Published in final edited form as:

Dev Psychol. 2012 March ; 48(2): 477–491. doi:10.1037/a0025889.

Stability of Language in Childhood: A Multi-Age, -Domain, -Measure, and -Source Study

Marc H. Bornstein and Diane L. Putnick

Child and Family Research Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Public Health Service

Abstract

The stability of language across childhood is traditionally assessed by exploring longitudinal relations between individual language measures. However, language encompasses many domains and varies with different sources (child speech, parental report, experimenter assessment). This study evaluated individual variation in multiple age-appropriate measures of child language derived from multiple sources and stability between their latent variables in 192 young children across more than 2 years. Structural equation modeling demonstrated the loading of multiple measures of child language from different sources on single latent variables of language at ages 20 and 48 months. A large stability coefficient ($r = .84$) obtained between the 2 language latent variables. This stability obtained even when accounting for family socioeconomic status, maternal verbal intelligence, education, speech, and tendency to respond in a socially desirable fashion, and child social competence. Stability was also equivalent for children in diverse childcare situations and for girls and boys. Across age, from the beginning of language acquisition to just before school entry, aggregating multiple age-appropriate methods and measures at each age and multiple reporters, children show strong stability of individual differences in general language development.

Language Development by Individual and Group: Two Developmental Psychologies

Development is in many ways synonymous with growth and change. In the realm of language development, growth and change are especially salient. As the toddler emerges out of the infant and the child out of the toddler, one of the most prominent and readily observable developmental characteristics is growth and change in the child's language. But the coin in this as in other realms of development has two sides. The complements of growth and change in development are continuity and stability. Although development may be commonly identified with change, some features of human development are theorized to remain (more or less) consistent over time. Furthermore, transformation versus constancy in developmental science is crossed with two similarly related and complementary temporal concerns: group average performance across time and individual variation around that average (Cairns, 1979; Hartmann, Pelzel, & Abbott, 2011; Wohlwill, 1973). In developmental study, group mean level consistency or inconsistency (continuity/discontinuity) is often contrasted with individual order consistency or inconsistency (stability/instability). This paper is centrally concerned with stability through time in the important developmental domain of child language.

Individual Variation

Within any group at every age, human beings tend to vary dramatically amongst themselves on any given psychological characteristic (construct, structure, function, or process). It is commonly understood that variation among individuals in diverse characteristics appears in normal (Gaussian) distributions in the population. So, to continue our example, at virtually every age, children vary in terms of individual differences in their language (e.g., Bornstein & Haynes, 1998; Feldman et al., 2000; Fenson et al., 1993, 1994, 2000; Thal & Bates, 1990). Bornstein, Cote, et al. (2004) reported that 20-month-olds around the world have a range in their reported productive vocabulary from as few as 1 to as many as 487 spoken words (see also Feldman et al., 2000). Similarly, Le Normand, Parisse, and Cohen (2008) reported standard deviations near 200 words (suggesting a normal range, $\pm 2 SD$, of 0 to over 1000 words) for children at 48 months. The first question addressed in this paper concerned *variation* associated with domains, measures, and sources of child language.

Developmental Stability and its Moderators

Developmental science is centrally concerned with a critical issue about this individual variation: its *stability* (Bornstein & Bornstein, 2008). Stability describes consistency in the relative standing or ranks of individuals in a group on some characteristic through time. Stability in language obtains when some children display a relatively high level of language at one point in time vis-à-vis their peers and continue to display a high level at a later point in time, where other children display consistently low levels at both times. Instability in language obtains when individual children in a group do not maintain their relative order through time. Longitudinal developmental designs are requisite to addressing questions about stability.

The study of individual stability is important for several reasons. One is that findings of stability tell us about the overall developmental course of a given characteristic. Whether individuals maintain their order on some characteristic through time informs not only about individual variation, but contributes to understanding more about the origins, nature, and future of the characteristic as well (Appelbaum & McCall, 1983; McCall, 1981). Past performance is often the best predictor of future performance (Sternberg, Grigorenko, & Bundy, 2001). Two additional reasons that understanding stability is important are that child characteristics --especially stable ones -- signal developmental status to others and they affect the child's environment (Lerner & Busch-Rossnagel, 1981). For example, children's vocalizations and words used during social interactions have been employed to quantify how children socialize with others (Eckerman, Davis, & Didow, 1989; Guralnick, 1980; Howes & Matheson, 1992). Furthermore, interactants often adjust to match another's consistent characteristics. So, adults modify their language to harmonize with the language of children. For example, mothers fine-tune the semantic and syntactic contents of their utterances in concert with their children's level of language understanding (Bellinger, 1980; Clarke-Stewart, Vanderstoep, & Killian, 1979; Cross, 1977, 1978). The mean length of mothers' utterances tends to match the mean length of those of their young children (McLaughlin, White, McDevitt, & Raskin, 1983). Last, some degree of stability is prerequisite to establishing that a measure constitutes a suitable individual-differences metric. To be meaningful, a characteristic normally needs to show some consistency across time (Hartmann et al., 2011). In a nutshell, developmentalists are broadly interested, not only in how characteristics manifest themselves, but also in their group and individual developmental course -- their continuity and stability through time.

Although stability is unequivocally central in developmental science, it has many instantiations and interpretations (Bornstein & Bornstein, 2008; Hartmann et al., 2011;

Wohlwill, 1973). Notably, stability itself varies, and variation in stability generically, and here of language specifically, reflects many factors. Two concern content, three procedure, and two time. Estimates of language stability vary with all these factors: the domain and measure of child language, the method, source, and context, as well as the age of the child and the temporal interval between assessments. In this paper, we re-visit the issue of stability in language development across early childhood, and, faithful to these three considerations, we do so out of a multi-age, -domain, -measure, and -source framework. Here, we briefly (not exhaustively) review and illustrate the extant literature in language stability pertinent to these three considerations in the approximate age range we studied, and we describe how the present effort attempts to advance our understanding of stability in early child language development.

With respect to content, child language is comprised of many domains (phonology, lexicon, grammar, pragmatics, and so forth), and individual measures (expressive or receptive vocabulary, number of different word roots, or mean length of utterance from child speech) typically target only one. Empirically, then, stability studies usually examine consistency of single measures of a single domain of language. However, some contents and procedures yield more stable estimates than others, and generally the less information that is available, the lower the stability estimate (Hartmann et al., 2011).

With respect to procedural considerations, method, source, and context all moderate stability. Moreover, method is crossed with source in most designs. The three principal options to study child language include assessing children directly by recording what children say, seeking out those people closest to children (like their parents) to report about them, and interviewing or testing children directly through experimentation. Each source proffers valid insights on a child's language, but each also offers a unique perspective with its own limitations and its own implications that can be expected to yield different stability estimates. Brief discussion of each illustrates why.

Observation of what children themselves naturally and spontaneously say has the undeniable appeal of ecological validity and is direct and objective. However, records of free speech can underestimate a child's language, and numerous decisions about recording and analysis (when to observe, how frequently, where, under what conditions, and with whom) necessarily frame the resulting picture of child language (Bornstein, Painter, & Park, 2002). Aside from a few notable exceptions (Cameron-Faulkner, Lieven, & Theakston, 2007, who used a "dense" database), researchers usually sample only a small fraction of the child's everyday life, and so stabilities based on naturalistic sampling may be constrained. (Stability relies on correlation which can be attenuated by restriction of range.) Other limitations of observation are that it cannot access much about children's comprehension and a lack of production masks comprehension skill.

An alternative approach to child language employs reporters, and much of the classic information about child language development since Darwin (1877) has derived from parental reports (e.g., Dromi, 1987; Leopold, 1949; Weir, 1962). Diaries, interviews, and questionnaires (or checklists) can provide sources of data not readily available to observation or testing. Parental reports offer comprehensive information about children because they come from observers who know the child best and are with the child all the time (e.g., Thomas, Chess, Birch, Hertzog, & Korn, 1963). They have provided revealing information for rapid overall evaluation (Dale, Bates, Reznick, & Morisset, 1989), when contextual information is informative or required (Gopnik & Meltzoff, 1986), and in cases of rare occurrence (Bowerman, 1985). Language checklists are efficient, economical, and can be based on extensive sampling across a wide range of situations and times (Bates, Bretherton, & Snyder, 1988; Thal & Bates, 1990).

All that said, reports tend to be unsystematic and are often retrospective; they may include subjective components reflective of extraneous reporter characteristics (e.g., employment status, achievement orientation, personality, parity, and so forth); and they place multiple demands on untrained parents to detect, observe, and interpret various aspects of communication; parents *qua* reporters may underestimate children or they may be overgenerous (Bates & Bayles, 1984). Who is reporting makes a difference as well. Children spend time with many adults, including mothers, fathers and childcare providers (Clarke-Stewart & Allhusen, 2002; Parke, 2002), and the situations and activities in which children spend time with these different people vary. Different conversants and different situations and activities can be expected to give rise to the use of different language forms. As a consequence any one adult familiar with a particular child may know that child's communicative abilities, but only on the basis of his or her own unique contacts with the child, and asked to report about that child's communicative ability, different reporters will not necessarily agree (De Houwer, Bornstein, & Leach, 2004). Each of these factors can modify estimations of the child (Achenbach, McConaughy, & Howell, 1987; Bates & Bales, 1984; Bornstein, Gaughran, & Homel, 1986; Kazdin, 1988; Verhulst & Koot, 1992). Moreover, commonly used fixed-item checklists are representative indexes of children, not exhaustive diaries of their knowledge, and so do not give a complete picture of each child's language (Bates et al., 1994).

Standardized tests, the last method typically used to assess child language, include related items or tasks, and children receive scores based on the number they complete successfully (relative to other children the same age). In testing, the context is normally controlled, and assessing children in such a structured way is seemingly equal and fair. However, child performance may be affected, in part at least, by the testing method as well as by noncognitive factors, such as personality, motivation, adaptive functioning, and self-efficacy (Sternberg et al., 2001; Zigler, Abelson, & Seitz, 1973). Young children sometimes prove to be poor participants in formal assessments, often reluctant to interact with strangers or unwilling to cooperate during testing. There are manifold difficulties inherent in administering formal tests to infants and toddlers. Structured tests have also been criticized as providing only a limited picture of the richness and complexity of the child's language. Also, testing might show that some capacity or performance is possible, but it does not show whether the capacity or performance is typical. For these reasons, questions inevitably arise about the generalizability of standardized test results in terms of what they reveal about the child's language as it occurs and is used in everyday communication (Leonard, Prutting, Perozzi, & Berkeley, 1978; Owens, 1995).

Beside the method and source of information, another procedural characteristic can be expected to moderate stability: whether assessments are made across consistent or inconsistent contexts (the former enhances stability and the latter attenuates stability). For example, free play with parents might elicit one set of verbal skills from children, whereas structured interactions, such as at a meal or while learning, might elicit quite a different set (Bornstein, Tamis-LeMonda, & Haynes, 1999).

Finally, with respect to temporal parameters, a characteristic may not be stable at one age in the life course, but stabilize at a later age. Generally, infancy and early childhood are thought to be less stable or predictive periods in life (Bayley, 1949), and people are thought to become increasingly consistent in relation to one another as they age (Roberts & Del Vecchio, 2000; Sternberg et al., 2001). Furthermore, the shorter the inter-assessment interval, normally the greater the stability estimate of a characteristic (the Guttman, 1954, simplex).

In summary, different assessment contents, procedures, and times used in measuring child language can contribute to different stability estimates; however, if applied conjointly as we do here, these different parameters might also complement one another to bring into focus a more coherent picture of the child. Because no one approach to developmental assessment is superior to all others under all situations, in this study of language stability in young children we investigated the shared contribution of diverse approaches.

A Latent Variable of Child Language

The second question addressed in this paper concerned whether these different paths to child language converge on a consistent picture. This study combined several domains, measures, and sources to extract *latent variables* of child language at each of two ages. The main advantage of using a latent variable to measure the stability of language is that the latent variable captures shared variance among its indicators. Thus, any variance uniquely associated with, for example, rater bias, systematic measurement error, or random error for a particular indicator is relegated to its error term. The resulting latent variable is a purer measure of the construct that incorporates the perspectives of multiple raters and domains of child language, and so stability of that construct can be assessed more precisely (Kline, 1998).

The data available to date on concurrent relations between different approaches to language measurement are fragmentary, but suggestive. As the number of possible permutations across studies in this field is large, any summary sketch can only convey a sense of the extant literature. Nonetheless, concurrent convergence among measures on the same children tends to be high: Parent reports correlate with spontaneous child speech (Bornstein & Haynes, 1998; Corkum & Dunham, 1996; Dale et al., 1989; Miller, Sedey, & Miolo, 1995); parent checklists and reports correlate with laboratory measures and experimenter assessments (Bates & Carnevale, 1993; Bornstein & Haynes, 1998; Chaffee, Cunningham, Secord-Gilbert, Elbard, & Richards, 1990; Dale, 1991; Fenson et al., 1994; Miller et al., 1995; O'Hanlon & Thal, 1991; Saudino et al., 1998); parent diaries correlate with checklists (Reznick & Goldfield, 1994); and observed child speech correlates with experimenter assessments (Bornstein & Haynes, 1998). To address the second question posed in this study we determined whether and how well several different measures from different sources converge on a single latent variable of general child language at each of two ages.

Stability of Child Language, "Third Variables," and Gender

In this study, the same children were first seen at 20 months and again at 48 months of age. The third question concerned *stability* between language latent variables at these two ages. Because of the plethora of methodological permutations, the extant body of work on stability of language does not submit to easy summary either; but a series of examples published over the last quarter century serves to convey a sense of this literature. All of the following studies (arrayed chronologically) fall in the approximate age range of the present investigation and have reported significant (if varying levels of) temporal stability of individual differences for a diversity of language measures: Sparrow, Balla, and Cicchetti (1984) for the Communication Domain of the Vineland Adaptive Behavior Scales (VABS) between 3 years and 4 years, 11 months; Olszewski (1987) for verbal fantasy play in 3- to 5-year-old children; Bates et al. (1988) for referential style between 13 months and at 50 words; Reznick (1990) for vocabulary comprehension to vocabulary production between 14 and 20 months; Blake, Quartaro, and Onorati (1993) for MLU between 18 months and 57 months; Beals and De-Temple (1993) for language level to oral production and comprehension between 3 and 5 years; Martlew and Sorsby (1995) for letter naming to level of emergent writing and familiarity with the alphabetic system between 4 and 5 years; Pine,

Lieven, and Rowland (1996) for observational and checklist measures of vocabulary between 1 and 2 years; Gavin and Giles (1996) for MLU in children 31 months to 46 months; Burgess (1997) for phonological awareness skills between 3 and 5 years; Bornstein et al. (1999) for vocabulary production across 2 contexts between 13 and 20 months; Feldman et al. (2000) for phrases understood, vocabulary comprehension, vocabulary production, and total gestures to vocabulary production, irregular word forms, overregularized words, length of the three longest sentences, and sentence complexity from 12 to 24 months; Dickinson and Tabors (2001) for oral language and literacy between 3 and 14 years; Storch and Whitehurst (2002) for oral and code-related skills to literacy between 4 and 10 years; Winsler, René de León, Wallace, Carlton, and Willson-Quayle (2003) for private speech of 3- and 4-year-olds over a 6 month interval; and Bornstein, Hahn, and Haynes (2004) for age-appropriate maternal questionnaires, maternal interviews, teacher reports, experimenter assessments, and transcripts of children's spontaneous speech from 1 year, 1 month to 6 years, 10 months. In overview, temporal intercorrelations of individual differences across a variety of tasks and intervals for a number of different language variables and metrics show significant, but variable, stability.

These studies purport to report stability of specific language measures, but as a collection they suffer certain noteworthy shortfalls. Some draw from the same reporter (other than the child), and so shared source variance may inflate stability correlations. Many reuse the same measures and so provide only a narrow view of language at the same time they capitalize on practice effects and shared method variance. Most do not take other endogenous or exogenous factors into consideration to assign stability to the child more unambiguously (unmeasured third variables may carry a bivariate association). Bornstein et al. (1999) attempted to account for such parameters; they reported stability of both word types and MLU in children's spontaneous speech from 13 months to 20 months, taking into consideration maternal word types and verbal responsiveness. Likewise, Aram (2005) reported stability of literacy measures between 5½ and 8 years that held beyond family measures of SES, maternal literacy, and literacy tools at home. Furthermore, stability study doubtless suffers the "file drawer" problem, it being unlikely that nonsignificant stabilities have appeared in the published literature (Rosenthal, 1979).

Stability of a target characteristic (such as language) is usually ascribed to consistency of that characteristic in the individual. However, stability might also be attributable to other stable endogenous (genetic, biological, maturational) characteristics in the individual that are related to the target characteristic, or stability in the target characteristic might be attributable to a stable environment that supports consistency in the target characteristic (Bornstein, Putnick, & De Houwer, 2006; Roberts & DelVecchio, 2000). (Many characteristics like language are sensitive to experience; Bornstein et al., 1999; Hart & Risley, 1992, 1995.) With respect to endogenous characteristics, children's social competence as well as maternal verbal intelligence have been implicated in child language (Colledge et al., 2002; Dionne, Dale, Boivin, & Plomin, 2003; Hardy-Brown, Plomin, & DeFries, 1981; Irwin, Carter, & Briggs-Gowan, 2002; McGregor & Capone, 2004; Pinker, 2007; Winsler et al., 2003). With respect to exogenous factors, on the one hand maternal education and parenting, especially language addressed to the child, is acknowledged to be relatively stable (Belsky & Jaffe, 2006; Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008; Dallaire & Weinraub, 2005; Holden & Miller, 1999), but on the other hand young children today experience changing rearing and language environments, as from one to another daycare setting, "before and after" (wrap around) childcare, in the numbers of children and caregivers, and so forth (NICHD, 2002; Peisner-Feinberg et al., 2001). Only some children in the age range of this study were at home full time, and the language environments that children are exposed to can be highly variable in terms of caregivers, other children, and so forth, regardless of child SES.

The fourth question therefore asked whether child language is stable in itself or if some *third variable* covaried with child language and accounted for child language stability. To address this question, we collected and analyzed a number of covariates and recalculated stability. Covariate analyses took into account family socioeconomic status, maternal verbal intelligence, education, speech, and tendency to respond in a socially desirable fashion, as well as children's own social competence. Furthermore, we assessed whether children in stable vs. unstable childcare situations (and therefore stable vs. unstable language environments) exhibited similar levels of language stability.

Finally, early language development is thought to differ in girls and boys (Bauer, Goldfield, & Reznick, 2002; Feldman et al., 2000; Fenson et al., 1994; Huttenlocher, Haight, Bryk, Seltzer, Lyons, 1991; Van Hulle, Goldsmith, & Lemery, 2004): Girls might have a linguistic advantage based on differential developmental timetables, gender-typed interests, and learning opportunities (Bornstein, Hahn, et al., 2004). It might be that gender differences are also manifest in stability. Therefore, the fifth question addressed in this paper concerned *gender* and asked if stability in child language is equivalent in girls and boys.

The Present Study

This study adds to the extant child language literature by assessing multiple domains using age-appropriate measures and sources to evaluate stability between latent variables of general language from the end of infancy to the start of formal schooling. Systematic multimethod investigation of stability with a relatively large sample is still rare in child language research, and Sroufe (1979) has described an organizational perspective on development that posits that the best way to study developmental consistency is to examine age-appropriate different yet conceptually related behaviors. We assessed the fit of several structural models to the data: one for the common convergence of multiple measures from different sources on single latent variables of child language at 20 and 48 months and the stability between those latent variables, and a second for stability of the language latent variables, controlling multiple covariates. We also evaluated separate stability fits for children with stable and unstable childcare situations and for girls and boys.

Method

Participants

All language assessments in the two longitudinal waves were completed for 192 firstborn children (87 girls and 105 boys). Children averaged 20.07 months ($SD = .20$) at the first assessment and 48.53 months ($SD = .93$) at the second assessment. All but 5 children were born at term (within 3 weeks of the due date); the 4 preterms and 1 postterm did not emerge as outliers and so were retained in the analyses. All children were typically developing healthy monolingual English speakers. Children varied in the number of hours they spent in non-parental childcare per week, $M = 25.20$, $SD = 21.02$, range = 0 – 72 at 20 months, and $M = 26.52$, $SD = 19.90$, range = 0 – 74 at 48 months. This analysis draws on a published longitudinal data base (Bornstein, Hahn, et al., 2004; Bornstein & Haynes, 1998) that used two cohorts. Data for the first cohort at 20 months were collected between 1989 and 1993 and for the second cohort between 1995 and 1999.

Language emerges in the second year of life, and so we began this investigation of language stability at that time. In the second year, children typically exhibit rapid increases in receptive and expressive language when they begin to understand and to say sound sequences that function as true naming, they shift from context-restricted purely performative uses of words or phrases to flexible usage across contexts, and general semantic meanings used to express possession, location, and action regularize, and word

combinations appear. In the child's fourth year, language in all domains has blossomed, and children are sophisticated in individual language and as conversants, having acquired the ability to communicate verbally about their thoughts, beliefs, and desires; yet, most children this age have still not experienced the homogenizing influences and intellectual and social rigors of mandatory formal schooling. As a consequence, the period bracketed by children's second and fourth birthdays seemed to us to constitute a formative time to investigate the nature of stability in child language.

Mothers averaged 31.34 years ($SD = 5.93$) at the first assessment. Over half (65.8%) of mothers had completed a standard college or university degree. Most mothers were working at the times of the two visits (64.58% at 20 months and 62.90% at 48 months), and hours of employment averaged 21.70 ($SD = 18.74$; range = 0 – 60) at the first visit and 21.64 ($SD = 19.38$; range = 0 – 65) at the second visit. Most (94.3%) families were intact, and family SES ranged from 20.50 to 66.00 ($M = 51.79$, $SD = 12.01$) on the Hollingshead (1975) Four-Factor Index of Social Status. On average, families were middle to upper-middle class, but the sample included families from nearly the full range of SES. Families of girls and boys did not differ on any of the sociodemographic variables. We recruited English-speaking, European American families in the United States because research in child language with English-speaking European Americans provides a familiar reference point; they are also currently the majority cultural group in the United States (Tilton-Weaver & Kakihara, 2008; U.S. Census Bureau Population Division, 2001). Our community sample was socioeconomically heterogeneous in terms of maternal education and family SES, but ethnically homogenous as a first step in understanding child language and its stability before embarking on more complex studies and analyses with diverse samples. By including only European American children, we intentionally avoided ethnicity variation and an ethnicity-SES confound that might cloud any findings.

Procedures

At both 20 and 48 months, multiple measures of child language were obtained. We chose the following measures to represent one-to-two indicators of child language per source (child observation, maternal report, experimenter assessment) as well as diverse age-appropriate domains of language (utterance length, vocabulary, receptive communication, sentence structure, adaptive communication, word associations, and written communication).

20 months: Child observation—Children's language was derived from transcripts of their spontaneous speech in a free-play interaction with their mothers at home. The first 10 min available from videorecords after the start of free play were transcribed verbatim by professional transcribers naive to all factors in the study. Each transcript was then checked against the record for accuracy by another researcher. Utterances that were unintelligible to transcribers or whose content transcribers could not agree on were marked as unintelligible. Transcripts were then coded to permit their analysis using Computerized Language Analysis (CLAN) following the conventions of the Codes for the Human Analysis of Transcripts (CHAT; MacWhinney, 2000). Free and bound morphemes were parsed to allow calculation of mean length of utterance in morphemes. Following CHAT rules, an utterance was defined as a unit of speech representing a complete thought as indicated by intonation and/or pauses. Multiple utterances per turn are possible in a conversation. Two measures of each child's language were calculated. Mean length of utterance (MLU): The complexity of speech was assessed based on a count of morphemes in each complete and intelligible utterance, averaged across all utterances for a child. Different word roots (DWR): A count of the number of different lexical items (ignoring inflections) was divided by the total number of min in the session. The mean number of morphemes in young children's utterances has proven to be a reliable index of their grasp of language (Bornstein & Haynes, 1998), and

MLU is a reliable indicator of verbal complexity as well as grammatical development in child speech. Blake et al. (1993) reported that child MLU showed high convergent validity with measures of syntax (.88) and grammar clauses (.82), and Pan, Rowe, Spier, and Tamis-LeMonda (2004) reported that DWR (Types) had high convergent validity with other vocabulary measures.

20 months: Maternal report—The Vineland Adaptive Behavior Scales - Interview Edition, Survey Form (VABS; Sparrow et al., 1984) were used to obtain mothers' estimates of their children's receptive and expressive communication skills. This semi-structured interview with mother was carried out by trained staff. The Communication Domain of the VABS is the sum of raw scores for the Receptive, Expressive, and Written subdomains. However, because of the floating baseline and ceiling on the VABS, mothers of 20-month children were unlikely to be asked questions from the Written subdomain; 91% of the sample scored 0. The Receptive Subdomain of the VABS Communication Domain is based on responses to up to 13 items designed to assess children's ability to attend to, listen, and understand communications and to follow instructions. The Expressive Subdomain is based on responses to up to 31 items designed to assess children's preverbal and early speech activity and use of speech in interaction and in expressing abstract and complex concepts. The split-half reliability coefficient for the Communication Domain was .92 for a standardization sample of 200 children in their 2nd year, and test-retest reliability was .92 for 70 children between 6 and 35 months of age (Sparrow et al., 1984).

Mothers also completed the Early Language Inventory (ELI; Bates et al., 1988) at home. This checklist, a forerunner of the MacArthur Communicative Development Inventory (CDI; Fenson, Thal, & Bates, 1990; Fenson et al., 1993, 1994), provides a measure of children's expressive vocabulary. Mothers indicated which words of 643 (including nouns, action verbs, and adjectives that are commonly used by children of this age) they had heard their child spontaneously produce. Fenson et al. (1993, 1994) reported high (r s above .90) 6-week test-retest reliability for vocabulary production on the MacArthur CDI-Words and Sentences. The total number of words that mothers reported their children produced on the ELI was calculated for each child.

20 months: Experimenter assessment—An experimenter administered the Verbal Comprehension Scale 'A' and the Expressive Language Scale of the Reynell Developmental Language Scales-Second Revision (RDLS; Reynell & Gruber, 1990). The RDLS assesses language comprehension and production in children aged 1 year, 6 months through 6 years. In the Verbal Comprehension Scale 'A', the child is asked to demonstrate an understanding of increasingly difficult verbal expressions ranging from labeled objects to higher-order concepts. The Expressive Language Scale uses two subscales (appropriate to this age level): structure and vocabulary. From the child's spontaneous expressions during the visit, language structure is assessed on a scale which ranges from "vocalizations other than crying" to the "use of complex sentences." In the vocabulary subscale, the child is asked to name familiar objects and actions from pictures. Split-half reliability coefficients for the Comprehension and Expressive Scales for children between the ages of 2 years and 2 years, 5 months are both .93 (Reynell & Gruber, 1990). Children's raw scores for comprehension and production were used in the analyses.

48 months: Child observation—Measures of children's spontaneous language were derived from transcripts of a videorecorded storytelling session. The goal of the activity, to "tell a story about a Bear Family," was explained to the child. A set of toy props was introduced in a consistent order (Bear Family, living room, kitchen, park with rabbit and duck, policeman, and doctor), and each prop was verbally labeled on presentation to facilitate familiarization. The props were arranged in a standard configuration within easy

reach of the child who was seated at a low table. The child played with the props for 10 min. At the end of the play time, the child was prompted to tell a story about the Bear Family. If the child's story did not appear to be coming to an end after 5 min, the experimenter prompted the child by asking how the story ends. If the child produced a fragment of a story (i.e., a single act or event), and then fell into silence, the experimenter prompted the child with a question: "What happens next?" A total of 3 such prompts was allowed. If the child did not spontaneously offer a story, the experimenter followed with a series of prompts intended to coax the child into telling a story (e.g., "What is the Papa Bear doing?"). For the final prompt, the experimenter told a standard beginning of a story and the child was encouraged to finish the story. The child's MLU and DWR were calculated over the entire storytelling session in the same manner described for 20 months.

48 months: Maternal report—As at 20 months, the Vineland Adaptive Behavior Scales—Interview Edition, Survey Form (VABS; Sparrow et al., 1984) Communication Domain was used to obtain mothers' estimates of their children's communication skills. Mothers were asked to indicate to what extent their children had mastered a variety of verbal skills in the Communication Domain, including Receptive, Expressive, and Written language skills. The Receptive and Expressive Subdomains were identical to those at 20 months. The Written Subdomain is based on responses to up to 23 items designed to assess children's written language skills. The split-half reliability coefficient for the Communication Domain was .93 for a standardization sample of 200 4-year-old children, and test-retest reliability was .86 for 74 children between 36 and 59 months of age (Sparrow et al., 1984). The raw Communication Domain was used in analyses.

48 months: Experimenter assessment—An experimenter administered 2 verbal subtests of the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989). The WPPSI-R is an individually administered intelligence test for children from 3 years to 7 years, 3 months. It consists of 12 subtests that are divided into Verbal and Performance scales. Two Verbal subtests (Information and Similarities) were administered on the basis of their correlations with Verbal IQ scores (r s ranged from .71 to .82; Wechsler, 1989). The Information subtest consists of 27 items that measure the child's knowledge of objects and events by having the child point to pictures or provide a short verbal response to spoken questions. In the Similarities subtest, children point to 6 pictures of items that share a common feature, complete 6 sentences involving similar items, and explain how 8 objects are related. Raw scores were used. In a standardization sample of 200 48-month-old children, split-half reliability coefficients were .88 for Information and .89 for Similarities.

Covariates

We assessed the possibilities that family SES and mothers' verbal intelligence and education influenced child language, mothers' speech during mother-child interactions influenced child speech, mothers' tendency to respond in a socially desirable fashion influenced maternal report, and child social competence influenced experimenter assessments. We evaluated these several measures for use as covariates. Family SES and maternal education were obtained from the Hollingshead (1975).

Maternal verbal intelligence—The mother was administered the Peabody Picture Vocabulary Test--Revised (PPVT-R Form L; Dunn & Dunn, 1981). This instrument provides a standardized index of maternal vocabulary intelligence. In adult standardization samples, split-half reliability was .88, alternate-forms reliability was .81, short-term alternate-forms reliability was .68, and validity coefficients with full IQ scale scores ranged from .58 to .72 (Dunn & Dunn, 1981).

Maternal speech—Maternal speech was derived from transcripts of the free-play session at 20 months and a mother-child picture book-reading session at 48 months. Mothers' MLU and DWR were derived in exactly the same way as for children (see above).

Maternal social desirability—The Social Desirability Scale (SDS; Crowne & Marlowe, 1960) uses 33 items to assess adults' tendency to respond to questions in a socially desirable fashion (Johnson, Shavitt, & Holbrook, 2011). Mothers completed this social desirability scale to check for reporting bias on the VABS and ELI. The SDS has significant test-retest reliability ($r = .89$) and internal consistency ($\alpha = .88$; Crowne & Marlowe, 1960).

Child social competence—At 20 months, observed child sociability was used as the measure of social competence. In a child solitary play session, three behaviors were double coded (11% of the tapes) by trained coders: frequency of vocalizations directed to an experimenter ($\kappa = .82$), frequency of positive vocalizations with no specific object ($\kappa = .92$), and amount of time the child spent within 2 ft of the experimenter ($\kappa = .74$). Based on a behavioral coding system developed by Mullen, Snidman, and Kagan (1993), an aggregate sociability index was computed by standardizing and averaging the three measures. At 48 months, child anxiety and hyperactivity-distractibility were used as measures of (lack of) social competence. Mothers rated their children on the 9-item anxious-fearful scale and the 4-item hyperactive-distractible scale from Behar and Stringfield's (1974) Preschool Behavior Questionnaire. Items were rated on a 3-point scale (*Doesn't Apply*, *Applies Sometimes*, and *Certainly Applies*) and summed to form the scales. Internal consistency (α) was .63 for the anxious-fearful scale and .77 for the hyperactive-distractible scale.

Results

Preliminary Analyses and Analytic Plan

Prior to data analysis, variable distributions were examined for univariate normality (Tabachnick & Fidell, 2007). Most variables approximated the normal distribution, but ELI and DWR at 20 months were reexpressed using $\log_{10}$ and cube-root transformations, respectively, to improve their distributions. Descriptive statistics are presented in the variables' original metrics to aid interpretation. Despite attempts to approximate univariate normality, problems emerged with multivariate skewness = 17.86, $z = 9.34$, $p < .001$, and multivariate kurtosis = 160.25, $z = 5.60$, $p < .001$. Therefore, we employed LISREL 8.72 (Jöreskog & Sörbom, 2005) to compute asymptotic covariance matrices, we used robust maximum likelihood estimation, and we report the normal theory weighted least squares chi-square (χ^2). We also report the comparative fit index (CFI), non-normed fit index (NNFI), and root mean square error of approximation (RMSEA) as indicators of model fit. A model was considered to have good fit if the χ^2 test was nonsignificant ($p > .05$), the CFI and NNFI were at or above .90 (Bentler, 1990; Marsh, Balla, & Hau, 1996), and the RMSEA was .06 or smaller (Hu & Bentler, 1999), but we gave greater weight to the incremental fit indices than to χ^2 because the χ^2 value is known to be sensitive to sample size (Cheung & Rensvold, 2002) and the size of the correlations in the model (Miles & Shelvin, 2006). Standardized path coefficients are presented in text and figures.

After fitting an initial model on the full sample, we performed multiple-group models by caregiving situation and child gender to assess whether the full model fit for children in low and high numbers of caregiving situations and for girls and boys. For multiple-group models, we report the difference in χ^2 statistics and CFI values (Cheung & Rensvold, 2002) for 2 nested models (Vandenberg & Lance, 2000). If the $\Delta\chi^2$ between the unconstrained and constrained models was nonsignificant ($p > .05$), and the $\Delta\text{CFI} \geq .01$ (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000), the model was deemed to fit equally well in

pairs of groups. For correlations and standardized path coefficients, we adopt conventional magnitudes of r corresponding to small, medium, and large effect sizes as .10, .30, and .50, respectively (Cohen & Cohen, 1983, p. 61).

Descriptive Statistics and Correlations of Language Measures at 20 to 48 Months

Table 1 displays the descriptive statistics and correlations among the 20-month and 48-month language measures. The standard deviations and ranges in Table 1 indicate that there was considerable variation in performance on all language measures. At both 20 and 48 months, children in the sample varied from no spontaneous speech and below-average receptive and expressive communication to long sentences, diverse vocabulary, and well above-average receptive and expressive communicative skills. Despite being capable of speech, two children at 48 months failed to spontaneously produce any words in the storytelling task. These children scored well within the normal range on concurrent maternal report and experimenter assessments of language and did not emerge as outliers in analyses. Excluding these scores did not change the results of the study, and part of language competence is being able to think of something to say; therefore, we decided that the 0 scores for MLU and DWR reflected true aspects of each child's language and we retained the 2 children in the sample.

Correlations among the 20- and 48-month measures indicate medium to large convergent validity within ages (average correlation = .55, $SE = .07$, at 20 months, and average correlation = .23, $SE = .07$, at 48 months) as well as medium average stability from 20 to 48 months across all measures (average correlation = .27, $SE = .07$). The average correlation across measures from the same source (shaded cells in Table 1) was .30 ($SE = .07$).

Stability of Language from 20 to 48 Months

We tested an initial structural equation model with the 6 language indicators at 20 months loading on a single latent variable of 20-month language, the 5 language indicators at 48 months loading on a single latent variable of 48-month language, and the stability coefficient running from 20-month language to 48-month language. Although all paths were significant (except for the loading of 48-month DWR on the 48-month language factor), this initial model was a relatively poor fit to the data, $\chi^2(43) = 157.02$, $p < .001$; CFI = .93; NNFI = .91; RMSEA = .11, 90% CI = .092-.013. Modification indices suggested that 4 conceptually meaningful error covariances were warranted. Therefore, we added error covariances between MLU and DWR at both waves to account for shared source variance associated with direct observation, ELI and VABS at 20 months to account for shared source variance associated with maternal report, and WPPSI-R Verbal Information and Similarities at 48 months to account for shared method variance.

After accounting for these four error covariances, the model fit was acceptable, $\chi^2(39) = 73.91$, $p < .001$; CFI = .98; NNFI = .97; RMSEA = .066, 90% CI = .041-.090. The final model of the stability of child language between the two latent variables is presented in Figure 1. All paths were significant at $p < .01$ (except for the loading of 48-month DWR on the 48-month language latent variable). The robust standardized estimate of language stability from 20 to 48 months was .84, $p < .001$.

Stability of Language from 20 to 48 Months, Controlling Covariates

As a check against threats to the validity of the model, we explored whether several covariates were associated with child language measures. We correlated (1) family SES and maternal verbal intelligence and education with all 11 language measures; (2) maternal MLU and DWR with child MLU and DWR, respectively; (3) maternal social desirability bias with maternal reports on the VABS and ELI; (4) observed child sociability at 20 months

with 20-month MLU, DWR, and the Reynell scales; and (5) child social competence (mother-rated child hyperactivity-distractibility and anxiety) with 48-month MLU, DWR, and the WPPSI-R scales. Maternal verbal intelligence was correlated with WPPSI-R Verbal Information, $r(190) = .25$, $p < .001$; maternal education was correlated with the subtests of the WPPSI-R, $r_s(188) = .20$ to $.26$, $p_s < .001$; family SES was correlated with the subtests of the WPPSI-R, $r_s(185) = .29$ to $.40$, $p_s < .001$; maternal DWR at 20 months was correlated with children's DWR, $r(190) = .17$, $p < .05$; and children's hyperactivity-distractibility at 48 months was correlated with MLU, $r(182) = -.20$, $p < .01$, and both subtests of the WPPSI-R, $r_s(182) = -.33$ to $-.37$, $p_s < .001$. No other potential covariates were associated with the child language measures. Prior to recomputing the final model in Figure 1, we residualized 20-month DWR and 48-month MLU, VABS Communication, and WPPSI-R Verbal Information and Verbal Similarities for their associations with the above covariates. The same pattern of relations observed in the uncontrolled model held when residualized variables were used, $\chi^2(39) = 67.31$, $p < .01$; CFI = .98; NNFI = .97; RMSEA = .060, 90% CI = .033-.086. In the residualized model, the robust standardized estimate of language stability from 20 to 48 months was .79, $p < .001$.

Metric Equivalence across Childcare Situations

To assess whether child language was stable for children in relatively stable versus unstable language environments, for each child we counted the number of distinct childcare situations between the ages of 20 and 48 months and recalculated stability in low and high caregiving variability groups. Mothers' reports were coded into different caregiving situations based on the identity of the caregiver, location of care, number of hours, and number of other children present. A change in any one of these variables constituted a new caregiving situation. For example, a change from one caregiver to another was counted as a new situation, even though the location, number of hours, and number of other children might remain the same. Preschool entry was included as a new caregiving situation. The number of distinct caregiving situations was used as an indicator of variability of the child's language environment. Between 20 and 48 months, children were in 0 (maternal care only) to 10 different caregiving situations. We defined low caregiving variability as 0-2 situations ($n = 101$) and high caregiving variability as 3-10 situations ($n = 85$) across the 28-month interval and used this variable in a multiple-group model. We then tested whether the uncontrolled final model in Figure 1 fit equally well for children in low and high numbers of childcare situations. The difference in fit between (1) a model with all loadings and the stability coefficient constrained to be equal in the low and high caregiving variability groups and (2) a model with all paths free to vary in the low and high groups was nonsignificant, $\Delta\chi^2(10) = 12.23$, ns , $\Delta CFI = .00$, indicating that the meanings of the language latent variables and the stability of language were similar for children in relatively consistent and relatively variable childcare situations and (presumably) language environments.

Metric Equivalence in Girls and Boys

Last, we tested whether the uncontrolled final model in Figure 1 fit equally well for girls and boys. The difference in fit between (1) a model with all loadings and the stability coefficient constrained to be equal in girls and boys and (2) a model with all paths free to vary in girls and boys was nonsignificant, $\Delta\chi^2(10) = 13.33$, ns , $\Delta CFI = .00$, indicating that the meanings of the language latent variables and the stability of language were similar for girls and boys. Figure 2 shows the scatter plot of the language latent variable scores at 20 and 48 months for girls and boys.

Discussion

In overview, clear evidence emerged for strong stability of individual variation in general language across early childhood. Stability obtained independent of child sociability and childcare history, maternal intelligence, education, speech, and social desirability of responding, as well as family socioeconomic status. Stability was also similar in girls and boys. When multiple domains, measures, and sources are used, child language emerges as a robust and stable individual-differences characteristic. This study points to the value of using multiple informants and diverse components of language measured in different ways to obtain a picture of stable language development in children. Multiple assessments take more aspects of language into account and give more precise estimates.

Interindividual Variability

It comes as no surprise that children in their second and fourth years varied in each of several language measures. At 20 months, children's mean length of utterance, vocabulary, adaptive communication, and comprehension all varied widely. For example, on the ELI at 20 months, one child was reported to speak no words while another was reported to speak 466 words. At 48 months, children's storytelling, adaptive communication, and verbal intelligence also varied widely. For example, children's mean length of utterance in the storytelling task ranged from 0 to over 13 morphemes. Often, problems in distributions of dependent variables and narrow ranges in abilities of participants make it difficult to obtain high correlations. Our findings move away from restriction of range that may have plagued previous work. So, on the first question posed in this study, child language at different ages shows substantial interindividual variability in all the forms we measured.

Language Latent Variable

Surprisingly little is still known about covariance among different language measures at the phenotypic level of analysis because so few studies use a multimethod approach. In one exception, seven diverse measures of language proved to be intercorrelated and were used to create a general factor of language (Colledge et al., 2002). On the second question posed in this study, we found five to six different language measures from three data sources satisfactorily converged on consistent and coherent pictures of orderliness in language at each of two times more than 2 years apart across early childhood. Our findings show that the number of different word roots and mean length of utterance in children's conversation, mothers' checklist reports and ratings of child language, and assessed expression and comprehension by an experimenter shared significant amounts of variance at 20 months and again at 48 months. These findings are presumptive of an underlying latent variable for language in children at each age. The only exception to this common variance among potentially disparate measures and sources of language was for 48-month number of different word roots. It may be that the number of different words a child produces in an experimental storytelling task is not closely related to other indicators of child language in the preschool years because most typically developing children at that age command a substantial vocabulary. Preschool children who speak a smaller number of words in the storytelling task may do so because they are not sure what kind of story they are supposed to tell or because they cannot think of anything to say, not because they lack a minimum vocabulary to complete the task (as shown by the other language measures).

A multimeasure approach to child language has been applied productively in the past for predicting continued language delay (Olswang, Rodriguez, & Timler, 1998; Paul, 1996, 1997; Thal & Katich, 1996). Moreover, Wetherby, Allen, Cleary, Kublin, and Goldstein (2002) reported their strongest relations between parent reports and face-to face evaluations

of child language for a speech composite as opposed to single measures (see also Lyytinen, Poikkeus, Laakso, Eklund, & Lyytinen, 2001).

Stability in Child Language

The extent to which variation in language represents a consistent attribute in the child was addressed by examining stability (*qua* an individual-differences construct) across age. On the third question posed in this study, our longitudinal evaluation showed that the language latent variable at 20 months was highly stable with the language latent variable at 48 months. Heterotypic stability between different individual measures of child language may represent conservative (i.e., lower bound) estimates of stability of individual variation because of the variance introduced by differences in assessment measurements and procedures used at different times. Homotypic stability between the same individual measures of child language may represent liberal (i.e., upper bound) estimates of stability of individual variation because of shared source and method variance, practice effects, and the like. Aggregating across the individual stabilities in Table 1 produced a heterotypic estimate (average correlation of all dissimilar measures across time in Table 1) of .28, and a homotypic estimate (average correlation of the 3 consistent measures across time) of .25, whereas the model of child language latent variables produced a stability estimate of .84. Clearly, assessing stability in the common variance across different domains, measures, and sources of child language improves the stability estimate considerably.

There are several advantages of using a latent variable (Kline, 1998) that could contribute to the high level of stability we observed. First, the latent variables incorporate the perspectives of multiple reporters and domains of language development. Second, the latent variable is a purer measure of the underlying construct that associates only the shared variance among indicators. Any measurement error and unique variance that is not associated with the other variables that load on the latent variable is relegated to the error term. Third, the model accounts for systematic source and measurement variance (i.e., correlated error terms) that remove this “noise” from the latent variable. Fourth, the latent variables at each age are allowed to have different age-appropriate indicators and different loadings for the same indicators. Language ability at 20 and 48 months manifest differently. For example, successful communication at 20 months is largely indicated by comprehension, vocabulary, and the ability to combine words. At 48 months, successful communication is indicated by the abilities to relate complex and novel ideas verbally, to understand how words are related to one another, and to communicate in contextually and culturally appropriate ways. Using latent variables allows for the measurement of language ability to vary across time (as the construct does).

We would be remiss if we failed to point out that even such a strong stability result ($r = .84$) leaves substantial common variance unaccounted for ($\approx 30\%$). Here, theoretical perspective comes into play. Focusing on instability would lead to the singular but limited view that development is disorderly. Focusing alternatively on stability risks overlooking change in the developing organism. Next to stability, children still change in their status relative to one another over their 2nd and 3rd years in terms of their general language. They also manifestly change in terms of their group average language skills. The life-span perspective in developmental science specifies that human beings are open systems, and the plastic nature of psychological functioning ensures both stability and instability across the life course. Like many developmental processes, language is Janus-like with both stable and unstable, continuous and discontinuous aspects.

The large stability coefficient that emerged in this study might lead researchers and practitioners to conclude that language ability is set by 20 months of age. This is not necessarily the case. Ours was a sample of typically developing children without significant

language disabilities or delays. Children in this age range who require intervention have been shown to make significant progress in their language abilities in response to treatment (Bickford-Smith, Wijayatilake, & Woods, 2005; Kouri, 2005; Whitehurst et al., 1991). Furthermore, even in our normative sample, individual children increased and decreased in their relative standing by as much as 1.5 standard deviations across time. Overall, underlying language ability appears to be highly stable, and those children who are highly verbal at 20 months tend to be the same children who are highly verbal at 48 months, but the language abilities of individual children relative to their peers can still change across time. Moreover, child language in general increases dramatically in mean level across this same period.

Development is generally governed by genetic and biological factors in combination with environmental influences and experiences. Thus, stability of any individual characteristic could be attributable to factors in the individual, or stability might emerge through the individual's transactions with a consistent environment, or both. Cairns and Hood (1983) logically outlined several factors that feasibly support individual stability in the context of a developing system. First, specific and stable biological variables may contribute to stability. Such variables include genetic, hormonal, and morphological characteristics that might endure across developmental periods. Biological forces generally tend to reinforce homeostasis in the individual (Bornstein & Bornstein, 2008), and the role of the individual as an active agent in his or her own (stable) development (Lerner, 1982; Lerner & Busch-Rossnagel, 1981) means that individuals contribute to stability in their development by virtue of physical, socioemotional, cognitive, or behavioral characteristics that evoke consistent reactions in others. Thus, individual $\leftrightarrow$ context relational processes tilt to promote stability. For example, a verbally precocious child may elicit richer language input and interaction, thereby maintaining the child's linguistic advantage. In the psychological domain of the child, temperament, customarily defined as constitutionally based individual differences in positive and negative reactivity to stimulation, is hypothesized to be consistent over time and to influence social others (Rothbart & Bates, 2006; Wachs & Kohnstamm, 2001). Here we measured a facet of temperament, child social competence, and determined that stability of language obtained over and above its associations with language. Child stability also obtained separate from maternal verbal IQ. Notably, as Cairns and Hood (1983) went on to observe, "[I]t cannot be safely assumed that biological or genetic-based differences will persist, unmodified by social encounters or interchanges in which the individual engages" (p. 309).

A second factor that Cairns and Hood (1983) identified as potentially contributing to stability encompasses the social network in which development transpires. Here, we included family SES, maternal education and language, as well as children's childcare histories as representative of diverse milieus surrounding the child. All other factors being equal, similarities from one time to the next will be greatest when the social network in which development is enveloped remains constant. The extent to which stability reflects attributes of the individual or depends on circumstances is somewhat clarified by the multiple-covariate and childcare experience follow-up analyses we brought to bear on our original stability findings. Over half the mothers reported that they were working and that their children were in the care of other people. Stability in child language obtained at a high level separate from each and even in the face of non-trivial vicissitudes in children's caregiving environments.

In sum, with respect to our fourth question, in this study of the stability of child language we brought a number of covariates to bear to address several alternative endogenous and exogenous explanations, and we found that strong stability obtained in the child separate and apart from them all. The fact that stability of child language remained so high ($r = .79$) even

under multiple controls indicates that general language is a highly conserved and robust individual characteristic.

Gender

In answer to our fifth question, we found no systematic differences in stability between girls and boys; language in girls and boys is equally stable apart from reported mean differences in language level in girls and boys. The fact that stability of language in childhood transcends child gender further underscores the robustness of the basic result (see also Bornstein et al., 2004), and suggests that the mechanisms underlying language stability are similar for girls and boys.

Limitations

Despite its size and multivariate design, this study of child language is limited by the nature of the sample and the domains, measures, and sources it used. Children were all essentially the same ages at each of the two assessments; mothers who participated were primiparous at recruitment; and the families all English-speaking European American. Different patterns of stability could emerge in laterborn children, in children initially and terminally assessed at other ages, in families of different language and ethnic composition, or using features of child language different from those we measured. Note, too, the contexts for language samples at the two ages (e.g., play and bear story) differed. From one point of view, this procedural variation undercuts stability; from another, our findings are conservative.

Various language factors in the child (temperament), in mothers (reports child language), and in the environment (SES) might change little across our time span of data collection and carry stability. So, using multiple measures and multiple sources could overestimate the stability of child language. By controlling for family SES, mothers' verbal intelligence, education, speech, and social desirability bias, and child social competence we attempted to achieve a conservative estimate of child language stability that takes these several factors into consideration.

Conclusions

Stability assessment provides insight into individual variation and its development as well as the nature of the characteristic that is stable. Specifically, whether children maintain their rank order through time in terms of their language informs about individual children, but also contributes to understanding the nature and development of language. Insofar as language is ontogenetically stable, we know that children who do well or poorly at one time are likely to do well and poorly again later. So, in language a major predictor of developmental status at a given age is language at an earlier age, other things being equal.

The relatively high level of developmental stability observed in such a key achievement as general language prompts empirical, practical, and clinical implications. Empirically, our approach recommends the simultaneous investigation of multiple perspectives for other domains of child development in studies of stability. In that way, similarities and differences among all sources of information can be compared, and their implications adequately evaluated. Practically, whatever their developmental origins, meaningful individual differences in child language appear already established before the end of the second year. This finding suggests that, however and whichever experiential factors may be involved in motivating language development in very young children, maximizing their influence in the first 2 years of life may be advantageous for optimizing future child language. Clinically, our approach to estimate child language as a latent variable is not a quick tool for early diagnosis or screening; most clinicians do not have the benefit of a rich array of measures or

the technical support at hand to estimate latent variables. The present empirical findings contribute to clinical practice, however, insofar as we distinguish between language screening and the accuracy that the multiple domain, measure, and source approach yields to defining a latent variable. A screening instrument may be practically valuable, but the latent variable provides its knowledge base as well as a more fundamental understanding of human development. Finally, to be stable does not mean to be immutable or impervious to intervention. Change is an identifying characteristic of development, and children change in both their relative standing and mean level of language as they grow. Language ability is ultimately modifiable and plastic.

It is generally acknowledged that no single approach to measuring phenomena in child development is best, that no one representation of a developmental phenomenon predominates. Rather, assessment selection needs to be guided by tradition, tractability, goal, and convenience. Those who study developmental phenomena often advocate the wisdom of applying multiple assessments and employing converging operations of different strategies targeted to the same phenomenon. In this study, we used multiple ages, domains, measures, and sources to capture and evaluate individual variation and stability in child language. Employing observations, reports, and assessments together overcomes shortfalls associated with reliance on any single source. Multiple assessments also represent individuals better than do single assessments, and converging operations are necessary to evaluate whether the child's responses reveal an inferred capacity and that apparent performance is not simply an artifact of a given procedure. The agreement emergent among these diverse perspectives points to the value of evaluating stable individual variation in child language. One of the fundamental conceptual issues that has framed debates in theory and research across the history of developmental science has been the question of stability. General language capacity in young children is discontinuous in terms of its demonstrable growth in mean level through time; our study documents the very robust stability of general language capacity in young children in terms of individual order.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank D. Breakstone, O. M. Haynes, and J. Jager. This research was supported by the Intramural Research Program of the NIH, NICHD.

References

- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 101:213–232. [PubMed: 3562706]
- Appelbaum, MI.; McCall, RB. Design and analysis in developmental psychology. In: Mussen, P., editor. *Handbook of child psychology*. Vol 1: History, theory, and methods. New York: Wiley; 1983. p. 415–476.
- Aram D. Continuity in children's literacy achievements: A longitudinal perspective from kindergarten to school. *First Language*. 2005; 25:259–289.
- Bates, E.; Bretherton, I.; Snyder, L. From first words to grammar: Individual differences and dissociable mechanisms. Cambridge, England: Cambridge University Press; 1988.
- Bates E, Carnevale GF. New directions in research on language development. *Developmental Review*. 1993; 13:436–470.

- Bates E, Marchman V, Thal D, Fenson L, Dale P, Reznick JS, Reilly J, Hartung J. Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*. 1994; 21:85–124. [PubMed: 8006096]
- Bates JE, Bayles K. Objective and subjective components in mothers' perceptions of their children from age 6 months to 3 years. *Merrill-Palmer Quarterly*. 1984; 30:111–130.
- Bauer D, Goldfield B, Reznick J. Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics*. 2002; 23:313–335.
- Bayley N. Consistency and variability in the growth of intelligence from birth to eighteen years. *Journal of Genetic Psychology*. 1949; 75:165–196. [PubMed: 15403674]
- Beals DE, De-Temple JM. Home contributions to early language and literacy development. *National Reading Yearbook*. 1993; 42:207–215.
- Behar L, Stringfield S. A behavior rating scale for the preschool child. *Developmental Psychology*. 1974; 10:601–610.
- Bellinger D. Consistency in the pattern of change in mother's speech: Some discriminant analyses. *Journal of Child Language*. 1980; 7:469–487. [PubMed: 6969265]
- Belsky, J.; Jaffee, S. The multiple determinants of parenting. In: Cicchetti, D.; Cohen, D., editors. *Developmental Psychopathology*. Vol 3: Risk, disorder and adaptation. 2. New York: Wiley; 2006. p. 38-85.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]
- Bickford-Smith A, Wijayatilake L, Woods G. Evaluating the effectiveness of an early years language intervention. *Educational Psychology in Practice*. 2005; 21:161–173.
- Blake J, Quartaro G, Onorati S. Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*. 1993; 20:139–152. [PubMed: 8454679]
- Bornstein, MH.; Bornstein, L. Psychological stability. In: Darity, WA., Jr, editor. *International encyclopedia of social sciences*. 2. Vol. 8. Detroit, MI: Macmillan Reference; 2008. p. 74-75.
- Bornstein MH, Cote LR, Maital S, Painter K, Park S, Pascual L, Pêcheux M, Ruel J, Venuti P, Vyt A. Cross-linguistic analysis of vocabulary in toddlers: Spanish, Dutch, French, Hebrew, Italian, Korean, and English. *Child Development*. 2004; 75:1115–1139. [PubMed: 15260868]
- Bornstein, MH.; Gaughran, JM.; Homel, P. Infant temperament: Theory, tradition, critique, and new assessments. In: Izard, CE.; Read, PB., editors. *Measuring emotions in infants and children*. Vol 2: *Cambridge studies in social and emotional development*. New York: Cambridge University Press; 1986. p. 172-199.
- Bornstein MH, Hahn C-S, Haynes OM. Specific and general language performance across early childhood: Stability and gender considerations. *First Language*. 2004; 24:267–304.
- Bornstein MH, Haynes OM. Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*. 1998; 69:654–671. [PubMed: 9680678]
- Bornstein MH, Putnick DL, De Houwer AD. Child vocabulary across the second year: Stability and continuity for reporter comparisons and a cumulative score. *First Language*. 2006; 26:299–316.
- Bornstein MH, Painter KP, Park J. Naturalistic language sampling in typically developing children. *Journal of Child Language*. 2002; 29:687–699. [PubMed: 12109369]
- Bornstein MH, Tamis-LeMonda CS, Hahn C-S, Haynes OM. Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*. 2008; 44:867–874. [PubMed: 18473650]
- Bornstein MH, Tamis-LeMonda CS, Haynes OM. First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior & Development*. 1999; 22:65–85.
- Bowerman, M. Beyond communicative adequacy: From piecemeal knowledge to an integrated system in the child's acquisition of language. In: Nelson, K., editor. *Children's language*. Vol. 5. Hillsdale, NJ: Erlbaum; 1985. p. 369-398.
- Burgess S. The role of shared reading in the development of phonological awareness: A longitudinal study of upper class children. *Early Child Development and Care*. 1997; 127-128:191–198.

- Cairns, RB. Social development: The origins and plasticity of interchanges. San Francisco: Freeman; 1979.
- Cairns, RB.; Hood, KE. Continuity in social development: A comparative perspective on individual difference prediction. In: Baltes, PB.; Brim, OG., Jr, editors. Life-span development and behavior. New York: Academic Press; 1983. p. 301-358.
- Cameron-Faulkner T, Lieven E, Theakston A. What part of no do children not understand? A usage-based account of multiword negation. *Journal of Child Language*. 2007; 34:251–282. [PubMed: 17542158]
- Chaffee CA, Cunningham CE, Secord-Gilbert M, Elbard H, Richards J. Screening effectiveness of the Minnesota Child Development Inventory Expressive and Receptive Language Scales: Sensitivity, specificity, and predictive value. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*. 1990; 2:80–85.
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255.
- Clarke-Stewart, KA.; Allhusen, VD. Nonparental caregiving. In: Bornstein, MH., editor. *Handbook of Parenting*. Vol 3: Being and becoming a parent. 2. Mahwah, NJ: Erlbaum; 2002. p. 215-252.
- Clarke-Stewart KA, Vanderstoep LP, Killian GA. Analysis and replication of mother-child relations at two years of age. *Child Development*. 1979; 50:777–793.
- Cohen, J.; Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. 2. Hillsdale, NJ: Erlbaum; 1983.
- Colledge E, Bishop DVM, Koeppen-Schomerus G, Price TS, Happé FGE, Eley TC, Dale PS, Plomin R. The structure of language abilities at 4 years: A twin study. *Developmental Psychology*. 2002; 38:749–757. [PubMed: 12220052]
- Corkum V, Dunham P. The Communicative Development Inventory-WORDS Short Form as an index of language production. *Journal of Child Language*. 1996; 23:515–528.
- Cross, TG. Mothers' speech adjustments: The contribution of selected child listener variables. In: Snow, C.; Ferguson, C., editors. *Talking to children: Language input and acquisition*. Cambridge, England: Cambridge University Press; 1977.
- Cross, TG. Mothers' speech and its association with rate of linguistic development in young children. In: Waterson, N.; Snow, C., editors. *The development of communication*. New York: Wiley; 1978.
- Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*. 1960; 24:349–354. [PubMed: 13813058]
- Dale PS. The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Research*. 1991; 34:565–571. [PubMed: 2072681]
- Dale PS, Bates E, Reznick S, Morisset C. The validity of a parent report instrument of child language at twenty months. *Journal of Child Language*. 1989; 16:239–250. [PubMed: 2760125]
- Dallaire DH, Weinraub M. The stability of parenting behaviors over the first 6 years of life. *Early Childhood Research Quarterly*. 2005; 20:201–219.
- Darwin C. A biographical sketch of an infant. *Mind*. 1877; 2:286–294.
- De Houwer A, Bornstein MH, Leach DB. Assessing early communicative ability: A cross-reporter cumulative score for the MacArthur CDI. *Journal of Child Language*. 2004; 32:735–758. [PubMed: 16429709]
- Dickinson, DK.; Tabors, PO., editors. *Beginning language with literacy: Young children learning at home and school*. Baltimore, MD: Brookes; 2001.
- Dionne G, Dale PS, Boivin M, Plomin R. Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development*. 2003; 74:394–412. [PubMed: 12705562]
- Dromi, E. *Early lexical development*. Cambridge, England: Cambridge University Press; 1987.
- Dunn, LM.; Dunn, LM. *Peabody Picture Vocabulary Test -Revised Manual*. Circle Pines, MN: American Guidance Service; 1981.
- Eckerman CO, Davis CC, Didow SM. Toddlers' emerging ways of achieving social coordination with peers. *Child Development*. 1989; 60:440–453. [PubMed: 2924660]

- Feldman HM, Dollaghan CA, Campbell TF, Kurs-Lasky M, Janosky JE, Paradise JL. Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*. 2000; 71:310–322. [PubMed: 10834466]
- Fenson L, Bates E, Dale P, Goodman J, Reznick JS, Thal D. Measuring variability in early child language: Don't shoot the messenger. *Child Development*. 2000; 71:323–328. [PubMed: 10834467]
- Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ. Variability in early communicative development. *Monographs of the Society for Research in Child Development*. 1994; 59(Serial No. 242):1–173. [PubMed: 7845413]
- Fenson, L.; Dale, P.; Reznick, JS.; Thal, D.; Bates, E.; Hartung, J.; Pethick, S.; Reilly, J. The MacArthur Communicative Development Inventories: User's guide and technical manual. San Diego: Singular Publishing Group; 1993.
- Fenson, L.; Thal, D.; Bates, E. Normed values for the 'Early Language Inventory' and three associated parent report forms for language assessment (Technical Report). San Diego: San Diego State University; 1990.
- Gavin WJ, Giles L. Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research*. 1996; 39:1258–1262. [PubMed: 8959610]
- Guralnick M. Social interactions among preschoolers. *Exceptional Children*. 1980; 46:248–253.
- Gopnik A, Meltzoff AN. Relations between semantic and cognitive development in the one-word stage: The specificity hypothesis. *Child Development*. 1986; 57:1040–1053.
- Guttman, L. A new approach to factor analysis: The radex. In: Lazarsfeld, PF., editor. *Mathematical thinking in the social sciences*. Glencoe, IL: Free Press; 1954. p. 258-349.
- Hardy-Brown K, Plomin R, DeFries JC. Genetic and environmental influences on rate of communicative development in the first year of life. *Developmental Psychology*. 1981; 17:704–717.
- Hart B, Risley TR. American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*. 1992; 28:1096–1105.
- Hart, B.; Risley, TR. Meaningful differences in the everyday experiences of young American children. Baltimore, MD: Brookes; 1995.
- Hartmann, DP.; Pelzel, KE.; Abbott, C. Design, measurement, and analysis in developmental research. In: Bornstein, MH.; Lamb, ME., editors. *Developmental science: An advanced textbook*. 6. New York: Taylor & Francis; 2011. p. 107-195.
- Holden GW, Miller PC. Enduring and different: A meta-analysis of the similarity in parents' child rearing. *Psychological Bulletin*. 1999; 125:223–254. [PubMed: 10087937]
- Hollingshead, AB. The four-factor index of social status. Yale University; 1975. Unpublished manuscript
- Howes C, Matheson CC. Sequences in the development of competent play with peers: Social and social pretend play. *Developmental Psychology*. 1992; 28:961–974.
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:1–55.
- Huttenlocher J, Haight W, Bryk A, Seltzer M, Lyons T. Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*. 1991; 27:236–248.
- Irwin J, Carter A, Briggs-Gowan M. The social-emotional development of "late-talking" toddlers. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2002; 41:1324–1332. [PubMed: 12410075]
- Johnson, TP.; Shavitt, S.; Holbrook, AL. Survey response styles across cultures. In: Matsumoto, D.; van de Vijver, FJR., editors. *Cross-cultural research methods in psychology*. Cambridge, UK: Cambridge University Press; 2011. p. 130-178.
- Joreskog, K.; Sörbom, D. LISREL 8.72. Lincolnwood, IL: Scientific Software International; 2005.
- Kazdin AE. The diagnosis of childhood disorders: Assessment issues and strategies. *Behavioral Assessment*. 1988; 10:67–94.

- Kline, RB. Principles and practice of structural equation modeling. New York: Guilford Press; 1998.
- Kouri TA. Lexical training through modeling and elicitation procedures with late talkers who have specific language impairment and developmental delays. *Journal of Speech, Language, and Hearing Research*. 2005; 48:157–171.
- Le Normand M-T, Parrisé C, Cohen H. Lexical diversity and productivity in French preschoolers: Developmental, gender, and sociocultural factors. *Clinical Linguistics & Phonetics*. 2008; 22:47–58. [PubMed: 18092219]
- Leonard LB, Prutting CA, Perozzi J, Berkeley R. Non-standardized approaches to the assessment of language disorders. *Asha*. 1978; 20:371–379.
- Leopold, WF. Speech development of a bilingual child. Evanston, IL: Northwestern University Press; 1949.
- Lerner RM. Children and adolescents as producers of their own development. *Developmental Review*. 1982; 2:342–370.
- Lerner, RM.; Busch-Rossnagel, NA. Individuals as producers of their development: Conceptual and empirical bases. In: Lerner, RM.; Busch-Rossnagel, NA., editors. *Individuals as producers of their development: A life-span perspective*. New York: Academic Press; 1981. p. 1-36.
- Lyytinen P, Poikkeus A-M, Laakso M-L, Eklund K, Lyytinen H. Language development and symbolic play in children with and without familial risk for dyslexia. *Journal of Speech, Language, and Hearing Research*. 2001; 44:873–885.
- MacWhinney, B. The CHILDES Project: Tools for Analyzing Talk. Mahwah, NJ: Erlbaum; 2000.
- Marsh, HW.; Balla, JR.; Hau, K-T. An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In: Marcoulides, GA.; Schumacker, RE., editors. *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Erlbaum; 1996.
- Martlew M, Sorsby A. The precursors of writing: Graphic representation in preschool children. *Learning Instruction*. 1995; 5:1–19.
- McCall RB. Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development*. 1981; 52:1–12.
- McGregor KK, Capone NC. Genetic and environmental interactions in determining the early lexicon: Evidence from a set of tri-zygotic quadruplets. *Journal of Child Language*. 2004; 31:311–337. [PubMed: 15495843]
- McLaughlin B, White D, McDevitt T, Raskin R. Mothers' and fathers' speech to their young children: Similar or different? *Journal of Child Language*. 1983; 10:245–252. [PubMed: 6841496]
- Miles J, Shelvin M. A time and a place for incremental fit indices. *Personality and Individual Differences*. 2006; 42:869–874.
- Miller JF, Sedey AL, Miolo G. Validity of parent report measures of vocabulary development for children with Down syndrome. *Journal of Speech and Hearing Research*. 1995; 38:1037–1044. [PubMed: 8558873]
- Mullen M, Snidman N, Kagan J. Free-play behavior in inhibited and uninhibited children. *Infant Behavior and Development*. 1993; 16:383–389.
- National Institute of Child Health and Human Development Early Child Care Research Network (NICHD). The relation of child care to cognitive and language outcomes. *Child Development*. 2002; 71:960–980.
- O'Hanlon, L.; Thal, D. MacArthur CDI/Toddlers: Validation for language impaired children; Paper presented at the meeting of the American Speech-Language-Hearing Association; Atlanta. 1991 Nov.
- Olsewski P. Individual differences in preschool children's production of verbal fantasy play. *Merrill-Palmer Quarterly*. 1987; 33:69–86.
- Olswang L, Rodriguez B, Timler G. 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*. 1998; 7:23–32.
- Owens, REJ. *Language disorders: A functional approach to assessment and intervention*. 2. Boston: Allyn & Bacon; 1995.

- Pan BA, Rowe ML, Spier E, Tamis-LeMonda C. Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language*. 2004; 31:587–608. [PubMed: 15612391]
- Parke, RD. Fathers and families. In: Bornstein, MH., editor. *Handbook of parenting*. 2. Vol. 3. Mahwah, NJ: Erlbaum; 2002. p. 27-73.
- Paul R. Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology*. 1996; 5:5–21.
- Paul R. Understanding language delay: A reply to van Kleeck, Gillam, and Davis. *American Journal of Speech-Language Pathology*. 1997; 6:40–49.
- Peisner-Feinberg ES, Burchinal MR, Clifford RM, Culkin ML, Howes C, Kagan SL, Yazejian N. The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*. 2001; 72:1534–1553. [PubMed: 11699686]
- Pine JM, Lieven EV, Rowland C. Observational and checklist measures of vocabulary composition: What do they mean? *Journal of Child Language*. 1996; 23:573–590.
- Pinker, S. *The language instinct: How the mind creates language*. New York: Morrow; 2007.
- Reynell, JK.; Gruber, CP. *Reynell Developmental Language Scales U S Edition*. Los Angeles, CA: Western Psychological Services; 1990.
- Reznick JS. Visual preference as a test of infant word comprehension. *Applied Psycholinguistics*. 1990; 11:145–166.
- Reznick JS, Goldfield BA. Diary vs. representative checklist assessment of productive vocabulary. *Journal of Child Language*. 1994; 21:465–472. [PubMed: 7929690]
- Roberts BW, DeVecchio WF. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*. 2000; 126:3–25. [PubMed: 10668348]
- Rosenthal R. The “file drawer problem” and tolerance for null results. *Psychological Bulletin*. 1979; 86:638–641.
- Rothbart, MK.; Bates, JE. Temperament. In: Eisenberg, N.; Damon, W.; Lerner, RM., editors. *Handbook of child psychology*. Vol 3: Social, emotional, and personality development. 6. Hoboken, NJ: Wiley; 2006. p. 99-166.
- Saudino KJ, Dale PJ, Oliver B, Petrill SA, Richardson V, Rutter M, Simonoff E, Stevenson J, Plomin R. The validity of parent-based assessment of the cognitive abilities of 2-year-olds. *British Journal of Developmental Psychology*. 1998; 16:349–363.
- Sparrow, SS.; Balla, DA.; Cicchetti, DV. *Vineland Adaptive Behavior Scales Survey Form Manual (Interview Edition)*. Circle Pines, MN: American Guidance Service; 1984.
- Sroufe LA. The coherence of individual development: Early care, attachment, and subsequent developmental issues. *American Psychologist*. 1979; 34:834–841.
- Sternberg RJ, Grigorenko EL, Bundy DA. The predictive value of IQ. *Merrill-Palmer Quarterly*. 2001; 47:1–41.
- Storch SA, Whitehurst GJ. Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*. 2002; 38:934–947. [PubMed: 12428705]
- Tabachnick, BG.; Fidell, LS. *Using multivariate statistics*. 5. Boston: Allyn & Bacon; 2007.
- Thal, D.; Bates, E. Continuity and variation in early language development. In: Colombo, J.; Fagan, J., editors. *Individual differences in infancy*. Hillsdale, NJ: Erlbaum; 1990. p. 359-383.
- Thal, D.; Katich, J. Predicaments in early identification of specific language impairment: Does the early bird always catch the worm?. In: Cole, K.; Dale, P.; Thal, D., editors. *Assessment of communication and language*. Baltimore, MD: Brookes; 1996. p. 1-28.
- Thomas, A.; Chess, S.; Birch, H.; Hertzog, M.; Korn, S. *Behavioral individuality in childhood*. New York: New York University Press; 1963.
- Tilton-Weaver, LC.; Kakihara, F. United States of America. In: Arnett, JJ., editor. *International encyclopedia of adolescence*. New York: Routledge; 2008. p. 1061-1076.

- U.S. Census Bureau. Table 1. Total population by age, race and Hispanic or Latino origin for the United States: 2000. 2001. Retrieved August 17, 2011 from <http://www.census.gov/population/cen2000/phc-t9/tab01.pdf>
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–69.
- Van Hulle C, Goldsmith H, Lemery K. Genetic, environmental, and gender effects on individual differences in toddler expressive language. *Journal of Speech, Language, and Hearing Research*. 2004; 47:904–912.
- Verhulst, FC.; Koot, HM. *Child psychiatric epidemiology: Concepts, methods, findings*. Newbury Park: Sage; 1992.
- Wachs, TD.; Kohnstamm, GA. *Temperament in context*. Mahwah, NJ: Erlbaum; 2001.
- Wechsler, D. *Wechsler preschool and primary scale of intelligence--Revised: Manual*. San Antonio, TX: The Psychological Corporation; 1989.
- Weir, RH. *Language in the crib*. The Hague: Mouton; 1962.
- Wetherby AM, Allen L, Cleary J, Kublin K, Goldstein H. Validity and reliability of the Communication and Symbolic Behavior Scales Developmental Profile with very young children. *Journal of Speech, Language, and Hearing Research*. 2002; 45:1202–1218.
- Whitehurst GJ, Fischel JE, Lonigan CJ, Valdez-Menchaca MC, Arnold DS, Smith M. Treatment of early expressive language delay: If, when, and how. *Topics in Language Disorders*. 1991; 11:55–68.
- Winsler A, René de León J, Wallace BA, Carlton MP, Willson-Quayle A. Private speech in preschool children: Developmental stability and change, across-task consistency, and relations with classroom behavior. *Journal of Child Language*. 2003; 30:583–608. [PubMed: 14513469]
- Wohlwill, JF. *The study of behavioral development*. New York: Academic; 1973.
- Zigler E, Abelson WD, Seitz V. Motivational factors in the performance of economically disadvantaged children on the Peabody Picture Vocabulary Test. *Child Development*. 1973; 44:294–303.

Figure 1. Model of stability between the two general child language latent variables from 20 months to 48 months. ** $p < .01$ *** $p < .001$.

Figure 2.
Scatterplot of 20-month and 48-month language latent variable scores by child gender.

Table 1
Language Measures at 20 and 48 Months: Descriptive Statistics and Correlations

	1	2	3	4	5	6	7	8	9	10	11
<i>20 months</i>											
1. MLU	--										
2. DWR	.50***	--									
3. VABS Communication	.31***	.57***	--								
4. ELI	.37***	.64***	.74***	--							
5. Reynell Comprehension	.21*	.37***	.52***	.51***	--						
6. Reynell Expression	.40***	.76***	.69***	.79***	.54***	--					
<i>48 months</i>											
7. MLU	.15*	.22**	.29***	.31***	.26**	.31***	--				
8. DWR	.09	.13	.17*	.18*	.15*	.15*	.40***	--			
9. VABS Communication	.16*	.39***	.45***	.45***	.35***	.42***	.14	.01	--		
10. WPPSI-R Information	.05	.13	.35***	.33***	.41***	.33***	.24***	.14	.30***	--	
11. WPPSI-R Similarities	.06	.26***	.41***	.40***	.37***	.38***	.16*	.04	.29***	.55***	--
<i>M</i>	1.37	2.29	41.47	132.11	19.26	17.33	5.43	8.12	78.48	16.92	11.42
<i>SD</i>	.33	1.52	8.31	111.09	7.12	6.02	1.74	3.27	6.55	3.51	4.78
<i>Range</i>	0-2.88	0-8.40	25-65	0-466	5-42	1-36	0-13.35	0-15.88	62-102	1-25	1-23

Note: Shaded cells contain 20- to 48-month stability coefficients for scales from the same sources. Italicized stability coefficients are homotypic stabilities of the same measures across time.

MLU = mean length of utterance in morphemes. DWR = Different word roots. VABS = Vineland Adaptive Behavior Scales. ELI = Early Language Inventory. WPPSI-R = Wechsler Preschool and Primary Scales of Intelligence - Revised.

* $p < .05$.

** $p < .01$.

*** $p < .001$.