Child Development, March/April 2000, Volume 71, Number 2, Pages 323–328

Measuring Variability in Early Child Language: Don't Shoot the Messenger

Larry Fenson, Elizabeth Bates, Philip Dale, Judith Goodman, J. Steven Reznick, and Donna Thal

Feldman et al. criticize the MacArthur Communicative Development Inventories (CDIs) as having too much variability, too little stability, and insufficient ability to predict early language delay. We present data showing that these characteristics of the CDI are authentic reflections of individual differences in early language development rather than measurement deficiencies. We also respond to their critical assertions concerning sociodemographic influences on the CDI scores.

INTRODUCTION

Feldman et al. (this issue), present MacArthur Communicative Development Inventories (CDI) data from an unusually large and demographically well-balanced sample of children ($N = 2{,}156$). This large data set offers fresh CDI data on growth trends, variability, gender effects, and social class effects, with findings that largely corroborate the CDI normative study and subsequent reports. However, Feldman et al. also raise some methodological challenges that would (if their interpretations are correct) greatly limit the utility of this parent report instrument for the study of early language development. In this commentary, we explain why we believe that Feldman et al. have been too pessimistic.

DEVELOPMENTAL SENSITIVITY OF THE CDI: A REPLICATION

One of the most significant contributions of Feldman et al. is their confirmation of the ability of the CDI to index the growth of language skills. The major scales show generally increasing monotonic growth in the age periods from 10 to 13 and 22 to 25 months (Tables 3 and 5, respectively, in Feldman et al.). Moreover the means and standard deviations obtained in the Pittsburgh study are comparable to those obtained in the CDI normative study, as shown in Table 1 in this paper. They also find gender differences similar in size and direction to the ones that we report (in our sample, girls are one month ahead of boys on average, but the difference accounts for less than 2% of the immense variation observed within and across ages). We were pleased to see such striking correspondence between the data presented by Feldman et al. and our own, and we view this as a satisfying replication that ought to increase confidence in the reliability and utility of the CDI. And yet Feldman et al. reach very different conclusions, as follows.

Commentary on Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky, & Paradise, "Measurement Properties of the MacArthur Communicative Development Inventories at Ages One and Two Years."

IS THE VARIANCE REAL?

Among other things, Feldman et al. are worried that the huge individual differences represented in our Table 1 reflect serious flaws in the CDI that limit its applicability. For example, they note repeatedly that the standard deviations match or exceed mean scores for many CDI measures, and they imply that such standard deviations are unacceptably large. We believe instead that these figures reflect true variation in language development from 8 to 30 months (see Fenson et al., 1994). To illustrate this point, compare Figures 1 and 2, based on a longitudinal study (Jahn-Samilo, Goodman, Bates, & Appelbaum, 1999). Figure 1 presents parent-report data for 36 object names (taken from the CDI word production checklist, administered monthly from 8 to 30 months). Figure 2 displays child performance for the same 36 words in an elicited word production task administered monthly in the laboratory from 12 to 30 months.

Although parents report word knowledge earlier than was observed in the laboratory, both measures show great variability between children from 16 to 30 months of age. Further, the standard deviation is greater than the mean at some ages for the laboratory measure as well. Thus, the vast variability exhibited by the CDI does not reflect any psychometric deficiency; rather, it reflects a fact about early language growth that any valid measure must faithfully record.

STABILITY OF INDIVIDUAL DIFFERENCES ACROSS THE SECOND YEAR OF LIFE

Feldman et al. examined correlations across a one-year interval between four different infant scale mea-

 0009-3920/2000/7102-0005

Table 1 MacArthur Communicative Development Inventories (CDI) Scores in Two Samples

	Age in Months			
CDI: Words and Gestures	10	11	12	13
Phrases understood				
Pittsburgh study	14.0 (6.3)	15.0 (6.4)	16.0 (6.4)	17.4 (6.2)
Norming study	11.5 (6.7)	13.3 (5.8)	15.5 (5.6)	17.6 (6.5)
Vocabulary comprehension				
Pittsburgh study	83.6 (68.8)	92.3 (74.0)	105.0 (77.2)	119.0 (77.4)
Norming study	66.8 (60.2)	78.4 (75.1)	86.4 (49.2)	121.8 (68.9)
Total gestures				
Pittsburgh study	21.8 (8.1)	23.2 (8.8)	26.6 (9.2)	29.7 (9.5)
Norming study	18.4 (8.2)	23.0 (8.0)	27.8 (9.9)	33.2 (11.1)
	Age in Months			
CDI: Words and Sentences	22	23	24	25
Vocabulary production				
Pittsburgh study	249.2 (165.8)	281.9 (178.5)	302.4 (172.8)	312.4 (168.9)
Norming study	268.9 (167.3)	334.9 (156.8)	311.7 (173.7)	366.0 (161.0)
Mean sentence length				
Pittsburgh study	3.0 (1.6)	3.5 (1.9)	3.8 (1.9)	3.9 (1.9)
Norming study	3.8 (1.9)	4.7 (2.2)	4.7 (2.7)	5.5 (2.7)
Sentence complexity				
Pittsburgh study	8.7 (8.1)	9.2 (8.3)	10.2 (8.7)	9.8 (8.5)
Norming study	6.4 (7.3)	10.5 (10.2)	9.1 (9.6)	11.4 (10.2)

Note: Portions of this table have been adapted from Feldman et al. (2000).

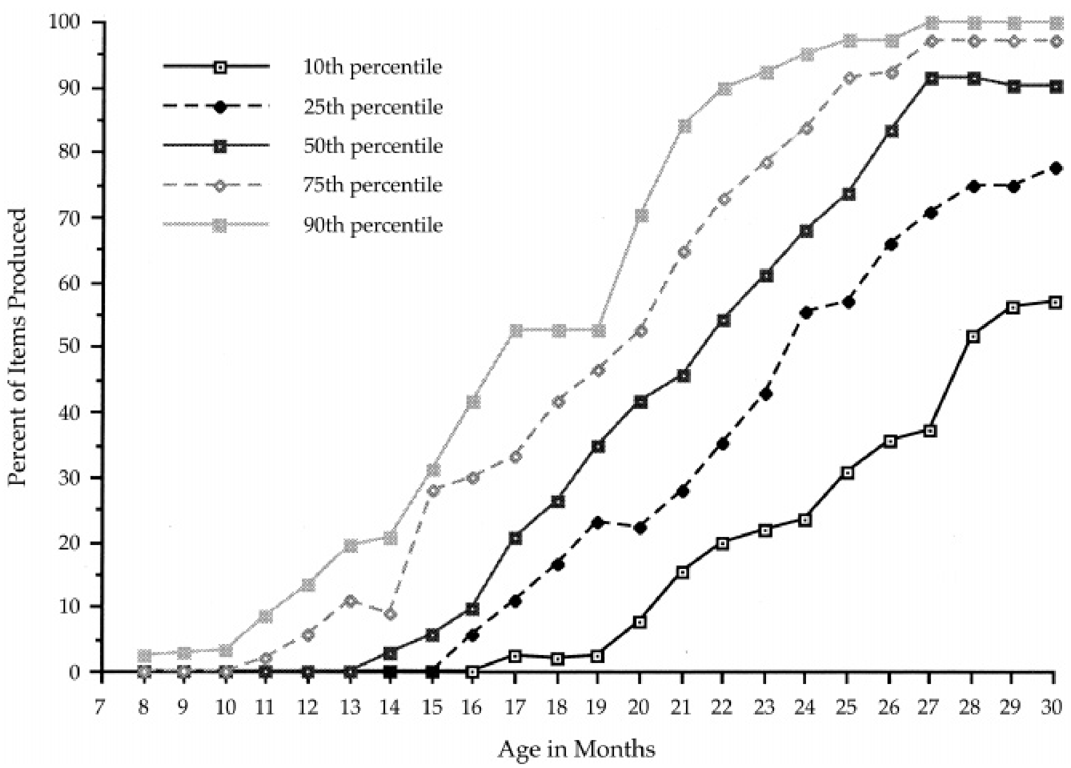

Figure 1 Vocabulary growth on Communicative Development Inventories for the same 36 items used in the laboratory test.

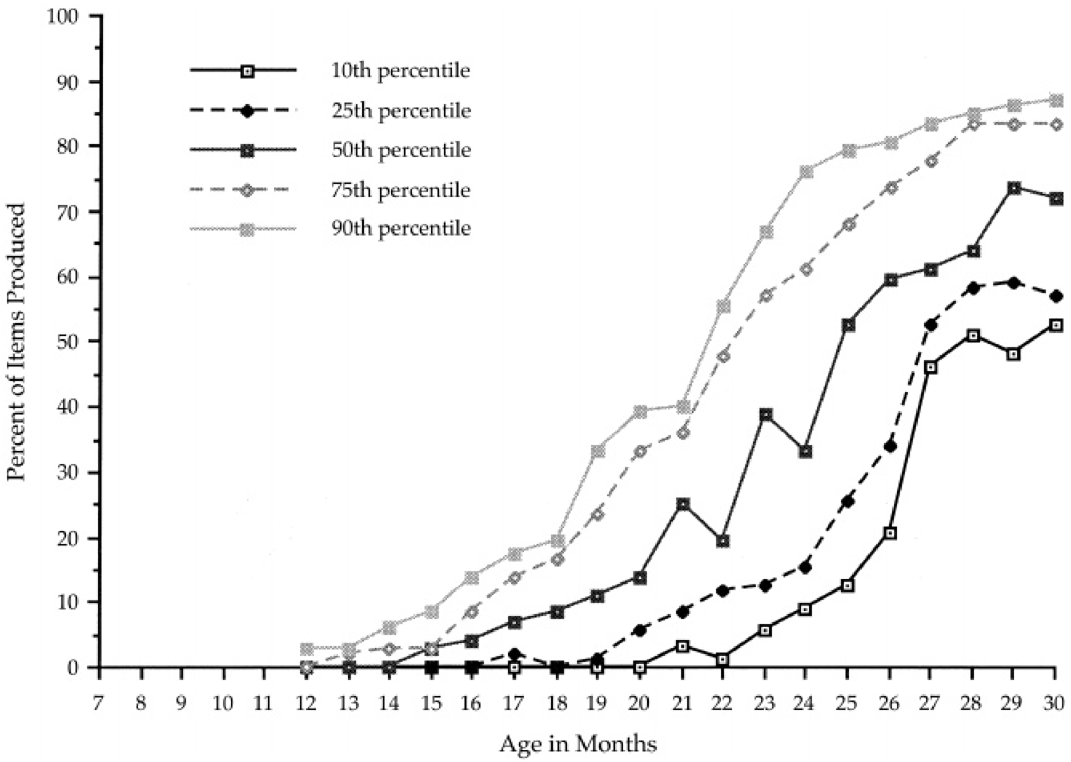

Figure 2 Vocabulary growth in the laboratory task (percent of 36 items).

sures (phrases understood, vocabulary comprehension, vocabulary production, and total gestures) and five toddler scale measures (vocabulary production, irregular word forms, overregularized words, length of the three longest sentences, and sentence complexity). The intercorrelation matrix, presented in their Table 8, shows significant but weak associations between the infant and toddler scales.

Feldman et al. imply that the stability should be much higher, and that some other measure of language development might do a better job. But is this the case? To our knowledge, no one has ever demonstrated stronger correlations in language skills from 1 to 2 years, by *any* method—parent-based, naturalistic, or experimental. For some pairs of variables, there is no theoretical reason to expect a high correlation (e.g., from 12-month gesture to 24-month grammar). For other pairs (e.g., expressive vocabulary from 12 to 24 months), higher correlations might be expected on theoretical grounds but have proven elusive with any technique. For example, vocabulary production between 12 and 24 months on the CDI is correlated at .40 in the Jahn-Samilo et al. (1999) longitudinal sample, but the corresponding correlation for elicited vocabulary production in the laboratory is only .20 in the same children. We should never give up the search for better measures, but we must also be open to the possibility that the finding is real, that is, that individual differences in language ability are quite unstable in this age range.

PREDICTING LANGUAGE DELAY

Feldman et al. are disappointed in the low correlations from 1 to 2 years of age because this means that the instrument cannot be used to predict language delay as early as 1 year of age. We agree that 1 year of age may be much too early to identify individual children who are at risk for language delay; indeed, we have said so ourselves in several publications including the CDI manual itself (Fenson et al., 1993, p. 32). Language skills may simply not be sufficiently developed at age 1 to make accurate assessments. However, Feldman et al. use the low correlations from 1 to 2 years of age to reach a much stronger conclusion: that the CDI has limited utility as a screening device at any age. They present no evidence in support of this extrapolation (e.g., no evidence regarding the low predictive power of the CDI at 24 months for language outcomes one or two years later). We disagree with their conclusion, for two reasons.

First, the low predictive power of the CDI at 12 months is logically and empirically independent of its predictive power at 24 months or, for that matter, at 16 to 18 months. At some point between 16 and 24 months of age, there is a marked increase in the stability of individual differences in many developmental measures (e.g., the substantial increase in prediction of later childhood IQ measures from infant test scores between the first and second year of life, as summarized by McCall, Hogarty, and Hurlburt, 1972). Hence, there are good reasons to believe that, as with other child measures, the predictive power of the CDI will increase across this period. Note that prediction to later measures is especially plausible if we use a measure with good internal reliability (as Spearman pointed out long ago, no measure can correlate any higher with anything else than it correlates with itself). In fact, Jahn-Samilo et al. (1999) have shown that the CDI has high split-half reliability at every age from 8 to 30 months, a psychometric virtue compared with most laboratory measures. For that reason, it is a particularly good candidate as a screening device given that stability increases between 16 and 24 months of age. Feldman et al. should not jettison the CDI or any other measure simply because its predictive power is low at 12 months.

Second, Feldman et al. assume that linear correlation across the full sample is the best and indeed the only valid way to evaluate a screening instrument. In fact, clinically significant language delay concerns only the bottom of the distribution. Note that children with serious biomedical complications who are at risk for language delay are typically excluded from normative studies (e.g. Feldman et al., 1999, and Fenson et al., 1993, both excluded children who were premature or very low birthweight, together with children with diagnosed forms of neurological impairment or mental retardation). As a result, authentic predictability may be missed in such studies. There is an even more important statistical implication for the evaluation of screening measures. To detect children who are (or will be) at the bottom of the distribution, a cutoff score is determined for defining delay at time 1 and at time 2 on the relevant measures. One then determines what proportion of children who fall below the cutoff score at time 1 also fall below the time 2 cutoff. Success rate as well as false positives and false negatives can then be determined. Researchers in this area have also found that a multimeasure approach is more appropriate for predicting continued delay (Olswang, 1998; Paul, 1996, 1997; Thal & Katich, 1996), for example, using word comprehension, word production, and gesture. To be sure, even these predictions are generally restricted to the group level; we are still not in a position to say whether an individual toddler will receive a diagnosis of language delay. But this puts us in a position no different from the one faced by most medical epidemiologists: Certain risk factors, used together, predict the probability of cancer or heart disease, but no one can (or should) tell an individual patient that his fate is sealed.

PERCENTILES, INTERVALS, AND STANDARD SCORES

Feldman et al. also criticize the CDI's reliance on percentiles rather than means and standard deviations. They note that the difference between scores at, say, the 50th and 90th percentiles at earlier ages is smaller than the difference between those ranks at later ages. They conclude that percentile scores are therefore misleading. Once again, however, we would counter that the CDI is simply reflecting the nonlinear nature of development. The growth curves in Figures 1 and 2 illustrate this progressive expansion of variance with age. Feldman et al. suggest that a parametric approach would equalize this difference. But this is not so. The same problem would be encountered if scores were reported in terms of means and standard deviations at each age level. The primary difference between a percentile approach and a *z*-score approach lies in the assumption of normality which underlies the latter—an assumption that clearly does not hold across the early stages of language development. Percentile scores take this "fan effect" and the highly skewed distribution of raw scores into account. Nonetheless, we hasten to add that percentile scores can be misleading when the variance is exceedingly small, which will inevitably happen when skills are just emerging. For that reason, we advise against the use of percentile scoring in early portions of the infant and toddler age ranges for skills that have not yet begun to blossom.

SOCIODEMOGRAPHIC INFLUENCES

The Feldman et al. sample is much more demographically diverse and balanced than our original CDI norming sample and (as far as we know) provides the first large-scale examination of social class effects on the CDI. For this reason, their apparently counterintuitive findings on SES effects are important. Feldman et al. found that maternal education was inversely correlated with two infant form scales at 12 months, phrases understood and vocabulary production, and with two toddler form scales at 24 months, irregular word forms and overregularized word forms. That is, for these measures, mean scores were highest for children of mothers without a high school diploma, intermediate for high school graduates, and lowest for college graduates. In contrast, and in keeping with the great majority of findings on maternal education and child development, Feldman et al. found a positive relation between maternal education and two other toddler/infant scale measures (length of the three longest sentences and sentence complexity).

These inverse correlations have precedent in our own published findings (Fenson et al., 1993). Although our original norming sample was highly skewed toward middle-income families, the demographic composition of the sample did permit a comparison between the children of high school graduates and those of parents with one or more years of college. On the infant form, significantly higher scores were found for one measure (vocabulary composition) for children of mothers with a high school education. This effect was limited to younger infants (from 8 to 12 months).

The combined findings from the Feldman et al. and CDI norming studies suggest that some lower-income parents do overestimate or overreport certain newly emerging language skills, especially in infancy. This may be more likely for boys than for girls (Reznick, 1990). In marked contrast, Arriaga, Fenson, Cronan, & Pethick (1998) found a striking positive relation between income level and CDI scores for toddlers—as one would expect—in that over 75% of 103 toddlers who were younger siblings of Head Start children fell below the median CDI scores for vocabulary production and sentence complexity. Moreover, more than 35% of the sample fell below the CDI 10th percentile on vocabulary production (24% for sentence complexity). 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.

These findings do not mean that the CDI should not be used with low-income samples. Indeed, one might be more concerned if the instrument failed to reveal social class differences. But it does suggest caution in applying and interpreting percentile scores with children at the lower end of the demographic continuum and at the youngest ages—the kind of caution we recommend in the user's manual. Finally, we note that other methods of studying language are also not immune from social class influences, for instance, Labov's (1970) celebrated demonstration of the underestimation of language abilities in African American children when speech is elicited in a formal and unfamiliar context by white investigators.

CONCLUDING REMARKS

Since the CDI was published in 1993, its use has grown markedly among researchers and practitioners. Its cost effectiveness and high validity has made it increasingly an instrument of choice to address a range of practical questions such as the effects of infant day care (NICHD Early Child Care Research Network, 1997) and early intervention (Healthy Steps for Young Children Program, 1999) as well as theoretical issues (e.g., Bates et al., 1994). It has also spawned the development of an expanding number of comparable parent-report forms in well over a dozen foreign languages. The existence of these instruments is, in turn, permitting cross-linguistic questions to be investigated with considerably larger sample sizes than have been characteristic of such studies in the past (Bates & Goodman, 1997; Caselli et al., 1995; Caselli, Casadio, & Bates, 1999). In each of these ways, the CDI has contributed to the study of child language. At the same time, we confess to some unease about the uncritical application of the CDI in some quarters, especially when it has served to replace rather than supplement direct observation and collection of laboratory data on language skills. Parent report (including the CDI) has many strengths, but like any measure, it also has limitations. We strongly encourage further research evaluating the relation between parent reports and other measures of development in language and in other domains (e.g., Smith, 1999), together with increased documentation of the arenas in which the CDI should be used with particular caution.

In sum, Feldman et al. have made an important contribution to evaluating both the nature of individual differences and "boundary conditions" for the use of parent reports, but some of their criticisms reflect facts about the nature of language development—especially its variability—that any research and clinical enterprise must confront.

ADDRESSES AND AFFILIATIONS

Corresponding author: Larry Fenson, Developmental Psychology Laboratory, San Diego State University, San Diego, CA 92182; e-mail: fenson@crl.ucsd.edu. Elizabeth Bates is at the University of California, San Diego; Philip Dale and Judith Goodman are at the University of Missouri, Columbia, MO; J. Steven Reznick is at the University of North Carolina, Chapel Hill, NC; and Donna Thal is at San Diego State University, San Diego, CA.

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 levels. *Applied Psycholinguistics*, *19*, 209–223.

Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. In G. Altmann (Ed.), Special issue on the lexicon, *Language and Cognitive Processes, 12*(5/6), 507–586.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P. S., Reznick, J. S., Reilly, J. S., & Hartung, J. P. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*, 85–123.

Caselli, M. C., Bates E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*, 159–199.

Caselli, M. C., Casadio, P., & Bates E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26*, 69–111.

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

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

Healthy Steps for Young Children Program. (1999). *Assessing the impact of pediatric-based developmental services on infants, families and clinicians: Challenge to evaluating the Healthy Steps Program.* Women's and Children's Health Policy Center, Johns Hopkins University School of Public Health, Baltimore, MD.

Jahn-Samilo, J., Goodman, J. C., Bates, E., Appelbaum, M. & Sweet, M. (1999). *Vocabulary learning in children from 8 to 30 months of age: A comparison of parental report and laboratory measures.* Manuscript in preparation.

Labov, W. (1970). The logic of nonstandard English. In J. E. Alatis (Ed.), *20th Annual Round Table* (pp. 1–43). Washington, DC: Georgetown University Press.

McCall, R. B., Hogarty, P. S., & Hurlburt, N. (1972). Transitions in sensori-motor development and the prediction of childhood IQ. *American Psychologist, 27*, 728–748.

NICHD Early Childcare Research Network. (1997, April). *Mother–child interactions and cognitive outcomes associated with early child care: Results from the NICHD Study.* Paper presented at biennial meeting of Society for Research in Child Development, Washington, DC.

Olswang, L., Rodriguez, B., & Timler, G. (1998). Recommending intervention for toddlers with specific language learning difficulties: We may not have all the answers, but we know a lot. *American Journal of Speech-Language Pathology, 7*, 23–32.

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5*, 5–21.

Paul, R. (1997). Understanding language delay: A reply to van Kleek, Gillam, and Davis. *American Journal of Speech-Language Pathology, 6*, 40–49.

Reznick, J. S. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics, 11*, 145–166.

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: Erlbaum.

Thal, D., Bates, E., Goodman, J., & Jahn-Samilo, J. (1997). Continuity of language abilities: An exploratory study of late- and early-talking toddlers. *Developmental Neuropsychology, 13*(3), 239–274.

Thal, D. & Katich, J. (1996). Predicaments in early identification of specific language impairment: Does the early bird always catch the worm? In K. Cole, P. Dale, & D. Thal (Eds.), Assessment of communication and language (pp. 1–28). Baltimore, MD: Brookes.