CONTINUITY OF LANGUAGE ABILITIES: AN EXPLORATORY STUDY OF LATE– AND EARLY–TALKING TODDLERS

Donna J. Thal
San Diego State University
University of California, San Diego
Children's Hospital Research Center, San Diego

Elizabeth Bates
University of California, San Diego
Children's Hospital Research Center, San Diego

Judith Goodman
University of California, San Diego
University of Missouri

Jennifer Jahn-Samilo
University of California, San Diego

Developmental Neuropsychology, **1997**, ***13*****(3)**, **239–273**

This work was supported in part by PHS 5 P50 DC01289-02, DC00089, DC00482, and the John D. and catherine T. MacArthur Foundation.

Abstract

Three exploratory studies were carried out to determine if there was continuity in the development of language in young children at the upper and lower extremes of the normal continuum, and if it was possible to use variables from an early assessment to predict their language status at a later date. Studies 1 and 2 examined continuity over 6-month period (from approximately 20 to 26 and 13 to 20 months of age, respectively); Study 3 examined continuity from 8 to 30 months of age. Results provided solid evidence for continuity at the group level but no evidence of ability to predict outcome for individual children using the vocabulary production, vocabulary comprehension, and gesture production variables included in this study.

Many parents wonder whether their child is normal. They wonder whether he or she is abnormally slow to develop, or so precocious that early celebration is warranted. During their child's first year of life, most parents worry about issues like sleeping, eating, and attainment of motor milestones (especially crawling and walking). During the second year, the focus switches to communication and language. This is true for physicians and other health care professionals as well, because disorders of higher cognitive functions in toddlers and preschoolers are often manifested as language disorders (Tuchman, Rapin, & Shinnar, 1991).

Until recently, developmental psychologists and psycholinguists have had little to say in response to these concerns, beyond some relatively limited norms established early in the twentieth century (Gesell, 1925; McCarthy, 1954). This is true in part because modern developmental research has focused on the universal characteristics of language learning, based upon an idealized "modal child" who acquires his or her native language in a standard sequence, on a standard schedule, with a single set of mechanisms (i.e., the Language Acquisition Device—for discussions of this point, see Fenson et al., 1994; Hardy-Brown, 1983; Plomin, 1989). In the absence of more systematic information about normal variability, a folk wisdom emerged among many health professionals, including an exaggerated belief in gender differences ("Boys are usually late..."), coupled with an optimistic view of the outcomes associated with language delays before 3–4 years of age ("Don't worry, he'll catch up....").

Fortunately, variability in early language development has become an active topic of research in recent years, rendering the notion of a "modal child" less tenable. Large variations in vocabulary size and rate of growth have been reported by a number of researchers (Bates, Bretherton, &Snyder, 1988; Fenson et al., 1994; Goldfield, 1987; Hampson & Nelson, 1993; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). In most cases, these variations are too large to explain with traditional biological variables (e.g., gender, rate of sensori-motor development) or social variables (e.g., birth order, social class, maternal style or quality of mother-child attachment). For example, Fenson et al. (1994) have looked at variability in rate of

development among more than 1800 infants between 8 and 30 months of age, the largest study of early language development to date. Among other things, they report that females are indeed ahead of males on most measures of language and communication. However, this difference is far smaller than one might expect based on the conventional wisdom. Gender differences account for less than 2% of the variance on any single measure of language and communication, with girls averaging about 1 month ahead of boys across the period from 8 to 30 months. Rate is not the only type of variability in language development. Qualitative differences in "style" of development have also been observed, leading to the suggestion that there may be qualitative differences among normal children in the mechanisms used to acquire language (Bates et al., 1988; Bates, Dale, & Thal, 1995; Plunkett, 1993; Shore, 1995; Vihman & Miller, 1988).

The problems that such variability presents for theories of normal language development have been discussed at length by Fenson et al. (1994). This variability also poses significant problems for professionals charged with determining whether a child who is under three or four years of age is exceptional, either precocious or delayed. If there are no clear criteria for identifying what is "normal", then it is especially difficult to be certain that a child is truly delayed or precocious. Given that professionals who serve young children are now charged by Public Law 99-457 (Federal Register, 1986) to evaluate communicative abilities and provide appropriate treatment for children from birth to three years of age, this creates a serious dilemma.

Although tremendous variability in rate and style of language acquisition has been documented, it remains unclear whether individual differences during the early period of language acquisition have any consequences for language abilities at a later point in time. We need to assess whether children who are delayed or accelerated in their language acquisition at one point in development remain so at a later time and whether variability of this sort has long-term consequences for the child's ability to function successfully. There is still no answer to the second question, but there is evidence for short-term stability in rate of development in early language and communication, across the normal range. As part of their large norming study with

more than 1800 infants, Fenson et al. obtained six-month follow-up data from approximately a third of the sample, including children whose parents filled out the 8- to 16-month MacArthur Communicative Development Inventory (CDI) Infant Scale at two time points (Infant-Infant sample), children whose parents filled out the 16- to 30-month Toddler Scale at 2 time points (Toddler-Toddler sample), and another group whose parents filled out the Infant Scale at Time 1 and the Toddler Scale at Time 2 (Infant-Toddler sample). Results indicate clearly that the individual differences in language comprehension and production observed in the larger cross-sectional study are relatively stable across a six-month period. Because the Fenson et al study uses the same methodology as the study of extreme groups presented below, it will be useful to consider their longitudinal results in a bit more detail.

In the first study, Fenson et al used the CDI Infant Scale at both data points. The CDI: Words and Gestures (Infant form) is normed for vocabulary comprehension, vocabulary production, and gesture production in children from eight to 16 months of age. The Infant form has two parts: One samples language and the other samples gestures. The major portion of Part I consists of a 396-item vocabulary checklist organized into 19 semantic categories. Ten of those categories are composed of nouns, ("animals," "vehicles," "toys," "food and drink," "clothing," "body parts," "furniture and rooms," "small household items," "outside things and places to go," and "people") and the remaining sections include "sound effects and animal sounds," "games and routines," "action words" (i.e., verbs), "words about time," "descriptive words" (i.e., adjectives), "pronouns," "question words," "prepositions and locations," and "quantifiers." Parents are asked to mark the appropriate space if their child comprehends or comprehends and produces each word, and to leave the space blank if they don't comprehend or produce it yet. Part II samples actions and gestures, examples of early communicative and representational skills that are not dependent on verbal expression. The gestures are divided into five subscales as follows: First communicative gestures (those that signal the onset of intentional communication like "giving," "showing," and "pointing"), games and routines (early social interactive gestures, like "pat-a-cake"), actions with objects (e.g., "sniffing" a flower, pretend "stir,"), pretending to be a parent

(among the first types of symbolic actions, like brushing a doll's hair, pretend "feeding" a doll), and imitating other adult actions (pretend "driving," pretend "sweeping" with a broom). These are behaviors that have been described as predecessors and/or correlates of early language (Bates et al., 1979) and as potential predictors of risk for language delay (Thal et al., 1991; Thal & Tobias, 1994).

The Fenson et al (1994) Infant-Infant sample demonstrated continuity in vocabulary comprehension, vocabulary production, and gesture production, in 62 children with a mean age of 9.61 months (SD = 0.72) at Time 1, and a mean of 16.34 months (SD = 0.81) at Time 2. In order to identify the independent stability of each of these three dependent variables over time, multiple regression analyses were conducted in which the relationship between that variable at Time 1 and its equivalent at Time 2 was tested after variance due to age, gender, birth order and SES was removed. In the longitudinal analysis of word comprehension, Time 1 comprehension scores accounted for 21.8% of the variance in Time 2 comprehension after other factors were controlled (all the Time 1 variables together accounted for 32.9% of the variance at Time 2). In a similar regression analysis using gesture production as the dependent variable, Time 1 gesture accounted for an additional 16.7% of the variance in Time 2 gesture (out of a total of 36.7% for all predictors together). In a third regression analysis, Time 1 word production accounted for an additional 8.1% of the variance in Time 2 production (out of a total of only 12% for all variables together on the final step). All of these unique contributions were highly reliable ($p < .001$). As Fenson et al. point out, the weaker longitudinal stability for vocabulary production in the infant data set probably reflects the fact that many of the subjects in this sub-sample did not produce any language at the first data point (i.e., a statistical floor effect).

The second study used the CDI: Words and Sentences (Toddler form) at the second data point. It is normed for vocabulary production, utterance length and grammatical complexity in children from 16 to 30 months of age. Part I of the Toddler form contains a 680-word vocabulary production checklist organized into 22 semantic categories. In this inventory parents

are only asked about the words a child produces. Part II contains five sections designed to assess morphological and syntactic development. Two of those, utterance length and grammatical complexity, were used in this study. The utterance length section provides an upper-limit measure. Parents are asked to provide examples of the three longest utterances that they have heard their child produce and the mean number of morphemes is calculated. We will refer to this measure as M3L in order to distinguish it from the more familiar mean length of utterance (MLU). The grammatical complexity section contains 37 sentence pairs in which one represents typical immature grammatical structures and the other more mature forms. Three levels of grammatical ability are represented within the 37 pairs: bound morphemes (e.g., "Daddy car." vs. "Daddy's car."), functor words (e.g., "Doggie table.", vs. "Doggie on table.") and early emerging complex sentence forms (e.g., "I sing song.", vs. "I sing song for you."). Parents are asked to choose the exemplar in each of the pairs that reflects the child's current level of speech.

In the Fenson et al. Infant-Toddler sample, mean age at Time 1 was 13.45 months (SD=1.71), while mean age at Time 2 was 20.15 months (SD=1.86). In this particular data set, word production was the only measure with equivalents at both time points. Using the same regression strategy described above, Time 1 vocabulary production alone accounted for a highly reliable 24.5% of the variance in Time 2 production after other factors were controlled (bringing the total for all Time 1 predictors to 54.1%). In other words, parent reports of vocabulary production have substantial predictive value from 13 to 20 months of age.

Finally, the Fenson et al. Toddler-Toddler sample yielded more evidence for short-term stability of individual differences in early language development. In this data set, the mean age of children at Time 1 was 20.26 months (SD=2.40), and the mean age at Time 2 was 26.88 months (SD=0.62). Regression analyses comparable to those just described for the Infant-Infant and Infant-Toddler samples were conducted for two key measures; total vocabulary and grammatical complexity. Vocabulary production at Time 1 added 32.3% to the variance in Time 2 vocabulary, out of a total of 60.5% from all the predictors together on the final step.

Grammatical complexity at Time 1 added 19.2% to complexity at Time 2, out of a total 47.8% of the variance from all predictors. All predictions were highly reliable ($p < .001$).

To summarize, although there are large individual differences in the timing of early language milestones, the Fenson et al. longitudinal data suggest that these differences are relatively stable across a six-month time period. Is this also true for children at the extreme end of the distribution? Do groups of children who are significantly delayed or very precocious in early language and communication maintain their standing across time? Perhaps more important from an applied perspective, can we use variables from an early assessment to predict whether individual children will remain "late talkers" or "early talkers" at a later point in time? These three questions will be explored in the studies reported below.

We have used the CDIs described above to examine the value of parental reports of vocabulary and gesture production for predicting precocity or delay. Experiments 1 and 2 use the same data base and group design analyzed by Fenson et al. (1994), but focus on children in the upper and lower ten percent of the normal distribution rather than mean or modal developmental trends. These are exploratory studies, designed to use data already gathered to help formulate hypotheses for further work. Our emphasis on late talkers is easy to justify on clinical grounds. However, early talkers are equally interesting. Their development is relevant to theoretical questions such as whether late and early status are equally stable, and whether the same factors that predict late status are the ones that predict continued precocity. They also provide useful information for the interpretation of language delay (e.g., as a control for psychometric factors like regression toward the mean). We chose a working definition of late talkers as children who fall at or below the tenth percentile and of early talkers as children who fall at or above the 90th percentile because these are the cut-off points adopted in several previous studies of children with atypical profiles (Robinson, Dale, & Landesman, 1990; Thal & Bates, 1988; Thal & Tobias, 1994). Experiment 3 reports on developmental trajectories of individual children from a new longitudinal study of 34 children who were followed monthly

across the period from 8 to 30 months of age. Here we will focus on children who are late or early talkers at any point across the period from 18 to 24 months (the time window used by Thal and her colleagues to identify late and early talkers in previous studies). We will look at where these children end up at 30 months of age (the end of the study), and how they fared on the infant measures in the months prior to their identification as late or early talkers. These three studies will provide further evidence for stability of individual differences across the period from first words to grammar, but they will also show that our ability to predict outcomes for individual children is still quite limited.

Experiment 1

Method

Subjects

The subsample from the complete cross-sectional norming study of the CDIs analyzed in Experiment 1 included 185 children whose parents completed the CDI Toddler form twice, once when they were between 16 and 25 months of age (mean = 19.73 months, SD = 2.08) and again when they were between 22 and 31 months of age (mean = 26.39 months, SD = 2.12 months). The average time between completing the two toddler CDIs was 6.65 months (SD = 0.34 months, range = 6.0-8.47). The samples were collected at three sites: New Haven, Connecticut; Seattle, Washington; San Diego, California. Together the subsample contained 24.3% of the original CDI: Toddler norming sample.

Children were excluded from the study if they were six or more weeks premature, had a genetic disorder such as Down or Williams syndrome, had extended illnesses or were subjected to serious medical or surgical procedures, or had other serious medical problems. Repeated ear infections were not an exclusionary criterion as they had not been for the normative sample. In the original sample 4.3% of the children were reported to have had repeated ear infections.

The ethnic and educational characteristics of the whole sample are described in detail by Fenson et al. (1993, 1994). These are representative of the subsample used in this study. In comparison to the 1990 census figures (Bureau of the Census, 1991) the educational level of the parents in this study is above the national average. The ethnic diversity was not significantly different from the 1991 Bureau of the Census figures.

The sample was restricted to children for whom English is the primary language. However, children with exposure to a second language were not eliminated from the normative study because bilingual environments are common for many children in the United States. A total of 12.2% of the families in the normative sample reported that their children had some exposure to a second language.

Children were placed into late-talker or early-talker categories based on percentile norms from Fenson et al. (1994). Although (as we noted earlier) gender differences are relatively small in this sample, they were significant and Fenson et al. recommend assignment of percentile scores based on separate norms for boys and girls. These recommendations were followed in the present study, so that late-talker status always means that the child obtained a vocabulary score at or below the 10th percentile for his or her age and gender; similarly, early-talker status always means that the child obtained a vocabulary score at or above the 90th percentile for his or her age and gender.

Procedure

As noted above, the CDI Toddler form was used for Experiment 1. Because of the exploratory nature of this work we examined stability of extreme scores in this sample with a number of statistical analyses, some of which are mathematically redundant. However, each analysis has an advantage for a different kind of question. The kinds of questions and the statistical analyses used to examine them follow:

(1) Breakdown of late and early talkers. Based on cross-sectional norms for the CDI, we determined how many longitudinal subjects fell at or below the tenth percentile (i.e., late talkers) and how many fell at or above the 90th percentile (i.e., early talkers), at both Time 1 and Time 2. Binomial tests were used to determine whether late or early talkers were under- or overrepresented in the longitudinal data set, compared with the cross-sectional norms.

(2) Subject and family variables. One-way analyses of variance were conducted to compare late talkers, early talkers and children who fell between the two extremes for age in days at both Time 1 and Time 2, birth order, socioeconomic status, maternal and paternal education, maternal and paternal vocation. Chi-square tests were used to compare the same three groups on four categorical variables (gender, ethnicity, presence/absence of limited exposure to a second language, presence/absence of mild medical complications).

(3) Stability of late/early status. Binomial tests were used to determine if children who were delayed (or early) at the first data point were likely to be delayed (or early) at the second data point. After determining whether and to what extent children maintained their extreme group status, we checked to see if those who stayed late (or early) differed from those who changed status on the above subject and family variables.

(4) Factors that predict Time 2 status. We then asked whether any of the Time 1 language measures predicted late- or early-talker status at Time 2, first across the sample as a whole and then only for those children who fell into extreme groups at Time 1. Analysis of variance comparing late talkers and early talkers with mid-range children at Time 2 on the major language variables at Time 1 were carried out, followed by post-hoc tests to determine which pairwise comparisons were significant when the ANOVA was significant. Regression analyses were used to explore whether

any of the major language variables at Time 1 contributed unique variance to Time 2 status (for the group as a whole, and for those who fell in the extremes at Time 1).

(5) Predicting individual cases. Finally, we conducted discriminant analyses using the same Time 1 predictors to classify children into Late or Early status at Time 2. Whereas analyses 1 - 4 asked about continuity at the group level, these discriminant analyses tell us how successful we are in predicting the short-term future for individual children.

Results

(1) Breakdown of Late and Early Talkers. At Time 1, when the children averaged just under 20 months of age, 24 of the 185 subjects (12.9%) fell at or below the 10th percentile according to the MacArthur cross-sectional norms. A binomial test indicated that this was no more than we would expect if the longitudinal sample had been drawn randomly from the cross-sectional population. Similarly, 25 of the 185 children (13.5%) fell at or above the 90th percentile, again no more than we would expect by chance on a binomial test. At Time 2, 17 of these 185 children (9.1%) qualified as late talkers, slightly below 10% but still no more than we would expect by chance on a binomial test. By contrast, 30 out of 185 children (16.2%) qualified as early talkers at Time 2, which is significantly more than we would expect by chance ($p < .007$) if children were drawn randomly from the cross-sectional norms. There are at least two possible explanations for the overrepresentation of early talkers in this longitudinal sample at Time 2, and they are not mutually exclusive: (1) parents who are willing to participate in a longitudinal study tend to be more attuned and attentive to language, and such individuals are more likely to have precocious children; (2) parents of precocious children are particularly proud of their offspring, and more motivated to continue at the second data point.

(2) Subject and family variables. We used the percentile rankings described above to divide the children into three groups (late talkers, middle range talkers, and early talkers), one for

each of the two time points. One-way analyses of variance were then conducted on all the subject and family variables that form an interval or ordinal scale (Time 1 age in days, Time 2 age in days, birth order, SES, maternal education, paternal education, maternal occupation, paternal occupation). Chi-square tests were conducted to compare the three groups on categorical variables, including gender, presence/absence of mild medical complications, presence/absence of exposure to a second language, and ethnicity (Caucasian, African-American, Hispanic, Asian, other).

The three groups identified at Time 1 differed significantly on three measures: Time 1 age ($F(2,182) = 3.42$, $p < .04$), Time 2 age ($F(2,182) = 3.11$, $p < .05$), and SES ($F(2,182) = 3.43$, $p < .04$). Post-hoc Tukey tests ($p < .05$) indicated that early talkers were older than children in the middle range at both time points. Late talkers were younger than children in the middle range at Time 1, but were not significantly different at Time 2 (Time 1: Late = 19.38, Mid = 19.62, Early = 20.71; Time 2: Late = 26.08, Mid = 26.26, Early = 27.35). It is possible that some of the early talkers crept to the 90th percentile or above because they were closer to the border for the next age-in-month bracket (where they might have received somewhat lower percentile scores). A similar (but reverse) phenomenon may have produced the effect for late talkers. The Tukey post hoc test for SES was not significant. Early talkers tended to come from families slightly higher and late talkers from families with slightly lower SES (Late = 43.1, Mid = 44.0, Early = 50.9; Hollingshead, 1965).

The three groups identified at Time 2 also differed on age, including their age at the earlier time period ($F(2,182) = 3.85$, $p < .023$, and their age at Time 2 ($F(2,182) = 3.20$, $p < .05$. Again, post-hoc Tukey tests ($p < .05$) indicated that early talkers were older than children in the middle range at both time points. However, late talkers were also significantly older than children in the middle range at both time points (Time 1: Late = 20.24, Mid = 19.49, Early = 20.55; Time 2: Late = 26.82, Mid = 26.17, Early = 27.26). Once again, some of the variance in the early talker group may be an artifact of age brackets, as noted for the previous set of

analyses. This cannot explain the late talker findings, however. None of the other effects reached significance in the Time 2 analyses, including the SES variable. This suggests that the weak SES advantage for early talkers at Time 1 is probably a transient phenomenon, or of little significance in both the statistical and the practical sense of the term.

(3) Stability of late- and early-talking status. For children who were early or late at Time 1, binomial tests were conducted to determine whether more of them retained that status at Time 2 than we would expect by chance (i.e., more than we would expect if there were no continuity of status across the six-month period). A complementary set of analyses was conducted to determine if children who were early or late at Time 2 had the same status at Time 1. Of the 24 children who were late talkers at Time 1, 12 (50%) retained that status at Time 2, far more than we would expect by chance ($p < .00001$). Of the 25 children who were early talkers at Time 1, 15 (60%) were also early talkers at Time 2 ($p < .0001$). Hence it is indeed the case that children who fall in the extremes around 20 months of age have a higher-than-chance probability of retaining that status six months later. It is also true, however, that 50% of the late talkers and 40% of the early talkers regressed back into the mid-range. Looking back in the opposite direction, we find that 12 out of 17 or 70.5% of children who were late talkers at Time 2 came from the same extreme group at Time 1 ($p < .00001$). Of the children who were early talkers at Time 2, 15 out of 30 or 50% were also early talkers at Time 1 ($p < .00001$). It is clear that there is some stability in language status across this six-month period, although between 30 - 60% of the children migrate in or out of the extremes across this period of rapid development.

Can we distinguish between late talkers who "stay late" and those who move into the mid-range, or between early talkers who "stay early" and those who move back toward the mean? In a first attempt to answer this critical diagnostic question, we conducted analyses comparing late talkers who stayed late with those who did not, and early talkers who stayed early with those who did not, on all the subject and family variables reviewed in the previous section. T-tests were used for the continuous or interval variables (age, birth order, various measures of

family status) and chi-square tests were used for categorical variables (gender, ethnicity, presence/absence of mild medical problems or second-language exposure).

For Time 1 late talkers, we found no significant differences on any of the subject or family variables between children who stayed late and children who moved into the middle, with two exceptions. First, children who stayed late tended to have more mild medical complications than children who moved into the middle (i.e., 4 of the 12 children who stayed late had some kind of reported complication, compared with none of the 12 children who regressed toward the mean, chi-square likelihood ratio, $p < .02$). Second, there was a tendency for children who stayed late to be slightly older at Time 1 than children who regressed toward the mean (Stay Late, Time 1 mean age = 20.23; Regress, Time 1 mean age = 18.52, $t = -1.79$, $p < .09$). This latter finding may reflect outcome for children who were having more serious difficulty. That is, the older late talkers had a longer time to regress toward the mean (into the middle range) but had not done so. Alternatively, it may be an artifact of assigning percentile scores based on age-in-month brackets. That is, some of the children may have been classified as late talkers at Time 1 because they were particularly young for their age bracket, and would not have fallen below the 10th percentile if they were just a few days younger.

For Time 1 early talkers, we found only one reliable difference between those who stayed early and those who regressed toward the mean. Specifically, children who remained precocious across this six-month period had mothers with significantly lower rankings on an occupational scale ($t = 3.69$, $p < .001$; Stay Hi = 4.67; Regress = 7.5).

(4) <u>Predicting Time 2 status from Time 1 language measures.</u> In this section, we want to consider whether children who were late or early at Time 2 differed on specific language measures at Time 1. Five variables were used 1) ability to combine words (scored as never, sometimes, or often), 2) sentence complexity, 3) mean length (in morphemes) of the three longest utterances, percentile score (based on vocabulary size rather than age) for the proportion of words in their vocabulary which were grammatical function words (an attempt to assess

"closed-class style"—Bates et al, 1994; Shore, 1995), and 5) total number of words (i.e., the measure that was used to assign late- and early-talker status at each time point). One-way ANOVAs were conducted comparing the late, early, and middle range groups at Time 2 on each of these Time 1 language variables, followed by Tukey tests ($p < .05$) to identify significant pairwise differences. Results are summarized in Table 1. We then carried out regression analyses to determine whether any of the individual variables contributed unique variance to late or early status when other variables were controlled.

There was a significant main effect for all of the language variables except closed-class percentile score (see Table 1). For children who were late talkers at Time 2, post-hoc tests demonstrated significant differences for all Time 1 language measures for which there was a main effect in the ANOVA (Table 1). Similarly, for children who were early talkers at Time 2, significant differences were also observed for the same Time 1 language measures (Table 1). Hence children who are late or early talkers around 26 months were (as a group) already well behind or ahead of their peers, respectively, at 20 months in vocabulary and grammar.

- - insert Table 1 about here - -

Our next question pertains to the unique contribution of individual Time 1 measures to extreme group status at Time 2. Regression analyses were used because they allow us to evaluate the separate contributions that each of the correlated variables make. Two separate regression analyses were conducted, one for late talker status and the other for early talker status at Time 2. In both of the analyses we evaluated the joint and unique variance of the following eight predictor variables: Time 1 age, Time 2 age, SES, and five Time 1 language variables including total number of words produced, level of word combinations, sentence complexity, mean length of the three longest utterances, and closed-class scores. Unique variance was determined by running eight separate step-wise regressions with the same variables but changing the variable that was added last for each analysis.

In the analysis for Time 2 late talkers, the eight predictors together account for 14.64% of the variance, a highly reliable prediction ($p < .001$). However, only one of the eight predictors added significant unique variance when it was entered into the equation last. Specifically, total vocabulary increased the prediction by an additional 3.66% ($p < .01$). This reflects a partial correlation of –.20, indicating that late talkers had smaller vocabularies than other children in the sample even after all the age, SES and grammar measures were controlled on the first step. However, age, SES and grammatical abilities did not add to the prediction when vocabulary size is controlled.

In the analysis for Time 2 early talkers, the eight predictors together accounted for a robust 30.9% of the variance ($p < .001$). Note that this is more than twice the variance accounted for when the same variables are used to predict late-talker status, suggesting that it is easier to predict who will stay ahead than who will stay behind. Once again, however, the only variable that made a significant unique contribution on the final step was total vocabulary size at Time 1 (7.3%, $p < .001$), with a partial correlation of +.31 indicating that early talkers at Time 2 were ahead of their peers in vocabulary size at Time 1 even after all the age, SES and grammatical predictors were controlled. Grammatical measures did not add significantly to the prediction, although there was a trend in that direction (with Time 1 utterance length contributing 1.2% on the last step, $p < .10$).

This brings us to our most important question: Can we use Time 1 language variables to predict whether children will maintain their status across a six-month period? To answer this question, we began with t-tests on the five language variables comparing Time 1 late talkers who stayed late with those who moved toward the mean, and Time 1 early talkers who stayed early with those who moved into the middle range. Results are summarized in Table 2.

In the late-talker analyses, none of the t-tests reached significance. There was a trend in the t-test on Time 1 production, but it was in the opposite direction from what we might predict if persistence and severity of delay were related. Specifically, children who stayed late had

slightly larger vocabularies at Time 1 (mean = 25.08), compared with children who caught up later (mean = 11.91, $t = -2.02$, $p < .065$).

- - insert Table 2 about here - -

In the early-talker analyses, there was no significant difference between those who stay early and those who regress toward the mean on vocabulary size, combining or closed-class proportion scores at Time 1. However, there was a significant difference on grammatical complexity (although the difference just missed significance when the Bonferroni correction was applied), and the difference for utterance length approached significance in the direction that we would predict if persistence and precocity were related. Specifically, children who stayed precocious across this six-month period had higher complexity scores at Time 1 ($t = -2.25$, $p < .04$; Stay Early, mean = 13.87; Regress, mean = 6.90), and they were reported to produce longer utterances ($t = -1.79$, $p < .09$; Stay Early, mean = 5.69 morphemes; Regress, mean = 6.55 morphemes).

Finally, we conducted regression analyses using the same eight predictors described above, to determine which factors jointly or uniquely predicted staying late and/or staying early.

In the analysis of late talkers, the eight predictors together accounted for 32.2% of the variance. However, the prediction as a whole did not reach significance. Furthermore, none of the individual variables made a unique prediction when it was entered on the final step. We repeated the regression using only two language measures, total vocabulary and grammatical complexity. Taken together, the two measures accounted for 21% of the variance in staying late, but this prediction also failed to reach significance ($p < .09$). When it was entered on the last step, total vocabulary size did make a significant contribution (16.7%, $p < .05$), but the partial correlation was positive (+.42), indicating that children who stayed late actually had slightly larger vocabularies at Time 1, in line with the report above using simple t-tests. We also conducted a regression using nothing but age and social class as predictors, and captured a

nonsignificant 29.2% of the variance in staying late ($p < .07$). Social class and Time 2 age did not make significant contributions on the last step, but Time 1 age did increase the prediction by a reliable 17.6% ($p < .04$). The partial correlation was +.45, which means that children who stayed late were somewhat older at Time 1. As we have already noted, this may reflect greater risk for continued language delay with increasing age as has been suggested by Rescorla & Schwartz (1990) or it may be nothing more than an artifact of age bracketing in the assignment of percentile scores. That is, some children were given lower scores because they were only a few days from the age boundary and those children are more likely to "catch up" at the next time point.

Analogous regressions were conducted on staying early. When all eight predictors are used together, the total variance accounted for was 36.4%, a figure that is not significant. None of the predictors made a reliable contribution on the final step. We also repeated the regression using only two language predictors, total vocabulary and grammatical complexity. The total prediction was only 15.4% ($p < .16$), and neither of these language variables made a unique contribution when the other was controlled. Hence there is no evidence to suggest that grammar makes a reliable contribution to persistence of early-talker status, above and beyond the variance shared by grammar and vocabulary around 20 months of age (see also Bates, Bretherton, & Snyder, 1988). Finally, we conducted a regression using age and SES as our only predictors. This equation accounted for a non-significant 14.5% percent of the variance in staying early, with no significant contribution from SES or age. There was a trend for SES, which added 12.5% to the prediction after Time 1 and Time 2 age were controlled ($p < .10$), but the partial correlation was negative, indicating that the families of children who stayed early tended to have slightly lower SES scores.

To summarize so far, we have evidence for continuity at the group level in late-talker and early-talker status from 20 - 26 months of age. However, we are not yet in a position to predict which late talkers will stay delayed, nor can we predict which early talkers will stay precocious.

We will attempt to partially resolve this problem, using discriminant analysis to determine whether the eight variables described above form a factor that distinguishes between late talkers and middle level children, and between early talkers and middle level children. By determining whether such a factor can predict group membership we will get a clearer picture of our ability to make short-term predictions for individual children with the subject, family and language variables that we have at hand (Munro & Page, 1993).

(5) <u>Discriminant analysis and classification.</u> We began by attempting to predict late-talker status at Time 2, from a factor composed of the same eight variables employed for the regression analyses described above. A chi-square analysis indicated that the prediction of late-talker status at Time 2 for the group as a whole was statistically reliable (chi square = 26.60, $p < .008$), as we would expect from the regression analyses. The critical contribution of discriminant analysis, however, comes from the classification tables because they provide information about the success of predicting outcomes for individual members of the group. Table 3 shows the classification data for prediction of Late Talker status at Time 2. The number of children actually classified as Late Talkers or middle level producers at Time 2 is listed on the left. The two columns on the right indicate the number (and percent) of those actually classified in either category that were predicted from our SES and language factor.

- - insert Table 3 about here - -

For the group as a whole, 74.7% of the cases were classified correctly. Of the children who actually qualified as late talkers at Time 2, 68.8% were correctly identified while 31.3% were misassigned to the normal range by the variables used in the analysis.. Of the children who should have been classified in the normal range, 75.3% were correctly assigned while 24.7% were misdiagnosed as late talkers. In short, we do not have good specificity or sensitivity for prediction of late-talker status with the variables we now have at hand, even though there is significant continuity over time at the group level.

In the early-talker analysis using the same factor, 79.89% of the cases were correctly classified for the group as a whole, significant by a chi-square statistic (chi square = 62.21, $p < .001$), and this is compatible with the regression analyses reported above. For those children who should have been identified as early talkers, 66.7% were correctly classified while 33.3% were misassigned to the normal range (see Table 4). For those children who should have been assigned to the normal range, 82.6% were correctly placed while 17.4% were incorrectly assigned to the early-talker group. Once again, even though we do have evidence for continuity at the group level, the sensitivity and specificity of our prediction is not very good.

- - insert Table 4 about here - -

Finally, we used the same combination of variables in discriminant analyses of children who were late or early at Time 1, to see how many cases of staying late or staying early we could predict. In the staying late analysis, 76.2% of the cases were correctly classified. However, the chi-square analysis evaluating this prediction was not reliable, analogous to our findings with regression (chi square = 5.83, n.s). The classification table for this analysis is presented in Table 5. Among the children who really did maintain late-talker status, 72.7% were classified correctly but 27.3% were misassigned to the normal range. Among the children who had moved into the normal range, 80% were correctly placed but 20% were classified as late talkers.

- - insert Table 5 about here - -

In the staying early analysis, 76% of all cases were correctly classified. However, in line with our regression results, a chi-square statistic evaluating this prediction failed to reach significance (chi square = 8.61, n.s.). Among the children who really did maintain early-talker status (Table 6), 66.7% were classified correctly but 33.3% were misassigned to the normal range. Among the children who moved out of early-talker status back toward the mean, 90% were correctly classified but 10% were assigned to the early-talker group.

- - insert Table 6 about here - -

We conclude that there is significant stability in late and early-talker status from 20 - 26 months at the group level. However, these predictions are not very sensitive at the level of individual children. Above all, we are not yet in a position to predict which late talkers will stay late or which early talkers will stay early across this period of development. We turn now to our findings for a sample of younger children in Experiment 2, children in an age range during which measures of language production may not be available for use as predictors.

Experiment 2

Method

Subjects

The subsample from the complete cross-sectional norming study of the CDI analyzed in Experiment 2 included 217 children whose parents completed the CDI Infant form when their children were between 10 and 16 months of age (mean = 13.45 months, SD = 1.71) and the CDI: Words and Sentences when they were between 16 and 25 months of age (mean = 20.15 months, SD = 1.86). The average time between completing the two CDIs was 6.71 months (SD = .068 months, range = 4.8-12.6 months).

Procedure

We used the data from the two CDIs to look backward in the Infant-Toddler sample to see what kinds of stability and predictive validity we obtain using infant measures around 13 months of age to forecast late- and early-talker status around 20 months.

The same analytic procedures used in Experiment 1 were followed for the Infant-Toddler data set, with one important exception. Because so many children produced little or no speech at the first data point (mean age = 13 months), there was no reliable way to assign late-talker status at Time 1, and hence no way to determine whether late-talker status is stable over time. Thus, we looked only at stability of early-talker status from 13 to 20 months of age.

Results

(1) Breakdown of Late and Early Talkers. At Time 1, when children average 13 months of age, 44 out of 217 children in the sample were at or above the 90th percentile (20%), exactly twice the number that we would expect by chance ($p < .00001$). At the second time point, when we can use the cross-sectional norms to identify both late and early talkers, a total of 30 out of 217 of the children (14%) qualified as late talkers, slightly (but not reliably) more than we would expect by chance on a binomial test ($p < .08$). In the same group, 23 out of 217 (11%) qualified as early talkers, no more than we would expect if children had been drawn randomly from the cross-sectional norms.

Comparing these results with the Toddler-Toddler sample, we may conclude that late talkers are not significantly overrepresented in either sample, at either time point. Early talkers are occasionally overrepresented, at Time 2 in the Toddler-Toddler sample (20 - 26 months) and at Time 1 in the Infant-Toddler sample (13 - 20 months). The sporadic nature of these findings is hard to reconcile with any single explanation, and it is possible that several different factors are at work.

(2) Subject and family variables. Similar to our procedure for the Toddler-Toddler sample, we constructed a single variable at Time 2 representing extreme group status. Because late-talker status could not be assigned at Time 1, analyses at that data point were restricted to t-tests and chi square comparing only early talkers with the rest of the sample. The full extreme group variable established for the Time 2 data point was used in one-way analyses of variance on interval and ordinal variables (birth order, SES, maternal and paternal education and occupation), and in chi-square analyses with categorical variables (gender, ethnicity, presence/absence of mild medication complications, presence/ absence of exposure to a second language).

In the analyses of groups identified at Time 2, the only significant main effect was that for age at Time 1 ($F(2,214) = 4.07$, $p < .018$). However, Tukey post hoc tests ($p < .05$) were not

significant. The main effect for age at Time 2 approached significance ($F(2,214) = 2.466$, $p < .087$). None of the other effects were reliable.

In the analyses of groups identified at Time 1 (which are restricted to early talkers compared with the rest of the sample), t-tests revealed significant effects of birth order ($t = 2.86$, $p < .005$; Early = 1.32, Not-Early = 1.63), SES ($t = 2.04$, $p < .05$; Early = 38.48, Not-Early = 43.23), and maternal vocation ($t = 2.24$, $p < .03$; Early = 3.11, Not-Early = 4.39). However, only the birth order effect remains significant after applying a Bonferroni correction. The birth order effect is in the direction that we might predict based on literature concerning language development in first vs. later borns. That is, more first borns than later borns are early talkers at 13 months of age. The SES effects, however, are in the opposite direction from what one might predict based on the literature concerning socioeconomic factors in language development; early talkers at 13 months came from families with lower SES.

(3) Stability of late- and early-talking status. Of the 44 children who qualified for early-talker status at Time 1, binomial tests were conducted to see if more than expected by chance retained that status at Time 2. Sixteen cases (36%) were still at or above the 90th percentile at Time 2, far more than we would expect by chance ($p < .00001$). Conversely, of the 23 children who qualified as early talkers at Time 2, a total of 16 (70% of the sample) were also early talkers at Time 1 ($p < .00001$). In other words, linguistic precocity is often maintained between 13 and 20 months of age, and children who are well ahead at 20 months usually had a very early start.

What factors differentiated those children who stayed ahead from those who regressed toward the mean? T-tests and chi squares were used to compare children who stayed early with those who fell behind on all of our subject and family variables. Significant differences were found on a number of variables, including age at Time 1 ($t = -2.11$, $p < .043$; Stay Early = 14.2, Regress = 13.1), paternal occupation ($t = -2.21$, $p < .034$; Stay Early = 6.69, Regress = 5.17), presence of mild medical complications (somewhat more likely in children who stay early, $p <$

.02 by a chi-square likelihood ratio), and some exposure to a second language (also somewhat more likely in children who stayed early, $p < .04$ by a chi-square likelihood ratio). However, none of the t-tests remained significant after application of the Bonferroni correction. There were also trends favoring the children who remained precocious on other family variables, including maternal education ($t = –1.86$, $p < .08$; Stay Early = 15.37, Regress = 14.25) and paternal education ($t = –1.94$, $p < .06$; Stay Early = 15.81, Regress = 13.86). Because we have obtained such conflicting findings on the effects of social class, a two-tailed test was used in every case. We may conclude that demographic factors may have some influence on the likelihood that children will retain their precocity across the period from 13 - 20 months, but the effects, if present, are very small.

(4) <u>Predicting Time 2 status from Time 1 language measures.</u> Analogous to our procedure for the Toddler-Toddler sample, we used measures of language and gesture at Time 1 to compare children classified as late and early talkers at Time 2 with children in the middle range of vocabulary production. We performed ANOVAS followed by Tukey tests ($p < .05$) for pairwise comparisons where appropriate. Four Time 1 language measures were used: total word comprehension, total word production, total gesture production, and the percent of all words in the child's receptive vocabulary that she or he also produced. These are the four measures used by Bates et al. (this issue) to evaluate language and communication in infants with focal brain injury, and they are related to measures used in previous studies of development in late talkers (Thal & Bates, 1988; Thal et al., 1991; Thal & Tobias, 1994). Results are summarized in Table 7. After these simple group comparisons on individual measures, we carried out regression analyses using the same four language variables as predictors, together with Time 1 age, Time 2 age and SES, to determine whether any of the individual measures contribute unique variance to early- or late-talker status when the others are controlled.

There was a significant main effect for all of the language variables measured. For children who were late talkers at Time 2, post-hoc Tukey tests demonstrated significant differences for all

Time 1 language measures except total comprehension (for which there is a trend towards significance, $p < .07$). For children who were early talkers at Time 2, significant differences were also observed for the same Time 1 language measures (see Table 7). Hence children who were late or early talkers around 20 months were (as a group) already well behind or ahead of their peers, respectively, at 13 months in vocabulary comprehension, vocabulary production, gesture production, and the percent of their comprehension vocabulary that they produced.

- - insert Table 7 about here - -

We also used t-tests (with the Bonferroni correction) to determine whether children who stayed precocious from 13 to 20 months differed from those who regressed toward the mean at 20 months in their initial status on language and gesture variables. Results indicate that this was indeed the case, i.e., the children who stayed precocious were already significantly ahead at the first time point, on all four measures (see Table 8).

- - insert Table 8 about here - -

Turning now to the regression analyses, an equation with seven predictors (Time 1 age, Time 2 age, SES, and the Time 1 language and gesture variables) accounted together for 15.2% of the variance in late-talker status at 20 months of age ($p < .001$). Analyses were repeated so that there was an opportunity to enter each variable into the equation last. No unique variance was contributed by age, SES, total comprehension or total production. However, there were significant unique contributions on the last step by total gesture (6.0%, $p < .0002$, partial correlation = –.26) and by the percent of receptive vocabulary that is also produced (3.4%, $p < .004$, partial correlation = –.20). The negative partial correlations mean that children who became late talkers were lower at 13 months in gesture, and in the proportion of words they knew that they also said, even after the other five variables were controlled. As we will point out in more detail in the discussion, these findings are particularly interesting in view of independent evidence that infants with right-hemisphere lesions are particularly poor at production of

gestures, while infants with left-hemisphere lesions find it difficult to produce the words that they understand (Bates et al., this issue).

Rather different findings were obtained in the regressions predicting early-talker status at Time 2. The seven predictors together accounted for 39.8% of the variance in early-talker status at 20 months ($p < .001$). Overall, this is a much better prediction than we obtained for late-talker status. However, when each of the variables was entered into the equation last, the only unique contributions came from total production (2.1%, $p < .008$, partial correlation = +.18) and total comprehension (1.8%, $p < .02$, partial correlation = +.17). Although these contributions were quite small, it is interesting that they are very different from the results for late talkers.

Finally, we carried out a regression analysis only on children who were early talkers at 13 months, predicting whether children would stay precocious or drop back toward the mean. The total variance accounted for was a large and reliable 45.2% ($p < .002$). However, none of the seven variables made a unique contribution when it was entered on the last step.

(5) <u>Discriminant analysis and classification.</u> In these analyses, we used the same seven predictors adopted in the above regressions. The purpose of discriminant analysis was (once again) to see how many individual cases were correctly classified using these predictors.

In the analysis of late-talker status at 20 months, 77.88% of the cases were correctly classified, and the overall prediction was highly reliable by a chi-square statistic (chi square = 34.86, $p < .00001$). Of the children who really qualified as late talkers at 20 months, 90% were correctly classified while 10% were misassigned to the normal range. Of those children who were in the normal range at Time 2, 75.9% were correctly assigned but 24.1% were misclassified as late talkers.

- - insert Table 9 about here - -

In the corresponding analysis of early-talker status at 20 months, 91.7% of the cases were correctly classified, a highly reliable prediction (chi square = 107.20, $p < .00001$). However, among those children who did qualify for early-talker status at 20 months, only 69.6% were correctly classified while 30.4% were misassigned to the normal range. Among those children who fell below the 90th percentile at 20 months, 94.3% were correctly classified while 5.1% were misassigned to the early-talker group.

- - insert Table 10 about here - -

Finally, we conducted a discriminant analysis using these seven variables to predict which children stayed precocious and which ones fell behind. The total prediction was reliable, in line with the regression analyses above (chi square = 23.18, $p < .002$). Overall, 81.82% of the cases were correctly assigned. Of those early talkers who maintained their precocity from 13 to 20 months, 68.8% were correctly classified while 31.3% were misassigned. Among those children who moved out of early-talker status into the normal range, 89.3% were correctly classified while 10.7% were misassigned to the early-talker group.

- - insert Table 11 about here - -

Once again, we have clear evidence for continuity at the group level for late- and early-talker status, but the ability to predict the outcome for individual children is not very good. In most of these analyses, approximately two-thirds of the target cases are correctly identified, but many children are misclassified. Whether these predictions are valuable or not depends upon one's point of view: How much is at stake if even one case is missed? What harm is done if none are detected? We will return to this point in the final discussion (described as "The Pediatrician's Dilemma"). Meanwhile, we turn to fine-grained longitudinal data in Experiment 3 to get a clearer picture of the developmental trajectories observed in children who do maintain late- or early-talker status across this period of development, compared with some cases in which

children make dramatic changes in rate of language development, and in their rank relative to other children in the study.

Experiment 3

Method

Subjects

Twenty-eight children (17 males and 11 females) were included in Experiment 3. The majority of the children (64%) were either first-born or had no siblings in the same household. Thirty percent of the group were second-born and six percent were third- or fourth-born.

The ethnic and racial distribution of the group was representative of the United States population as a whole. Nine percent of the children were African-American, 27% were of various racial and ethnic combinations, and 64% were Caucasian. A range of educational and socioeconomic levels was also represented including 12% of the sample from single-parent households.

Children were recruited through a subject pool which contained names of parents who had responded to newspaper advertisements asking if they wished to participate in developmental studies and through personal referral. Children were not enrolled in the study if they were exposed to a second language on a regular basis. However, during the course of the study some of the children were exposed to a second language because of child care arrangements.

Procedure

Experiment 3 used both of the MacArthur Communicative Development Inventories. CDIs were mailed to parents and returned to the experimenter by mail each month. After the first form was filled out the items checked on the form returned each month were filled in on the

next form before it was sent. Thus, the data were cumulative. However, parents were permitted to remove checks from the CDIs if they had changed their minds about the status of that item since the last CDI was returned, and many did so. Children were analyzed descriptively with respect to their percentile on the national norms at each monthly data point. Percentiles for each child were calculated from the CDI Infant form when the children were between 8 and 16 months old and from the CDI Toddler form when they were between 17 and 30 months old. Our goal was to examine continuity of status for individual children in a manner that would allow us to compare these children to the existing literature on late talkers (Paul, 1991; Rescorla & Schwartz, 1990; Thal et al., 1991; Thal, & Tobias, 1994; Whitehurst, Fischell, Arnold, & Lonigan, 1992) and to the two experiments reported above. In order to do that we identified the children in the sample who fell below the tenth percentile or above the ninetieth percentile between 18 and 24 months of age. Not only is this the age range typically used in studies of children at the lower extreme of language development, Experiment 2 has also shown that it is difficult to classify a child as delayed in language at 13 months because so few children have large enough expressive vocabularies. By using 18 to 24 months as the age of identification of expressive delay in Experiment 3, we were able to gather additional evidence for stability over time (i.e., are late-talkers still late when they are older) and also to look back to see if any variables were predictive of subsequent delay. To do this we looked at language production at 30 months of age to examine whether status at 18 to 24 months was related to status at 30 months. We also looked back at word comprehension and gesture production for all children, and also at word production for early talkers, to see if there was continuity from the earlier months to the 18- to 24-month period.

Results

Of the 28 children who participated in this experiment, five had word production scores below the tenth percentile at some point between 18 and 24 months of age, and six had word production scores above the 90th percentile for their age at some point within that period.

Seventeen scored consistently between the 10th and 90th percentiles across the 18- to 24-month window. Four of the late talkers were boys and one was a girl; four of the early talkers were boys and two were girls (see Table 12).

Of the five late talkers, only one scored below the tenth percentile at 30 months of age; the remaining children were at the 12th, 24th, 26th, and 73rd percentile, respectively. Even if we regard the 12th percentile as "close enough" to qualify as a late talker, then we may conclude that only 40% of the sample retains late-talker status in this longitudinal study—a result that is even poorer than the discriminant analysis results of Experiment 1.

- - insert Table 12 about here - -

By contrast, all six of the early talkers scored above the 90th percentile at 30 months of age (note that the 30 month data point is missing for one of the early talkers, so the assessment is based on the 29 month data point for that child). This is a much greater percentage than seen in the discriminant analysis from Experiment 1 (see Table 12). Hence, early-talker status at 18 to 24 months did predict early-talker status at 30 months of age in this longitudinal sample.

Of the 17 children who fell within the average range from 18 to 24 months, none had moved into the late-talker category by the end of the study (note that the 30-month data point is missing for four of these children, so this assessment is based upon the 29-month point for these four cases—see Table 12). However, five of these 17 children did move into the early-talker category, according to cross-sectional norms on the MacArthur CDI. This finding reflects an "upward drift" for the longitudinal sample as a whole. The reason for this upward drift is unclear, and may reflect the kind of fine-grained longitudinal method adopted in this study (cf. Goodman & Bauman, 1995; Goodman & Jahn-Samilo, 1995).

Prediction of outcome using word production

Figure 1 illustrates progress in word production from 8 to 30 months for children who qualify as late, early, and average talkers at some point within the 18- to 24-month window. A series of one-tailed t-tests was conducted to determine when the early and late groups each differed significantly from the sample as a whole ($p < .05$). Results suggest that the six early talkers were not significantly different from the rest of the group prior to 18 months, although they retained a large and reliable advantage for the rest of the study. By contrast, late talkers were significantly below the other children in the study from 10 months of age (from 15 months of age using the Bonferroni correction), and their disadvantage as a group was maintained for the remainder of the study. Hence, we may conclude that there is some stability in expressive vocabulary for late talkers across the period from 15 (perhaps 10) to 30 months, even if the majority of individual late talkers did move above the tenth percentile (i.e., they lost their "official late-talker status"); stability of early-talker status is not evident until 18 months, but remains high after that point. Apparent differences between Experiment 2 and Experiment 3 in the onset and maintenance of late and early status may be related to frequency of sampling (i.e., there were only two data points in Experiment 2, compared with up to 23 data points for each child in Experiment 3).

- - insert Figure 1 about here - -

Prediction of outcome using word comprehension and/or gesture production

Can we predict who will become a late or early talker from progress in comprehension and/or gesture before 18 months of age? Table 13 summarizes percentile scores for each child in the experiment at 16 months of age, in word production, word comprehension, and gesture production.

- - insert Table 13 about here - -

As can be seen in the table, the five children who qualified as late talkers during the 18- to 24-month window were all within the bottom ten percent in comprehension at 16 months (i.e., the final data point on the Infant form). Four of these five children were also in the bottom ten percent for gesture production at 16 months, and three were within the bottom ten percent on word production. The fact that production was the weakest predictor is undoubtedly due to floor effects for production within this age range (i.e., there is far less variance in word production for the sample as a whole, which necessarily means that production percentiles are less reliable).

Figure 2 illustrates progress in comprehension for the respective late, early, and average groups, while Figure 3 illustrates progress in gesture production. A series of one-tailed t-tests were conducted to determine when the late talkers diverged from the average group in comprehension and gesture production (t-tests for word production are reported above). These tests showed (with the Bonferroni correction) that the late talkers were significantly below the other children at every age from 8 through 16 months, in both comprehension and gesture production (see Figures 2 and 3). Hence, this particular sample of late talkers appears to be slow across the board, in all modalities.

- - insert Figures 2 & 3 about here - -

Among the six children who qualified as early talkers between 18 and 24 months of age, Table 13 shows that only two were in the top ten percent for vocabulary production at 16 months, three were in the top ten percent for comprehension, and two were in the top ten percent for gesture production. At the same time, several of the "early-talkers-to-be" had rather low scores for comprehension and/or gesture production during this early phase of development. For example, Subject #19 had a 16-month percentile score of 5 for comprehension, 17 for gesture production and 25 for word production; Subject #16 had a 16-month percentile score of 13 for comprehension, 11 for gesture production, and 30 for word production; Subject #7 did have a high score of the 97th percentile for comprehension, but scored only at the 45th percentile for gesture production and the 18th percentile for word production. In fact, only two of the "early-

talkers-to-be" started off high across the board: #22 had 16-month percentiles of 100 in comprehension, 99 in gesture production and 100 in word production and # 38 had 97 in comprehension, 95 in gesture production and 98 in word production. Once again, we conducted a series of one-tailed t-tests, this time to determine when the early talkers diverged from average group on either comprehension or gesture production (results of t-tests for word production are reported above). Results of these t-tests were very clear, and compatible with the wide variation documented in Table 13. That is, early talkers were not significantly different from the rest of the sample (collapsing across late and average talkers) on any measure, at any point from eight to 16 months. This finding is illustrated in Figures 2 and 3, which show that late talkers were late from the beginning on both comprehension and gesture production, but early talkers were indistinguishable from children who later made average progress in word production.

Finally, we looked at the 8- to 16-month data for the children who made average progress in expressive vocabulary within the 18- to 24-month window (see Table 13 and Figures 2 and 3). In comprehension at 16 months of age, one of the "average-to-be" children scored in the top ten percent and one was in the bottom ten. In gesture production at 16 months, a total of five children scored within the top ten percent and one was in the bottom ten. In word production, two children started out in the top ten percent and none were in the bottom ten percent. We may conclude that there is a certain amount of regression toward the mean in this group, with some children moving out of the extremes and into the normal range. There also appears to be some asymmetry in direction of movement. For example, high gesture scores were not a good index of eventual precocity in language, but low gesture scores were common in children who qualified as late talkers. In the same vein, low comprehension scores occurred even in children who ended up as early talkers, but very low comprehension scores characterized all of the late talkers in this particular subsample.

Results of Experiment 3 differ in detail from those of Experiments 1 and 2, but one conclusion emerges clearly from all three: There is some continuity in language status across the

period from 10 to 30 months of age, but our ability to predict outcomes for individual children is limited.

Conclusions

The three experiments reported above provide clear support for stability of individual differences for children at either extreme of the normal distribution from as early as 13 months of age at the group level, a finding that supports earlier work by Fenson et al. (1993, 1994). In both Experiment 1 and Experiment 2, children who were late or early at one of the data points were more likely to be late or early, respectively, at the other data point. However, the predictions were somewhat stronger for precocious toddlers than for late talkers.

In Experiments 1 and 2, we also attempted to identify variables which predicted stability of language status in both groups using demographic and communication factors. For the 20- to 26-month sample, however, few of the demographic variables were significant. For the 13- to 20-month sample, only early-talker status was reliably predicted, but many of the predictors were not in the expected direction. These findings reflect the current state of disagreement regarding predictors of outcome in the literature on late talkers (Paul, 1991; Rescorla & Schwartz, 1990; Thal et al., 1991; Weismer, Murray-Branch, & Miller, 1994; Whitehurst et al., 1992). Because of these conflicting findings we have concluded that within the full range of middle class toddlers sampled in Experiments 1 and 2, demographic factors have more effect on the likelihood that children will remain precocious rather than delayed but, in either case, the effects are relatively small -- at least within the socioeconomic range that we have sampled here.

Communication factors provided reliable predictors for both studies. For children examined at 20 and 26 months of age, total number of words produced, presence of word combinations, sentence complexity, and the mean of the three longest utterances were significantly related to remaining late or early. As a group, children who were late on these variables at 20 months were more likely to be late at 26 months and vice versa. However,

regression analyses in which each of these factors was entered into the equation last demonstrated that only size of expressive vocabulary contributed independent variance. The patterns of prediction were quite different for late and early talkers from 13 to 20 months. Again, all of the communication measures (production vocabulary, comprehension vocabulary, gesture production, and the percent of comprehension vocabulary produced) contributed to predicting outcome for both groups. However, regression analyses indicated that only production and comprehension vocabulary contributed unique variance for early talkers. For the late talkers, on the other hand, only gesture production and the percent of comprehension vocabulary produced made unique contributions to continued delay. This is particularly interesting in light of findings by Bates et al. (this issue) regarding children with focal brain injury. In that study gesture production was significantly worse in children with right-hemisphere lesions, and the percent of comprehension vocabulary that is produced was the only measure that was sensitive to left-hemisphere damage. In addition, left- and right-hemisphere damage was associated with different profiles of early language delay in the children with focal brain injury. These findings suggest that the nature of the delay in late talkers (and, potentially, in children with specific language impairment) may be due to central nervous system phenomena which are bilateral and diffuse. These speculations suggest interesting directions for future cross-population research.

Although the experiments reported above provide solid evidence for continuity at the group level, they did not provide evidence of ability to predict outcome for individual children, a finding which supports Weismer et al. (1994). Discriminant analyses in Experiment 1 and Experiment 2 indicated poor prediction of Time 2 language status from variables measured 6 months earlier, and the descriptive analyses of individual children in the monthly longitudinal study (Experiment 3) show equally poor predictability for all late talkers and for early talkers below 18 to 24 months of age. This is demonstrated dramatically by subjects 7 and 19, for example, who score at the tenth percentile for vocabulary production at 16 months and above the 90th percentile at 30 months of age.

Although predictive ability for individuals is not good, it is the case that the two late talkers who remained low at 30 months (one at the 10th percentile, the other at the 12th) were delayed in both vocabulary comprehension and gesture production at 16 months of age. Thus, although delay in these variables does not indicate that a child will be delayed in language production at 30 months, the reverse is possible. That is, delays in these factors at an earlier time may be necessary for there to be a language delay at 30 months. This is yet another research direction that appears worth pursuing.

Professionals who have the responsibility of deciding whether children with early language delays should be referred for services want to be able to diagnose late talkers with some certainty. The limited value of any of the variables described above for predicting future language status of individual children puts them in a difficult position. We have labeled this phenomenon “the Pediatrician's Dilemma”. Briefly, because only 10 percent of all children meet late-talker criteria, you can be right 90% of the time if you claim that late talkers do not exist. Additionally, since a limited percent of late talkers retain that status, chances of being right are even higher. The critical question is whether this matters: Does it really make a difference if exceptional children are not identified before they are three or four years old? For linguistically precocious children the answer is probably no. However, for children who are truly language delayed and who are destined for learning difficulties in school, the long-term consequences of later identification could be serious. Clearly, carefully designed long-term studies are necessary to determine whether this is the case. However, since we know that the age and language variables identified in this set of studies, along with other factors such as family history, are predictive at a group level, it behooves us to treat them as we treat risk factors for afflictions such as cancer and heart disease. That is, given one or two of the risk factors, children should be carefully evaluated for the presence of others and monitored more frequently throughout the preschool years for development of additional risk factors or a clear language disorder.

REFERENCES

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic Press.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* New York: Cambridge University Press.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J.S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21,* 85-124.

Bates, E., Dale, P.S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 96-151). Oxford: Basil Blackwell.

Bates, E., Thal, D., Aram, D., Eisele, J., Nass, R., & Trauner, D. (1995). From first words to grammar in children with focal brain injury. To appear in D. Thal & J. Reilly, (Eds.). Special issue on Origins of Communication Disorders. *Developmental Neuropsychology* .

Bureau of the Census (1991). *General population characteristics.* (Vol. 1). Washington, DC: U.S. Government Printing Office.

Federal Register (1986). *Education of Handicapped Act Amendment.* Washington, DC: 100 STAT. 1145.

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.

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

Gesell, A. (1925). *The mental growth of the preschool child: A psychological outline of normal development from birth to the sixth year, including a system of developing diagnosis.* New York: Macmillan.

Goldfield, B. (1987). The contributions of child and caregiver to referential and expressive language. *Applied Psycholinguistics, 8,* 267-280.

Goodman, J. C., & Bauman, A. (1995). Group uniformity and individual differences in the rate and shape of language development. *Society for Research in Child Development Abstracts, 10,* 112.

Goodman, J.C., Jahn-Samilo, J., & Bates, E. (1995). *Patterns of vocabulary acquisition: Evidence from a longitudinal study.* Manuscript in preparation.

Hampson, J., & Nelson, K. (1993). Relation of maternal language to variation in rate and style of language acquisition. *Journal of Child Language, 20,* 313-342.

Hardy-Brown, K. (1983). Universals and individual differences: Disentangling two approaches to the study of language acquisition. *Developmental Psychology, 19,* 610-624.

Hollingshead, A. (1965). *Two-factor index of social position.* New Haven, CT: Yale University Press.

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.

McCarthy, M (1954). Language development in children. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley.

Munro, B. & Page, E. (1993). *Statistical methods for health care research.* (2nd Ed.). Philadelphia: JB Lippincott

Paul, R. (1991). Profiles of toddlers with slow expressive language development. *Topics in Language Disorders, 11*, 1-13.

Plomin, R. (1989). Environment and genes: Determinants of behavior. *American Psychologist, 44*, 105-111.

Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language, 20*, 43-60.

Rescorla, L., &Schwartz, E. (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics, 11*, 393-408.

Robinson, N.M., Dale, P.S., & Landesman, S. (1990). Validity of Stanford-Binet IV with linguistically precocious toddlers. *Intelligence, 14*, 173-186.

Shore, C. (1995). *Individual differences in language development.* Thousand Oaks, CA: Sage Press.

Thal, D., & Bates, E. (1988). Language and gesture in late talkers. *Journal of Speech and Hearing Research, 31*, 115-123.

Thal, D., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., & Bates, E. (1991). Early lexical development in children with focal brain injury. *Brain and Language, 40*, 491-527.

Thal, D., &Tobias, S. (1994). Relationships between language and gesture in normally developing and late-talking toddlers. *Journal of Speech and Hearing Research, 37*, 157-170.

Thal, D., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: A one-year follow-up. *Journal of Speech and Hearing Research, 34:3*, 604-612.

Tuchman, R., Rapin, I., & Shinnar, S. (1991). Autistic and dysphasic children I: Clinical characteristics. *Pediatrics, 88,* 1211-1218.

Vihman, M., & Miller R. (1988). Words and babble at the threshold of language acquisition. In M. D. Smith & J. L. Locke (Eds.), *The emergent lexicon: The child's development of a linguistic vocabulary* (pp.151-183). New York: Academic Press, Inc.

Weismer, S.E., Murray-Branch, J., & Miller, J.F. (1994). A prospective longitudinal study of language development in late talkers. *Journal of Speech and Hearing Research, 37,* 852-867.

Whitehurst, G., Fischell, J., Arnold, D., & Lonigan, C. (1992). Evaluating outcomes with children with expressive language delay. In S. Warren & J. Reichle (Eds.), *Causes and effects in communication and language intervention* (pp. 277-324). Baltimore: Paul Brookes.

Table 1

Comparison of Children Who Were Delayed, Precocious, or Average in Language at the Second Data Point on Measures of Language at the First Data Point.

Time 1 Variables	Late Talkers	Middle Level	Early Talkers	F-value	p-value	Post-hoc results
Total Production	37.5	186	357	40.58	<.001	LT<MET
Combining	0.35	1.03	1.63	18.06	<.001	LT<MET
Complexity	0.17	3.36	8.67	22.26	<.001	LT<MET
Mean of the 3 Longest Utterances	1.67	2.80	4.57	25.66	<.001	LT<MET
Closed Class Percentile Score	6	8.1	7.17	1.74	NS	NA

Table 2

Comparison of Children Who Stayed Late or Early Over the Two Time Points to Those Who Regressed Toward the Mean on Language Variables at the First Data Point

Time 1 Variables	Late at Both Data Points	Normal at Follow-Up	t-value	p-value	Early at Both Data Points	Normal at Follow-Up	t-value	p-value
Total Production	25.08	11.91	-2.02	.065	474.7	427.7	-1.54	NS
Combining	.017	0.00	NA*		1.87	1.90	0.25	NS
Complexity	0.83	.000	NA		13.87	6.9	-2.25	.04
Mean of the 3 Longest Utterances	1.52	1.00	NA		5.69	4.39	-1.79	.09
Closed-Class Percentile Score	8.62	7.03	-.41	NS	8.55	6.55	-1.52	NS

Table 3

Prediction of Late Talker Status From Eight Predictor Variables.

Actual Group (Time 2)	No. of Cases	Predicted Group Membership (Time 2)	
		Late Talker	Normal
Late Talker	16	11 (68.8%)	5 (31.3%)
Normal	158	39 (24.7%)	119 (75.3%)

Table 4

Prediction of Early Talker Status From Eight Predictor Variables.

Actual Group (Time 2)	No. of Cases	Predicted Group Membership (Time 2)	
		Early Talker	Normal
Early Talker	30	20 (66.7%)	10 (33.3%)
Normal	144	25 (17.4%)	119 (82.6%)

Table 5

Prediction of Continuity of Late Talker Status Using Eight Predictor Variables.

Actual Group	No. of Cases	Predicted Group Membership: Late Talker Times 1 and 2	Predicted Group Membership: Late Talker Time 1, Normal Time 2
Late Talker Times 1 and 2	11	8 (72.7%)	3 (27.3%)
Late Talker Time 1, Normal Time 2	10	2 (20.0%)	8 (80.0%)

Table 6

Prediction of Continuity of Early Talker Status From Eight Predictor Variables.

Actual Group	No. of Cases	Predicted Group Membership: Early Talker Times 1 and 2	Predicted Group Membership: Early Talker Time 1, Normal Time 2
Early Talker Times 1 and 2	15	10 (66.7%)	5 (33.3%)
Early Talker Time 1, Normal Time 2	10	1 (10%)	9 (90%)

Table 7

Comparison of Children Who Were Delayed, Precocious, or Average in Language at the Second Data Point on Measures of Language at the First Data Point and on Demographic Variables.

Time 1 Variables	Late Talkers	Middle Level	Early Talkers	F-value	p-value	Post-hoc results
Total Comprehension	79.70	103.73	230.96	33.52	<.0001	LT=M<ET
Total Production	4.13	17.72	102.65	60.48	<.0001	LT<M<ET
Total Gesture Production	23.33	31.25	43.74	25.17	<.0001	LT<M<ET
% of Comprehension Vocabulary Produced	7.69	16.47	41.41	34.91	<.0001	LT<M<ET

Table 8

Comparison of Children Who Stayed Early Over the Two Time Points to Those Who Regressed Toward the Mean on Language Variables at the First Data Point.

Time 1 Variables	Early at Both Data Points	Normal at Follow-Up	t-value	p-value
Total Comprehension	230.13	138.32	-2.94	<.007
Total Production	131.00	40.07	-3.69	<.002
Total Gesture Production	42.69	33.75	-3.15	<.004
% of Comprehension Vocabulary Produced	51.93	30.18	-3.70	<.001

Table 9

Prediction of Late Talker Status Using Seven Predictor Variables.

		Predicted Group Membership	
Actual Group	No. of Cases	Late Talker Time 2	Normal Time 2
Late Talker Time 2	30	27 (90%)	3 (10%)
Normal Time 2	187	45 (24.1%)	142 (75.9%)

Table 10

Prediction of Early Talker Status Using Seven Predictor Variables.

Actual Group	No. of Cases	Predicted Group Membership: Early Talker Time 2	Predicted Group Membership: Normal Time 2
Early Talker Time 2	23	16 (69.6%)	7 (30.4%)
Normal Time 2	194	11 (5.1%)	183 (94.3%)

Table 11

Prediction of Continuity of Early Talker Status Using Seven Predictor Variables.

Actual Group	No. of Cases	Predicted Group Membership: Early Talker Times 1 and 2	Predicted Group Membership: Early Talker Time 1, Normal Time 2
Early Talker Time 1	16	11 (68. 8%)	5 (31.3%)
Early Talker Time 1, Normal Time 2	28	3 (10.7%)	25 (89.3%)

Table 12

Production Vocabulary Status at 30 Months of Age for 24 Subjects Identified as Late, Early, or Average Between 18 and 24 Months.

Subject	Gender	18-24 Month Status	Percentile	
			29 Month	30 Month
1	M	Late	Missing	12
2	M	Late	55	73
12	M	Late	8	8
28	F	Late	23	24
31	M	Late	17	26
7	F	Early	100	100
16	M	Early	100	100
19	M	Early	94	95
22	M	Early	98	99
27	F	Early	92	Missing
38	M	Early	96	98
4	F	Average	Missing	98
6	M	Average	60	54
10	F	Average	54	57
11	F	Average	87	88
13	M	Average	74	76
14	M	Average	70	Missing
15	F	Average	93	92
17	M	Average	Missing	64
24	F	Average	Missing	54
26	M	Average	95	96
30	F	Average	98	Missing
32	M	Average	Missing	12
34	M	Average	16	Missing
35	F	Average	92	Missing
36	F	Average	Missing	57
37	M	Average	17	13
41	M	Average	38	33

Note. Status is determined by percentile scores on the MacArthur Communicative Development Inventory.

Table 13

Comprehension Vocabulary and Gesture Production Status at 16 Months of Age for 24 Subjects Identified as Late, Early, or Average Between 18 and 24 Months.

Subject	Gender	18-24 Month Status	16 Month Production Percentile	16 Month Comprehension Percentile	16 Month Gesture Production Percentile
1	M	Late	1	1	10
2	M	Late	41	9	25
12	M	Late	26	1	9
28	F	Late	2	1	3
31	M	Late	1	1	9
7	F	Early	18	97	45
16	M	Early	30	13	11
19	M	Early	25	5	17
22	M	Early	100	100	99
27	F	Early	57	31	85
38	M	Early	98	97	95
4	F	Average	50	24	60
6	M	Average	82	26	17
10	F	Average	78	33	60
11	F	Average	52	58	23
13	M	Average	33	33	13
14	M	Average	94	100	97
15	F	Average	23	34	99
17	M	Average	83	76	99
24	F	Average	87	52	5
26	M	Average	92	71	63
30	F	Average	21	22	24
32	M	Average	74	15	75
34	M	Average	28	27	16
35	F	Average	24	89	65
36	F	Average	53	64	95
37	M	Average	15	7	50
41	M	Average	23	77	90

Note. Status is determined by percentile scores on the MacArthur Communicative Development Inventory.

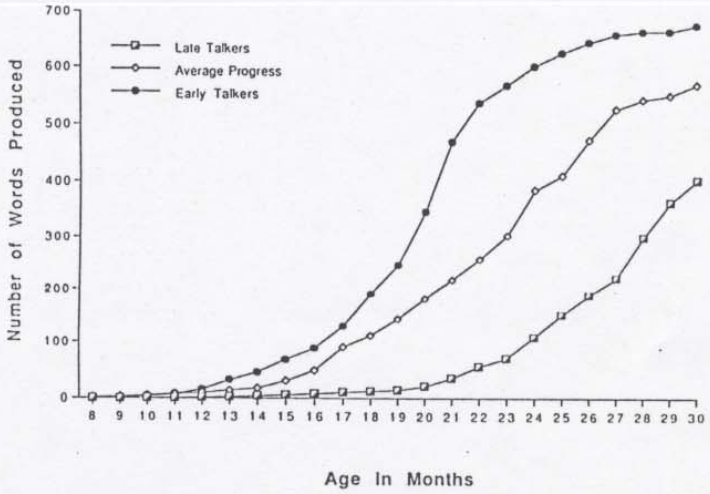

FIGURE 1 Mean number of words produced at each month from 8 to 30 months of age by children who were classified as late, early, or progressing at an average rate at some point between 18 and 24 months of age.

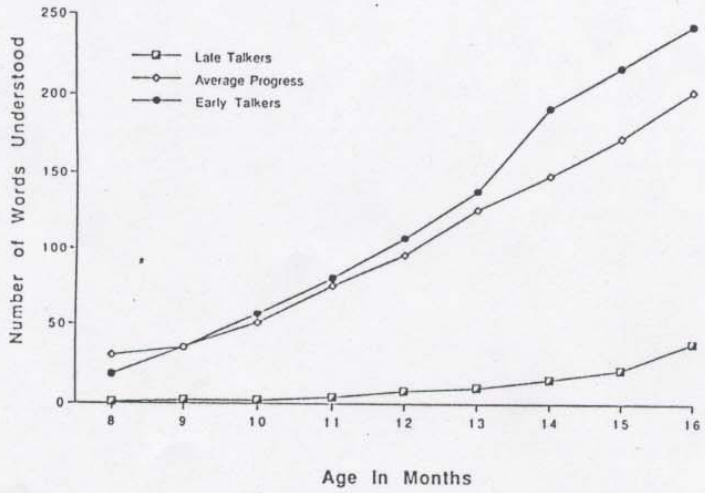

FIGURE 2 Mean number of words understood at each month from 8 to 30 months of age by children who were classified as late, early, or progressing at an average rate at some point between 18 and 24 months of age.

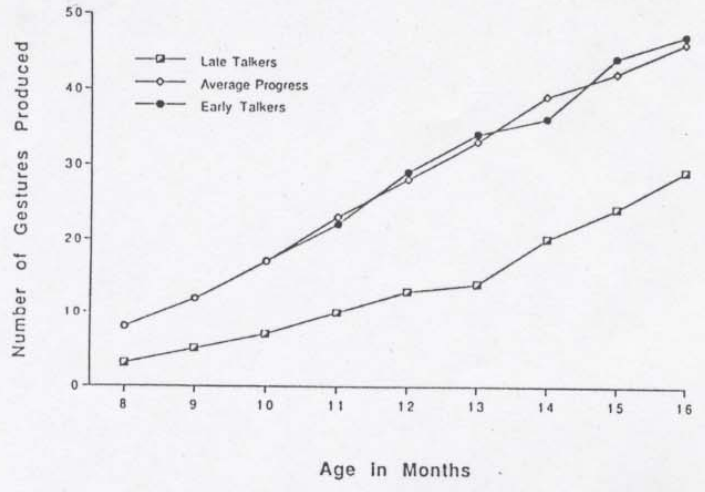

FIGURE 3 Mean number of gestures produced at each month from 8 to 30 months of age by children who were classified as late, early, or progressing at an average rate at some point between 18 and 24 months of age.