pearson product-moment correlations were computed comparing the percentage of words produced on the CDI subset with the total percentage of words produced in the laboratory. At 18 months, the correlation coefficient was high and significant ($r = +.63$, $p = .001$). For the remaining months of the study, the values fluctuated between a low of +.61 at 28 months and a very robust +.81 at 22 months, all the while displaying at least .05 significance levels. Figure 5 shows the correlation coefficients for each month along with the figures representing the internal reliability of our measures. Figure 6 also shows the correlation coefficients for each month along with the figures representing the stability of our measures.

Growth Functions

Two separate quadratic growth functions were calculated for each child. One was based on the subset of 36 CDI items and the other was based on the 36 items in the laboratory task (because all of the analyses that we have conducted so far indicate equivalence for the full CDI and the 36-item subset, we restricted these analyses to the CDI subset only). These curves provided us with four parameter estimates for each child, for each outcome measure: an intercept, a slope (linear component), a quadratic component, and a mean squared error. The corresponding components were then compared across measures to see how well the shape of growth on the CDI correlated with the shape of growth on our laboratory measure.

When we assessed the degree of relationship between the growth components for the CDI and those for the laboratory measure, both the linear ($r = +.70$, $p < .001$) and the quadratic ($r = +.66$, $p < .001$) components were highly correlated. In other words, individual differences in the shape of change on the corresponding laboratory and parental report measures were quantitatively very similar, a powerful form of developmental validation that goes well beyond the month-to-month correlations reported above. This shows that the shape of growth is the same whether it is measured by parent report or in the laboratory. However, the y-intercepts of these growth functions were not correlated ($r = +.17$), which means that the laboratory and CDI measures “got off the ground” at unrelated points in development. Nonetheless, the y-intercepts were not significantly different ($t(27) = -1.22$, $p = .24$). The mean squared errors for the CDI curves averaged .06, with a range from .02 to .14. The mean squared errors for the laboratory averaged .07, with a range from .02 to .11. Our two error terms were unrelated ($r = -.24$, $p = .22$, indicating that in general when the curve is a good “fit” for one measure, it isn’t necessarily a good “fit” for the other measure, and vice-versa. In other words, error variance is just error variance.

It is important to underscore what these correlations in components of growth add to the qualitative similarities observed in the mean curves illustrated in the first section of the results above (Figure 1). Because the components of growth were calculated for each individual child, for their respective laboratory and CDI scores at each month, these correlations reflect a tremendous range of individual growth patterns: They show that despite enormous individual variation, children who develop slowly when assessed with the CDI tend to show slow development when assessed with the laboratory task, and children who display steep, rapidly accelerating patterns of growth when assessed with the CDI tend to display similar patterns when assessed with the laboratory task as well. To illustrate this point, Figures 7a-c illustrate similarities in the growth functions on the CDI subset vs. the laboratory task for three children: one "late talker" (Figure 7a), one "average child" (Figure 7b), and one "early talker" (Figure 7c). In all three cases, we can see that the CDI provides an earlier estimate of vocabulary growth than the laboratory task; however, the patterns of growth are strikingly similar across measures. Hence Figures 7a-c provide a visual confirmation of the highly correlated growth components that we have just described.

CDI/ Laboratory Lag

The above analyses indicate that individual differences in profiles of growth are similar for the CDI and the laboratory measure, but the two measures are unrelated in their point of onset (as measured by the y-intercept). Specifically, as we also saw in qualitative comparisons of growth scores, parents report changes in vocabulary production before their children demonstrate those changes in the laboratory. To learn more about these differences in onset, we compared the CDI and grab bag task along another dimension – which we will call LD50. LD50 is defined as the point at which, for each of the growth functions for each child, growth on the measure is 50% of the distance from asymptote. In other words, since the maximum score possible on each measure was 36, we were interested in calculating the age at which the children had scored an 18. We computed an LD50 for both the CDI and the laboratory measure, for each child. We then correlated these two LD50 measures across children. The relationship between the LD50s for the CDI and the LD50s for the grab bag was positive and significant ($r = .74$, $p < .001$). Children tended to achieve half of the attainable words on the CDI and the grab bag task during the study at about the same time.

We were also interested in whether the ages at which the children reached the halfway point were related to our mean squared errors. The relationship between the LD50s for the CDI and the mean squared errors for the CDI was negative and significant ($r = –.41$, $p <.05$) as was the relationship between the LD50s for the laboratory and the mean squared errors for the laboratory ($r = –.53$, $p <.01$). These correlations indicate that those children who reached the halfway point earlier were the children whose data were more difficult to fit with the growth functions, regardless of which measure was used.

Next we computed a "lag" score, or the difference between the LD50s for the two measures. We were interested in the size of the disparity between the two measurements at the 50% boundary. The CDI-Laboratory lag scores ranged from a minimum of .02 to a maximum of .86 with a mean of .34 (SD = .22). This means that on average, the age difference between when the children reach 18 words on the CDI and the laboratory is 124.1 days or about 4 months. We also wanted to see whether the "lag" between the two measures was related to the goodness of fit (or error) in each of our growth curves. Correlation of the "lag" scores and the error scores indicate that in general, the size of the disparities between the two measurements was not related to whether the growth curves were a good "fit" for either the CDI ($r = .05$, $p = .80$) or the laboratory measure ($r = -.23$, $p = .23$). In addition, the CDI-Laboratory lag scores were not related to either 30-month CDI scores ($r = .30$, $p = .15$) or 30-month grab bag scores ($r = –.09$, $p = .67$). We may conclude that higher or lower vocabularies at the end of testing were unrelated to the disparities between the CDI and grab bag growth curves at the 50% boundary.

Virtual CDI

Finally, we computed the scores for another set of 36 CDI words that were matched to the overlapping CDI/grab bag words for age of acquisition. The scores for this "virtual CDI" set were then entered into the same analyses as the other two sets of CDI scores. Figure 8 shows that the virtual set of CDI vocabulary words yields results that are very similar to those obtained with the original subset of CDI items. The curve based on mean scores for the new set of words lies practically on top of the curve based on the original subset of 36 CDI items (Figure 8). In addition, the virtual set is also internally reliable and stable as indicated by Figure 9. Values obtained from split-half analyses for our virtual set of vocabulary words are high at the outset of testing ($r = +.73$) and do not fall below this level throughout the study. The reliability coefficients bounce around between a low (but still very impressive) +.89, $p <.001$, to a high +.99, $p <.001$. Correlation coefficients between the virtual set and the laboratory measure follow the same course as that between the overlapping set and the laboratory (Figure 9). At 18 months, the correlations are high and significant ($r = +.68$, $p < .001$) and remain at least at the .05-significance level until the end of the study at 30 months ($r = +.45$, $p < .05$).

DISCUSSION

We have shown in several ways that the CDI, a parental report, and the grab bag task, a laboratory measure of vocabulary development, yield similar patterns of growth. However, the curves constructed from the CDI were consistently higher (i.e. earlier onset and earlier growth) than the laboratory growth curves, indicating that parental report may provide the earliest indicators of vocabulary development, tapping into phenomena that are not yet measurable with common laboratory tasks.

Both the laboratory and parental report provide evidence for immense individual variation in vocabulary size and growth during this period of development. Feldman et al. (1999) have noted that the variance obtained with the CDI is huge, with standard deviations exceeding the mean at many months. They suggest that this fact may represent a psychometric limitation of the CDI. However, because we have observed equivalent variance on our laboratory measure, and because the individual differences on the laboratory measure and the CDI are highly correlated beyond 16 months of age (when the reliability ceiling of the laboratory measure finally lifts), we may conclude that the variance is not a psychometric artifact. Instead, these individual differences are real. Large variation in the rate of vocabulary acquisition is an accurate description of development.

Most important for our purposes here, the individual differences uncovered by these two measures were highly correlated. This was evident in the correlation between measures at each month after 16-18 months of age. These results are similar to those reported in other studies comparing the CDI to laboratory measures. Correlations were lower before that point, but a careful look at both the split-half and test-retest reliabilities for parental report vs. the laboratory task indicates that the low correlations prior to 16 months are due primarily to the lower internal reliability of the laboratory task, which limits its ability to serve as a criterion measure for the validation of parental report. Overlapping the reliability and validity graphs makes this point quite clearly (Figure 5, Figure 6). Laboratory performance, particularly at these early ages, is unreliable, highlighting the importance of valid and reliable parental report measures such as the CDI. A particularly strong form of developmental validation comes from our comparisons of the linear and quadratic components of the growth curves on the lab measure and the CDI, within and across individual children. The high correlations that we obtained on these two measures suggest that individual differences in the shape of vocabulary growth are similar for parental report and our laboratory task.

Although parental report and the laboratory task yield similar means and variances, with high correlations in patterns of growth, parental report consistently yields an earlier estimate of development both within and across children. On our growth curves, the y-intercept represents the child's age level at the point where vocabulary starts to rise. These y-intercepts were uncorrelated. Parents reported words prior to the children saying these words in the laboratory but the difference between the two ages was not statistically different. These y-intercept data were in striking contrast to the significant correlations observed between the linear and quadratic components of growth after that point, and to the high correlation between the LD50s for the two measures (i.e., the age at which children produced half of the items in the laboratory was highly related to the age at which they produced half of the same item subset on the CDI). Thus, individual differences in speed of growth and the "shape of change" in vocabulary learning appear to be very similar for parental report and our laboratory measure. However, the point at which parents first detect the onset of expressive vocabulary is not necessarily related to the point at which the same children will start to display those same skills in the laboratory. We may speculate that these differences in onset may reflect temperamental factors (in both the children and their parents) that are partially independent of individual differences in vocabulary learning.

Finally, we found that consistencies between the CDI and the laboratory measure were not limited to one particular set of 36 items. Rather, results based on another set of 36 CDI items, matched for age of acquisition to the original set, strengthen the argument that parental report can provide a valid estimate of individual differences in lexical ability. This analysis shows that differences in vocabulary acquisition are not tied to specific items. Scores on this “virtual set” of vocabulary items showed the same high level of consistency and internal reliability that was observed with the original set. In addition, the concurrent validity correlation coefficients based on our laboratory measure and this virtual set yielded similar values to correlations between identical lists of items.

In sum, the present results indicate that the CDI provides valid and reliable estimates of productive vocabulary across a period of dramatic language growth. We carried out six separate analyses and each supplied evidence of the similarities between the CDI and the laboratory measure. In some instances, the CDI proved to be the more reliable and internally consistent measure. Taken together, these results suggest that the laboratory measure and parental report yield the same patterns of vocabulary growth on slightly different timelines. The differences in timing may be due to depth of knowledge about the words or a child's temperament in a laboratory setting, but they should not lead to different interpretations about the way vocabulary develops. Given that it is inexpensive and can be more representative than laboratory measures, parental report is both a valuable tool for diagnosis of language delay and for many basic research questions.

Acknowledgments:
Data collection was supported by a grant from the John D. and Catherine T. MacArthur Foundation Research Network on the Transition from Infancy to Early Childhood. Support for later stages of data collection was provided by NIH/NIDCD Program Project P50 DC01289-0351 (Origins of Communication Disorders) to Elizabeth Bates. Please address correspondence to: Jennifer Jahn-Samilo, Department of Psychology, 0109, University of California at San Diego, La Jolla, CA 92093-0109 (858-822-1569 -- jahn@crl.ucsd.edu).

REFERENCES

Corkum, V. & Dunham, P. (1996) The Communicative Development Inventory-WORDS Short Form as an index of language production. *Journal of Child Language, 23,* 515-528.

Dale, P. S. (1991). The validity of a parent report measure on vocabulary and syntax and 24 months. *Journal of Speech and Hearing Research, 34,* 565-571.

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

Darwin, C. (1877). A biographical sketch of an infant. *Mind, 2,* 286-294.

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., Dale, P., Reznick, J.S., Bates, E., Thal, D.J., & Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59,* (5, Serial No. 242).

Fenson, L., Dale, P;., Reznick, J.S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *MacArthur Communicative Development Inventories: User's Guide and Manual.* San Diego: Singular Publishing Group, Inc.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Lanuage, 20,* 523-549.

Leopold, W.F. (1949). *Speech development of a bilingual child: A linguist's record (Vol. 3).* Evanston, IL: Northwestern University Press.

Miller, J.F., Sedey, A.L., & Miolo, G. (1995). Validity of parent report measures of vocabulary development for children with Down syndrome. *Journal of Speech and Hearing Research, 38,* 1037-1044.

O'Hanlon, L., & Thal, D. (1991, November). *MacArthur CDI/Toddlers: Validation for language impaired children.* Paper presented at the meeting of the American Speech-Language-Hearing Association, Atlanta.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Reznick, J.S., & Goldfield, B. A. (1994). Diary vs. representative checklist assessment of productive vocabulary. *Journal of Child Language, 21,* 465-472.

Ring, E.D., & Fenson, L. (in press). The correlation between parent report and child performance for receptive and expressive vocabulary beyond infancy. First Language.

Saudino, K.J., Dale, P.J., Oliver, B., Petrill, S.A., Richardson, V., Rutter, M., Simonoff, E., Stevenson, J., & Plomin, R. (1998). The validity of parent-based assessment of the cognitive abilities of 2-year-olds. *British Journal of Developmental Psychology, 16,* 349-363.

Smith, L.B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The emergence of language* (pp 277-303). Mahwah, NJ: Lawrence Erlbaum.

Stiles, J. (1994). On the nature of informant judgments in inventory measures: . . . and so what is it you want to know? *Monographs of the Society for Research in Child Development, 59* (5, Serial No. 242).

Tomasello, T., & Mervis, C. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development, 59,* (5, Serial No. 242).

Table 1
Age switched from Word and Gestures to Words and Sentences, of Missed CDIs and of Missed Laboratory Sessions

Participant	Sex	Age switched	CDI age Missed	Lab age missed
1	M	19[a]	8,28,29	27-30[b]
2	M	17	8,23	
4	F	17	8,13,22,25,29	22,25,29
6	M	17		
7	F	17		
10	F	17		22
11	F	17	8,14,18,25	14,18,25
12	M	20	13	
13	M	17		
14	M	17	28,30	
15	F	17	8	17
16	M	18	8	
17	M	17	12,13,15,18,21,23,25,26,29	15,17,19,21,23,25,26,28,29[c]
19	M	17		
22	M	17		
24	F	17	29	26,29
26	M	17		24
27	F	17	14,30	
28	F	23		
30	F	18		
31	M	26		
32	M	17	13,22,24,25,27-29	22,24,26,28[d]
34	M	17	30	19,29
35	F	17	20,30	18-30[e]
36	F	17	27,29	23,29
37	M	18	8,24,26,27	27-30[f]
38	M	17	21	21
41	M	18	8	

Note: Children who missed their 8-month CDIs were inducted into the study after their 8-month birthday.

[a]ID1 was switched at 19 months of age though his mother was asked to complete both inventories until 30 months of age.

[b]ID1 moved out-of-state prior to his 27-month session. His parent continued to complete the CDI even though he did not come to the laboratory.

[c]ID17 participated bi-monthly due to extensive driving distance.

[d]ID32 had transportation problems and often rode the bus to sessions.

[e]ID35 moved out-of-state prior to her 18-month session. Her parent continued to complete the CDI even though she did not come to the laboratory.

[f]ID37 moved out-of-state prior to his 27-month session. His parent continued to complete the CDI even though he did not come to the laboratory.

Table 2

Mean percentage scores and standard deviations on the total CDI, the matching set of CDI words and the laboratory

	Mean score (SD)		
Age in months	CDI total	CDI matched set	Laboratory
8	0.2 (0.7)	1.0 (3.7)	
9	0.3 (0.8)	0.8 (3.2)	
10	0.6 (1.1)	1.1 (3.3)	
11	1.0 (1.5)	1.9 (4.8)	
12	1.4 (2.0)	3.6 (6.6)	0.6 (1.7)
13	2.4 (3.5)	5.6 (8.5)	0.9 (1.9)
14	3.0 (4.5)	6.3 (9.2)	1.6 (2.8)
15	5.0 (7.1)	12.3 (14.3)	2.6 (3.4)
16	7.4 (8.7)	16.8 (15.5)	5.4 (5.2)
17	12.5 (11.6)	24.5 (17.9)	8.3 (6.8)
18	16.3 (14.2)	29.3 (18.4)	8.5 (8.0)
19	21.0 (15.3)	33.8 (18.4)	13.9 (12.5)
20	27.8 (19.4)	39.7 (21.0)	17.9 (14.4)
21	33.9 (23.4)	46.7 (24.4)	22.9 (14.1)
22	41.7 (26.6)	52.5 (25.2)	27.4 (19.8)
23	48.1 (26.9)	57.9 (25.2)	36.8 (22.8)
24	56.0 (27.0)	65.5 (24.6)	38.2 (24.6)
25	60.1 (28.8)	70.3 (24.5)	48.1 (24.3)
26	67.1 (27.4)	75.2 (22.9)	54.3 (21.3)
27	73.1 (27.1)	79.7 (22.6)	62.6 (16.3)
28	78.2 (24.5)	83.9 (17.4)	68.5 (14.8)
29	80.4 (23.0)	84.8 (16.5)	71.8 (13.7)
30	81.6 (21.6)	84.8 (16.0)	70.9 (13.6)

FIGURE CAPTIONS

Figure 1. Points represent the mean percentage of words on (a) the complete Words and Gestures form and the Words and Sentences form of the CDI (b) the subset of the Words and Gestures form and the Words and Sentences form of the CDI and (c) the laboratory measure.

Figure 2. Percentage of words on the complete Words and Gestures form and the Words and Sentences form of the CDI reported to be produced by children at each month-median values and spread of score distributions.

Figure 3. Percentage of words on the subset of the Words and Gestures form and the Words and Sentences form of the CDI reported to be produced by children at each month-median values and spread of score distributions.

Figure 4. Percentage of words produced on the laboratory measure of production at each month-median values and spread of score distributions.

Figure 5. Across the months of the study, the internal reliability of (a) reported production on the subset of the Word and Gestures form and Words and Sentences form and (b) production on the laboratory measure. Concurrent validity between the CDI forms and the laboratory measure.

Figure 6. Across the months of the study, the month-to-month stability of (a) reported production on the subset of the Words and Gestures form and the Words and Sentences form and (b) production on the laboratory measure. Concurrent validity between the CDI forms and the laboratory measure.

Figure 7a. For a late talker the percentage of words (a) reported to be produced on the CDI and (b) produced on the laboratory measure of production. CDI and laboratory measure medians.

Figure 7b. For an average talker the percentage of words (a) reported to be produced on the CDI and (b) produced on the laboratory measure of production. CDI and laboratory measure medians.

Figure 7c. For an early talker the percentage of words (a) reported to be produced on the CDI and (b) produced on the laboratory measure of production. CDI and laboratory measure medians.

Figure 8. Mean percentage of words reported to be produced on the (a) the subset of the Words and Gestures form and the Words and Sentences form and (b) the “virtual" subset of words from the Words and Gestures form and the Words and Sentences form. Mean percentage of words produced on the laboratory measure of production.

Figure 9. Across the months of the study, both internal reliability and month-to-month stability for the “virtual” subset of words from the CDI. Concurrent validity between the “virtual” subset of words from the CDI and the laboratory measure.

Fig. 1: Mean of overall CDI, subset of CDI and Laboratory

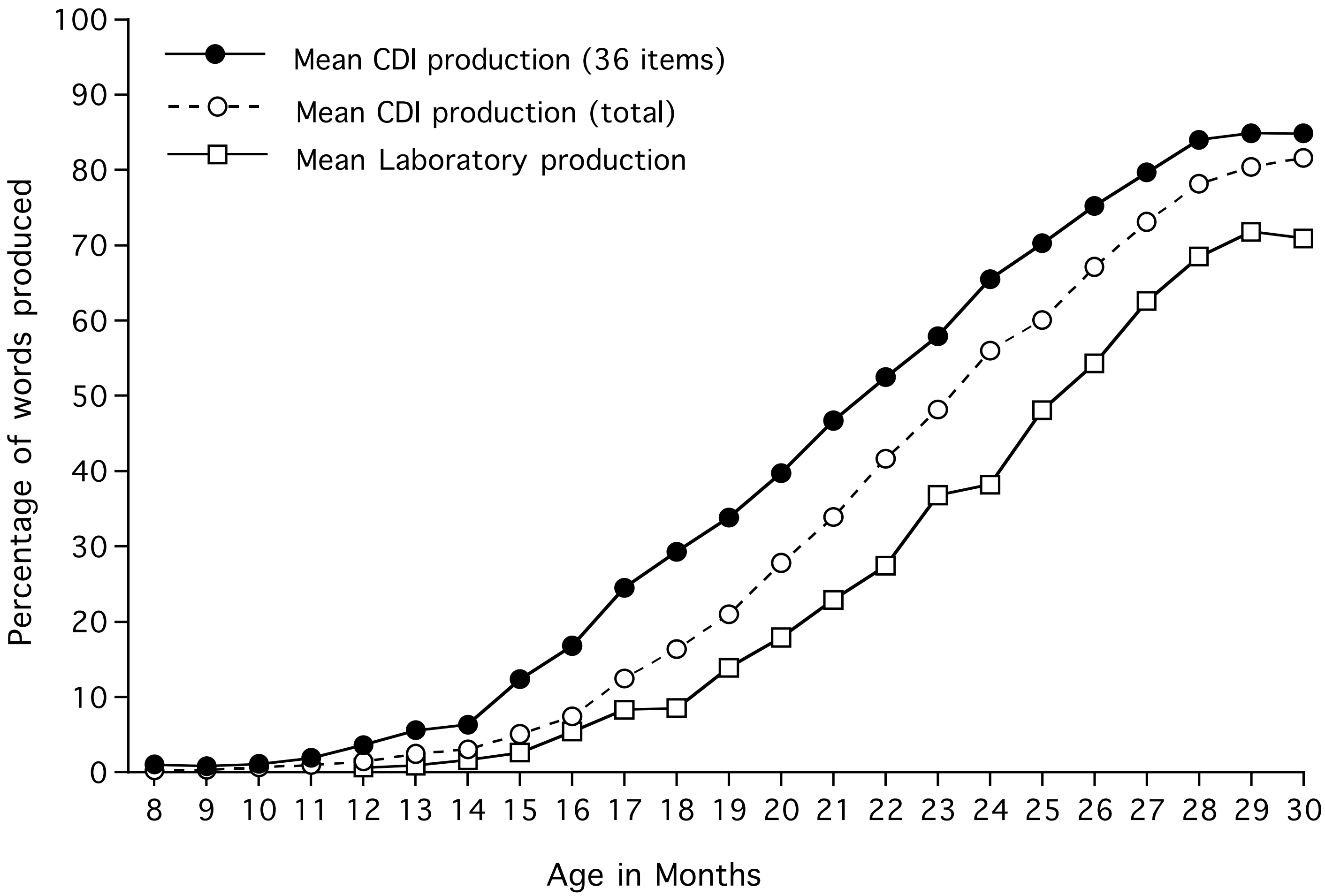


Fig. 2: Percentiles for CDI production (proportion by 680)

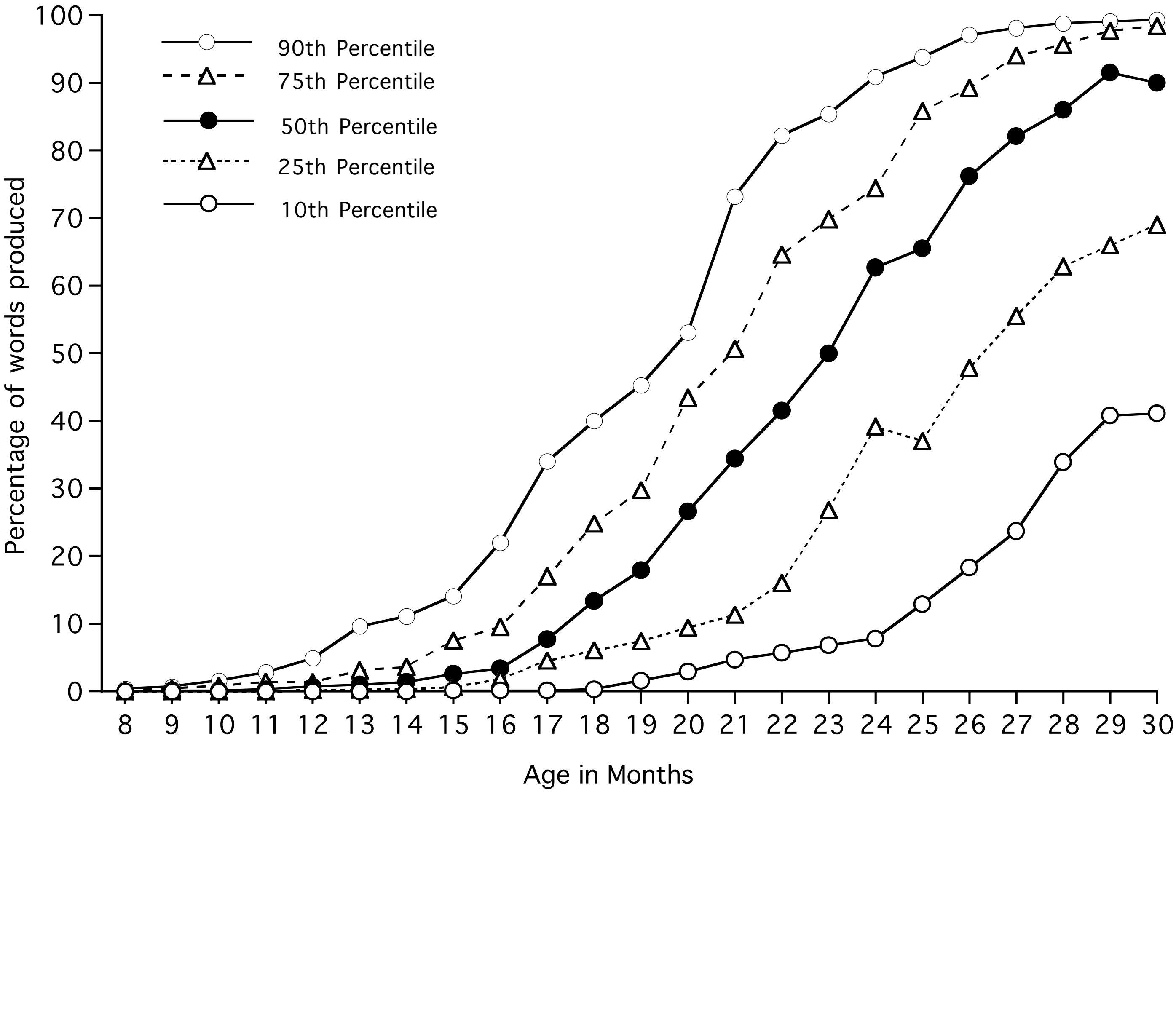


Fig. 3: Percentiles for CDI production matching items

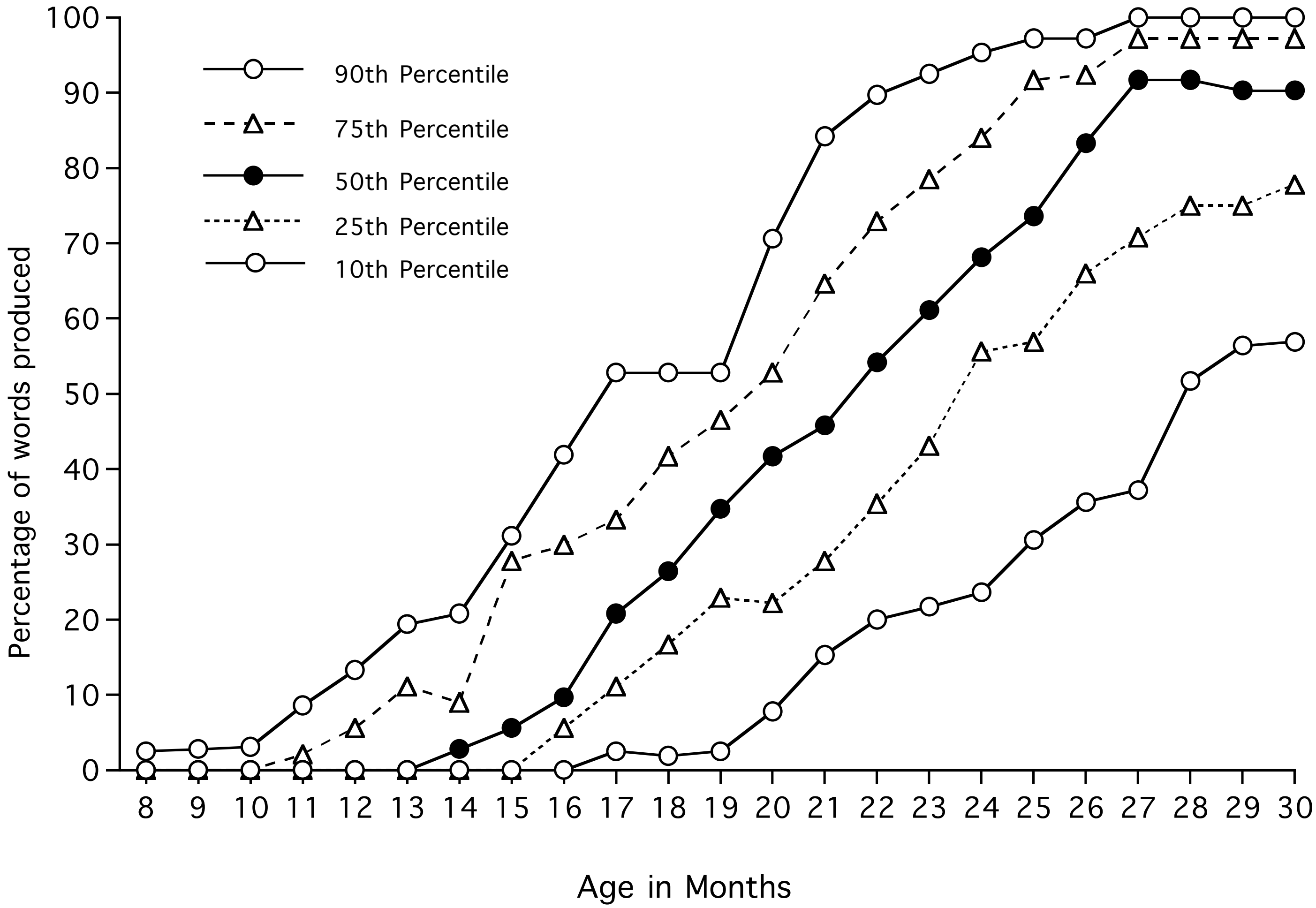


Fig. 4: Percentiles for lab production

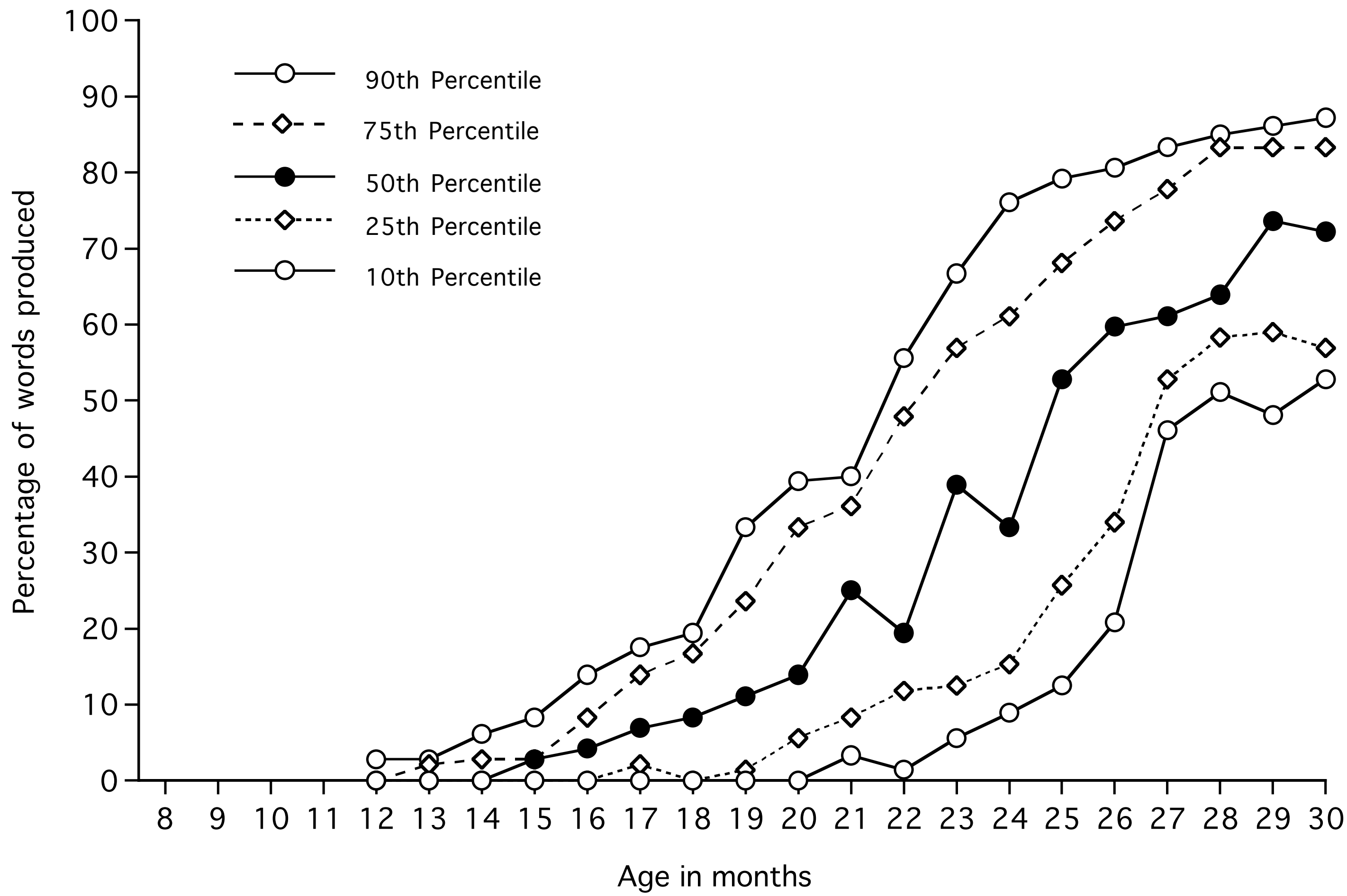


Fig. 5: Split half rel for CDI items and lab with validity

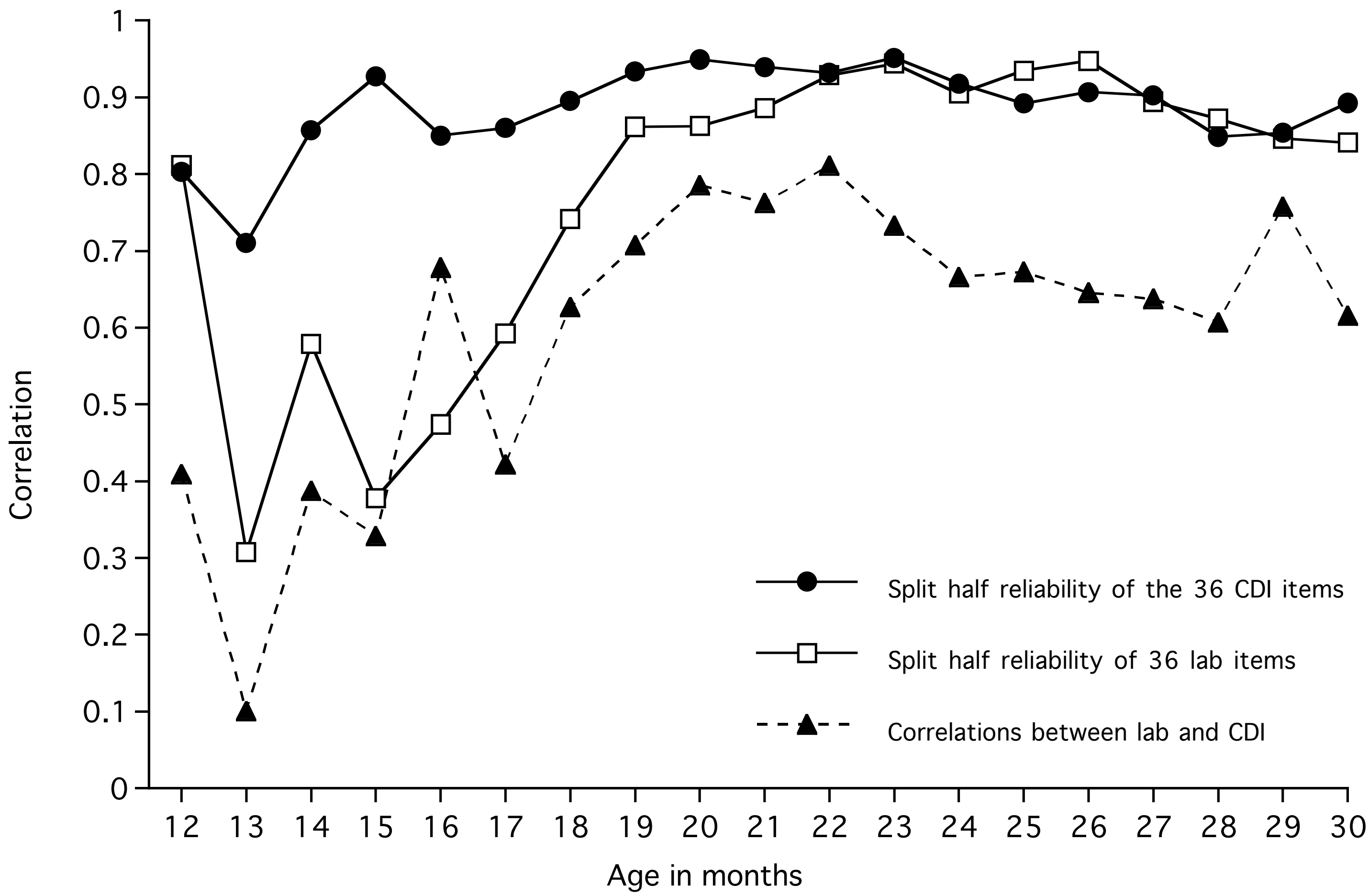
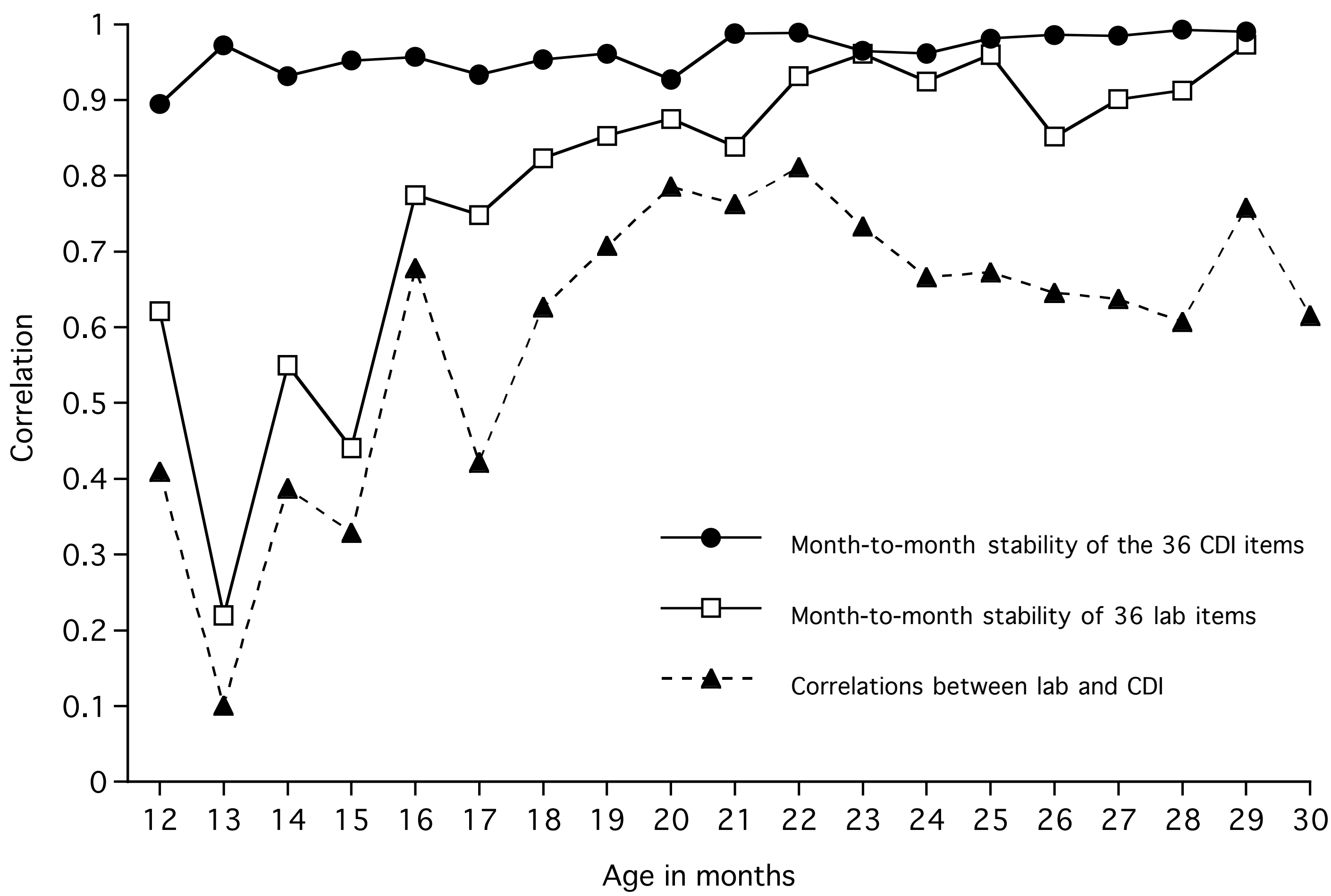


Fig. 6: Test retest rel CDI items and lab with validity



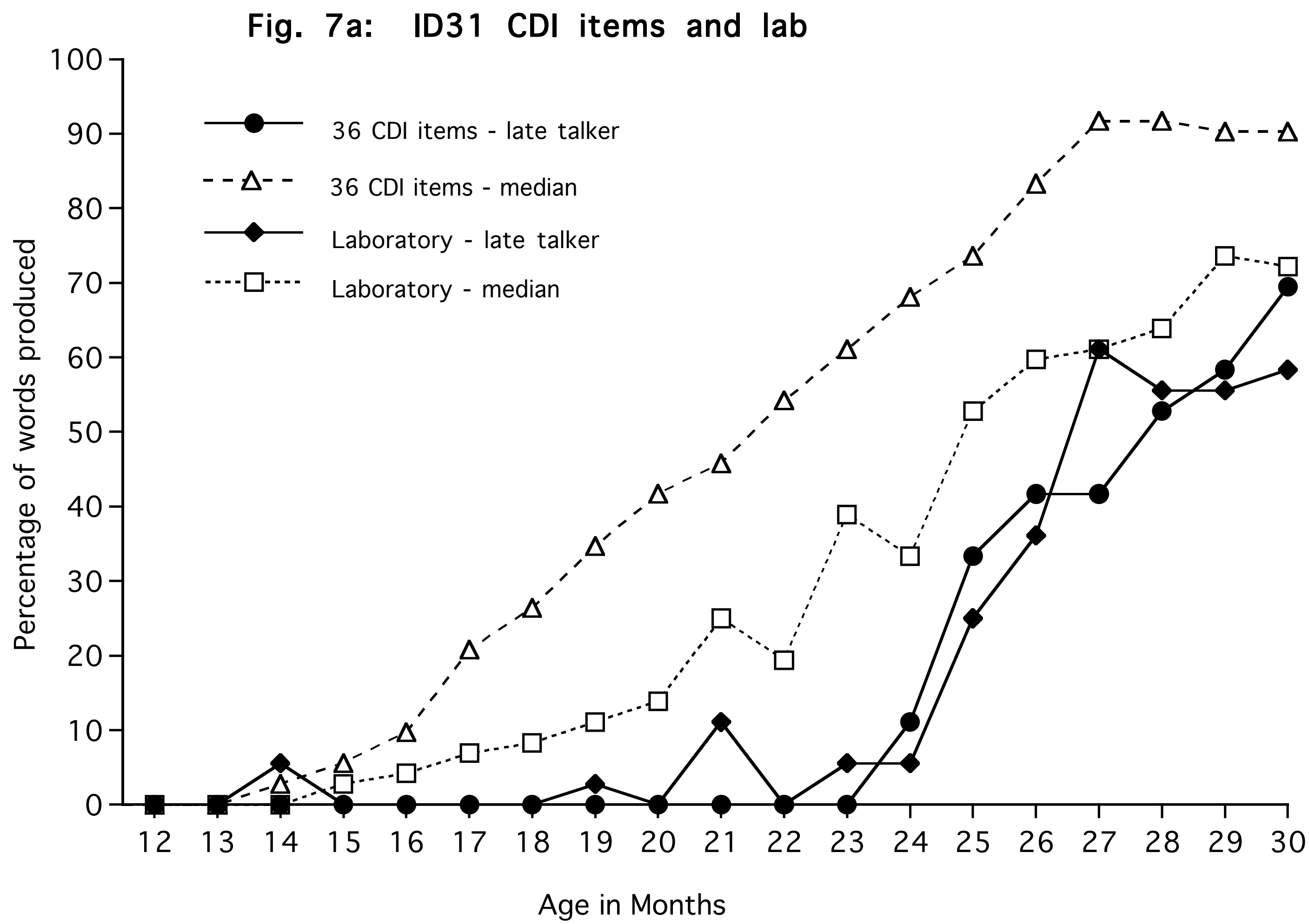


Fig. 7a: ID31 CDI items and lab
Percentage of words produced
0
10
20
30
40
50
60
70
80
90
100
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
Age in Months
36 CDI items - late talker
36 CDI items - median
Laboratory - late talker
Laboratory - median

Fig. 7b: ID15 CDI items and lab

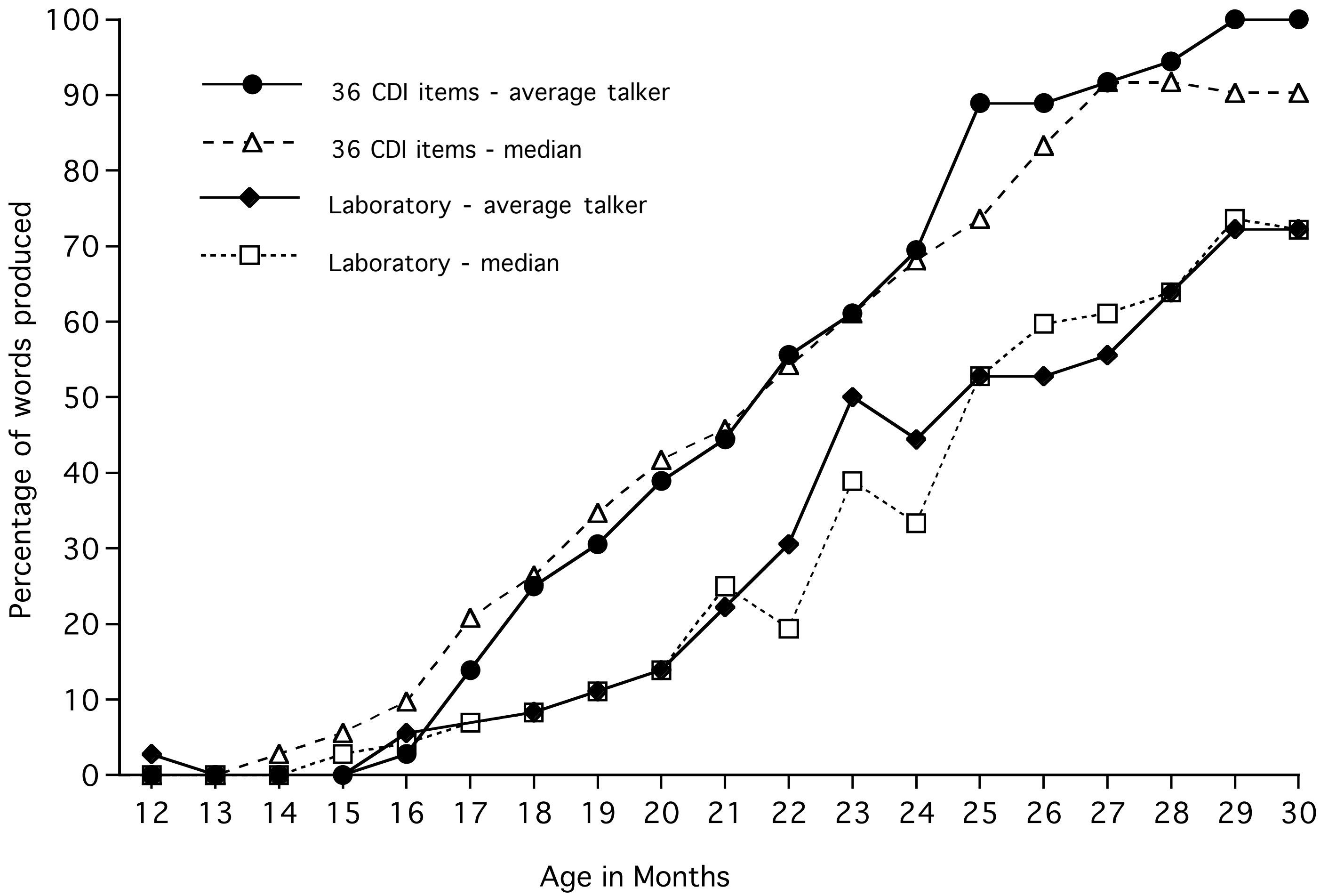


Fig. 7c: ID22 CDI and lab

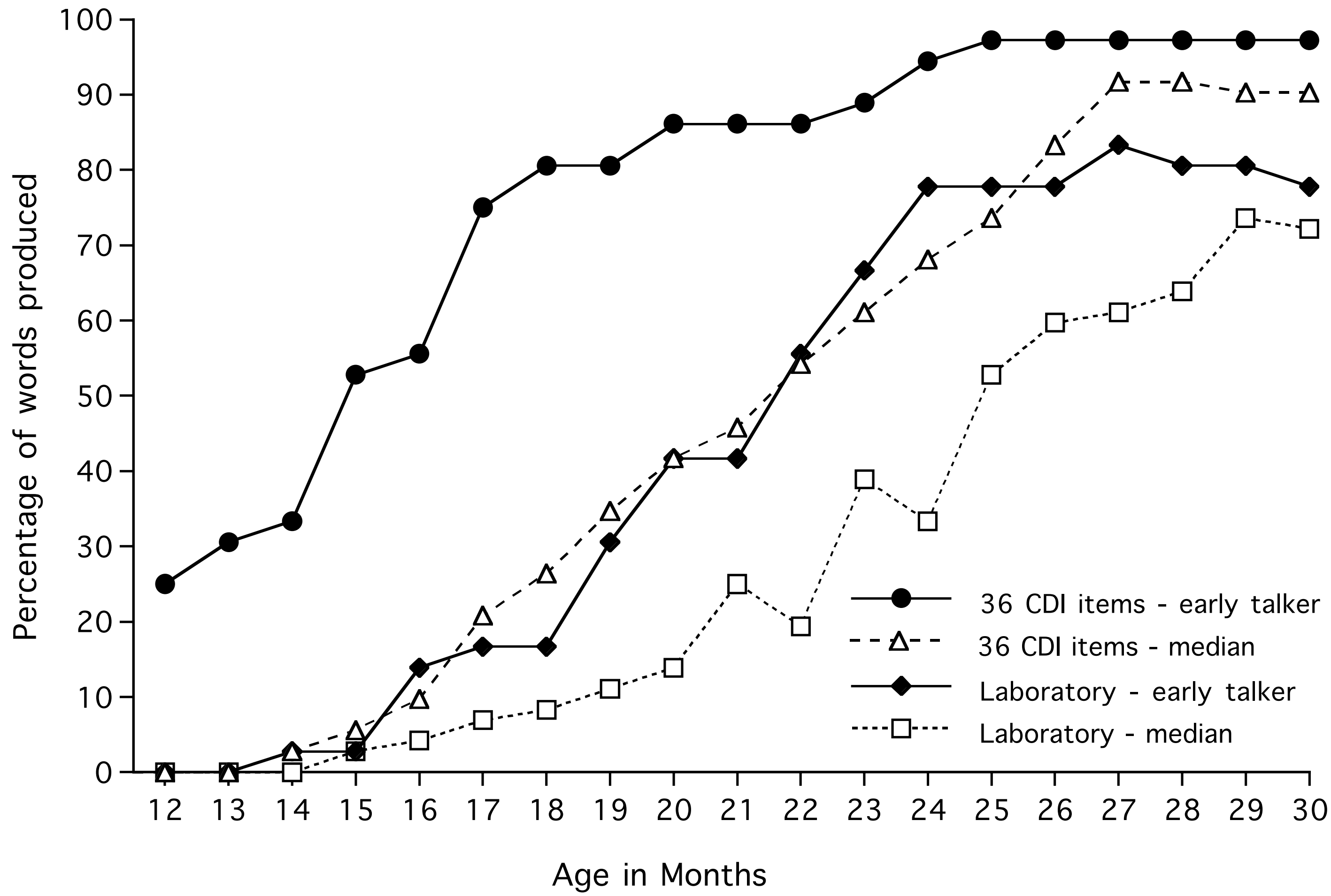


Fig. 8: Mean of overall CDI, subset of CDI and Laboratory

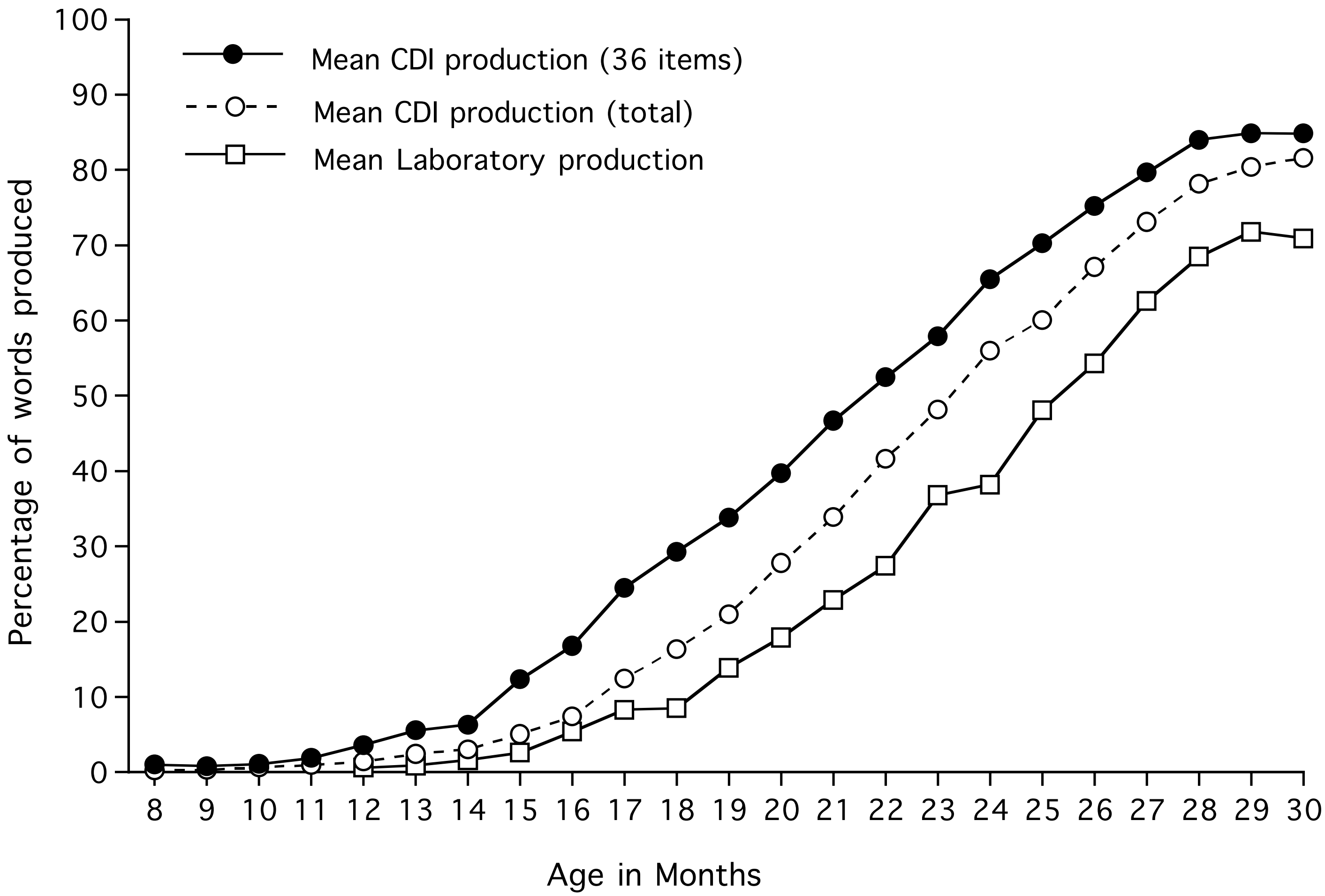


Fig. 9: Split half and test retest virtual with validity