



# HHS Public Access

Author manuscript

*Dev Psychol.* Author manuscript; available in PMC 2017 May 01.

Published in final edited form as:

*Dev Psychol.* 2016 May ; 52(5): 704–716. doi:10.1037/dev0000111.

## Long-Term Stability of Core Language Skill in Children with Contrasting Language Skills

Marc H. Bornstein, Chun-Shin Hahn, 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

This four-wave longitudinal study evaluated stability of core language skill in 421 European American and African American children, half of whom were identified as low ( $n = 201$ ) and half of whom were average-to-high ( $n = 220$ ) in later language skill. Structural equation modeling supported loadings of multivariate age-appropriate multisource measures of child language on single latent variables of core language skill at 15 and 25 months and 5 and 11 years. Significant stability coefficients were obtained between language latent variables for children of low and average-to-high language skill, even accounting for child positive social interaction and nonverbal intelligence, maternal education and language, and family home environment. Prospects for children with different language skills and intervention implications are discussed.

### Keywords

language; stability; development; skill

---

Individual variation in the many different dimensions of language--expressive and receptive domains of phonology, morphology, semantics, syntax, and pragmatics--is a hallmark of children's language acquisition (Feldman et al., 2000; Fenson et al., 1994; Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015; Rowe, Raudenbush, & Goldin-Meadow, 2012). That pervasive individual variation raises the central developmental question addressed in the present study: Is the individual variation in language in children with different language skills stable across early development?

Most stability research in language acquisition has focused on small middle-class community samples and individual components of language retested over short intervals (see, e.g., Blake, Quartaro, & Onorati, 1993; Bornstein, Hahn, & Haynes, 2004; Burgess, 1997; Feldman et al., 2000; Gavin & Giles, 1996; Olszewski, 1987; Pine, Lieven, & Rowland, 1996; Sparrow, Balla, & Cicchetti, 1984; Winsler, René de León, Wallace, Carlton, & Willson-Quayle, 2003). Resulting estimates are usually moderate but vary somewhat with the domain, measure, method, source, and context as well as the ages of assessment and temporal interval between assessments. This univariate approach constitutes

a logical initial step in understanding the ontogeny of child language that precedes studies and analyses of multiple aspects of language investigated simultaneously over longer-terms in larger samples with more diverse characteristics (Bornstein, Jager, & Putnick, 2013). Here we report such an omnibus study of long-term (15 months to 11 years) stability of core language skill in relatively large ( $N = 421$ ) skill-stratified samples.

## Individual Differences and Stability in Child Language

Individual differences and stability in language obtain when some children display relatively higher levels of language at one point in time vis-à-vis their peers and continue to display higher levels at later points in time, while other children display consistently lower levels. Individual differences tell us about the distribution of language skill in children, and stability tells us about the nature and overall developmental course of the skill. Whether children maintain their relative standing in language skill through time informs not only about the meaningfulness of individual variation, but deepens understanding of the origins, nature, and ontogeny of language skill as well. Insofar as language skill is distributed and developmentally stable, children who are skilled or not in language at one time are likely so again later. Moreover, stable early characteristics shape later emerging ones. Young children who know more words in the first years tend to know more words later and are at a longer-term advantage because knowing more words facilitates learning to read, improves verbal comprehension, and eventuates in better oral language and academic skills (Marchman & Fernald, 2008; Morgan et al., 2015; Rowe et al., 2012). Given the importance of language development in general and for later cognitive and socioemotional functioning (e.g., Bornstein, Hahn, & Suwalsky, 2013; Morgan et al., 2015; Petersen et al., 2013; Sénéchal, Ouellette, & Rodney, 2006; Snow, Burns, & Griffin, 1998), understanding the stability of individual differences in language skill during the early years of life is of interest to parents, psychologists, and practitioners.

Because language is componential and changes dramatically in mean level with development, and no one approach to measurement is superior to all others under all situations, studying stability of individual differences poses unique challenges. However, diverse components of language covary (Bornstein & Haynes, 1998; Colledge et al., 2002; Dale, Harlaar, Hayiou-Thomas, & Plomin, 2010; Harlaar, Hayiou-Thomas, Dale, & Plomin, 2008; Johnson et al., 1999; Tomblin & Zhang, 2006; Trouton, Spinath, & Plomin, 2002). A primary methodological issue is to identify sensitive, reliable measures of language derivable from varying domains, methods, sources, and contexts that track child age appropriately. To meet this developmental challenge, we turned to latent variables (LV) as a solution because LVs can accommodate multiple perspectives on child language as they extract shared variance among language characteristics which may differ phenotypically at each age; therefore, LVs permit comparison across ages. Latent variables of language skill allow different age-appropriate indicators and different loadings for the same indicators across age. LVs therefore permit the measurement of a core language skill to vary (appropriately) across time but preserve comparability prerequisite to stability assessment. In two previous studies (Authors, 2012, 2014), we identified a core language skill in children and found it to be stable from 20 to 48 months and 20 months to 14 years in middle-SES European American samples. Here, we attempted to externally replicate and extend that

work in a new and different sample of low-SES children stratified by language skill level (Duncan, Engel, Claessens, & Dowsett, 2014). To do so, we adopted a latent variable approach to measure long-term 15-month to 11-year stability of individual differences in language skill in low-income European American and African American children.

### Stability by Language Skill

The main focus of group comparison studies of child language has heretofore fallen on mean-level group differences, a decidedly non-developmental issue. Although group differences are informative, they speak only indirectly to developmental questions of stability. On the one hand, strictly statistically speaking, stability of two groups (e.g., variation around their slopes) is independent of mean differences between the groups (e.g., their intercepts). On the other hand, a higher or lower skill at an earlier point in development could instigate evocative effects that maintain or discourage stability over time. Furthermore, as Rowe and colleagues (2012) argued, information gathered at a single point in time can mislead; they found that velocity and acceleration in vocabulary development predicted later vocabulary. Thus, information about the trajectory that children follow in language acquisition can help predict children's outcome.

A focus on trajectories underscores the value of understanding the path of development and possibly the underlying mechanisms of change that create different trajectories. Differential stability in lower and higher language skill groups could have implications for explicating underlying processes. For example, if children with lesser skill were more stable in their language over time, it might imply limits to their ability to improve lagging language skills. Perhaps some underlying dysfunction inflects language achievement or early language delay compounds over time. By contrast, similar stability across different levels of language skill could suggest that similar processes maintain core language skill at different skill levels. With no previous moderation studies of language stability by skill to draw from, but given stability of language generally, we hypothesized that language skill would be stable, and similarly stable, across children who varied in language skill.

### This Study

This study adds to the extant child and language development literatures by (1) assessing multiple language domains using multiple age-appropriate measures across multiple methods, sources, and contexts to (2) evaluate their empirical covariation and latent variables of language skill at each of several child ages and (3) the long-term stability between latent variables of language from the end of infancy to the start of adolescence in (4) relatively large but comparable low and average-to-high language skill groups (5) composed of low-SES European American and African American children. Although language stability has been studied previously, stability by language skill has not (to our knowledge), and stability of language in low-SES versus high-SES children has been studied (Fernald, Marchman, & Weisleder, 2013; Hart & Risley, 1995; Pan, Rowe, Singer, & Snow, 2005), but not stability of language skill in low-SES ethnically diverse samples.

We assessed the fit of two structural models to the data. The structural models assessed the common convergences of multiple indices on single latent variables of child core language skill at 15 and 25 months and 5 and 11 years as well as the stability between those latent variables in low and average-to-high skill groups and the stability among the language latent variables controlling multiple covariates (child positive social interaction and nonverbal intelligence, maternal education and language, and the family home environment). Stability is usually ascribed to temporal consistency of a characteristic in the individual. However, a richer understanding of developmental stability and its accurate attribution necessitate simultaneous examination of factors that influence it or confound its interpretation. Most studies of language stability do not take other endogenous or exogenous factors into consideration to rule out sources of stability alternative to child language per se and to assign stability more unambiguously to the child. Here, we assessed whether child core language skill is stable in itself, or if any of several third variables that covary with child language account for stability in child language skill.

## Method

### Participants and Skill Groups

The current study included European American and African American children of normal (2,500g) or low (1,500–2,499g) birth weights from English-speaking households who provided language data at the 11-year wave of the national Early Head Start Research and Evaluation (EHSRE) study. The EHSRE is a federal program initiated in 1995 and designed to evaluate the impact of Early Head Start programs on low-income families with infants and toddlers (Love et al., 2005; Paulsell, Kisker, Love, & Raikes, 2002). We defined two language skill groups by children's 11-year language performance (see 11 years under the Child Language Measures). Records and Tomblin (1994) investigated the diagnostic decision-making standards used by practicing clinicians to diagnose a child as language impaired or normal. They found that 1.0 *SD* below a language composite *z* score represented a cutoff point where a child would be diagnosed as language impaired by the majority of clinicians in their study. Here, low language skill was assigned to a child who scored  $-1.0$  *SD* or lower on both standard scores of the PPVT-III (scored  $< 85$ ; normed  $M = 100$ ,  $SD = 15$ ; Dunn & Dunn, 1997) and national norms of ECLS-K reading IRT scale score (scored  $< 119.57$ ; national normed average = 147.07,  $SD = 27.50$ ; Najarian, Pollack, Sorongon, & Hausken, 2009). Average-to-high language skill was assigned to a child who scored at or above mean standard scores of the PPVT-III (scored  $\geq 100$ ) and the national normed average of ECLS-K reading IRT scale score (scored  $\geq 147.07$ ). We chose to use 11-year language (the end-point of this longitudinal study) to define the groups because two standardized measures of language were available then. The starting point of the study (15 months) was not appropriate for defining language skill because language is less differentiated at this early point in development, and low-income mothers have been reported to overestimate their children's language skill in infancy (Feldman et al., 2000; Fenson et al., 2000). Many children have yet to produce their first words by 15 months, and this lack of production may not predict later language deficits. By defining the groups based on the final time point in the study, we are able to determine the stability of language based on performance at the end of elementary school, when children have been exposed to years of language experience and

instruction. This approach asks whether the relative ordering of elementary school children's core language skill can be predicted from their core language skill earlier in development, and whether the stability estimates are similar for children who end up with low language skill and average-to-high language skill.

Altogether, data from 422 children whose 11-year language scores met the cut-offs for low ( $n = 202$ ) vs. average-to-high ( $n = 220$ ) language skill categories were used. Two children were identified as multivariate outliers: One was an influential case that contributed disproportionately to parameter estimates and was removed; the other was a noninfluential outlier (parameter estimates did not differ with or without this case in the models) and was retained. Reported statistics are thus based on 421 children in the final study sample. Altogether 66.4% of children were firstborns, 47.3% were girls, and 27.1% were identified as at risk of adverse developmental outcomes (examples included congenital birth defects, severe chronic diseases, and parental substance abuse). On average, children were 15.08 months ( $SD = 1.73$ ,  $n = 370$ ), 25.17 months ( $SD = 1.95$ ,  $n = 366$ ), 5.27 years ( $SD = 0.34$ ,  $n = 327$ ), and 11.09 years ( $SD = 0.31$ ,  $n = 421$ ) of age, respectively, at the four assessment waves. (N.B. Children's ages reported in the current study do not differ from those reported in the EHSRE Parent Interview, 15.00 and 25.13 months and 5.20 and 10.88 years, respectively, but do differ from what the EHSRE conventionally calls the waves.) The 5-year data collection was targeted in the spring preceding children's kindergarten entry; the 11-year follow-up took place in the spring of children's sixth year of formal schooling, thus most children were attending 5<sup>th</sup> Grade at the 11-year wave. As different states and districts had different age criteria for kindergarten entry, children's ages at the 5-year and 11-year waves varied more widely than they did on birthday-related waves of the EHSRE. All language measures were adjusted for the wide age spans either by using age-normed scores, where applicable, or by statistical control techniques (see Analytic Plan). Mothers averaged 22.15 years ( $SD = 5.67$ ) at the child's birth, and 38.0% of the mothers were teens when their children were born. At the time of program enrollment, 39.3% of the mothers lived alone with their children, 62.9% had at least a General Education Development diploma or high school degree ( $M$  years of education = 9.33,  $SD = 1.99$ ), and 27.3% were employed. At enrollment, 85.8% of the families had incomes below the poverty line, 55.0% were welfare recipients, and 53.7% of participating families received Early Head Start services.

Table 1 shows family and child characteristics by skill groups. Several demographic differences emerged between the two child language skill groups: The majority of children from the low skill group were African Americans, whereas the majority of the children from the average-to-high skill group were European Americans. More children from the low skill group were identified at birth as having biological/medical risks (e.g., chromosomal abnormality, congenital birth defect, sensory impairment, HIV/AIDS, congenital heart disease, diabetes, or a severe chronic illness) than children from the average-to-high skill group. Mothers of children from the low skill group were younger when they gave birth than were mothers of children from the average-to-high skill group. More mothers of children of average-to-high language skill lived with their husbands, had completed some education beyond high school, and were employed, whereas more mothers of children of low language skill lived alone with their children, had not completed high school, and were still in school or a training program. More children from the low skill group came from families that were

welfare recipients. There were no differences in child birth order, gender, or ages between the two language skill groups. The proportion of families that had incomes below the poverty line or received Early Head Start services did not differ between the two skill groups.

To be eligible for enrollment in the EHSRE, families had to meet the program's income guidelines, agree to random assignment, and be expecting a child or have a child under 12 months of age. Random assignment yielded sociodemographically equivalent groups, as verified in the similar baseline characteristics of program and control group members (ACF, 2002a).

### Procedures and Measures

In addition to the Baseline Data (collected at the time of program application), the EHSRE data reported here derived from Parent Interviews and Child and Family Assessments (videorecorded in-home observations and direct child assessments by center-trained data collectors) at the 15- and 25-month and 5- and 11-year waves. The EHSRE technical report includes descriptions and psychometric information of measures administered (ACF, 2002a&b).

### Child Language Measures

**15 months**—During the home visit, primary caregivers (99.8% of the respondents were mothers) were asked to report on children's language using the short-form of the MacArthur Communicative Development Inventory-Words and Gestures (CDI-W&G; Fenson, et al., 1994, 2000). The CDI-W&G was designed for children 8 to 16 months as a measure of emerging receptive and expressive vocabulary and the use of communicative or symbolic gestures. First, respondents were asked to mark, from an 89-word list, the words their children understood or said, yielding separate indexes of the counts of words understood and words produced. The second part of the form asked respondents if the children had performed 18 communicative and symbolic gestures *often*, *sometimes*, or *not at all*. An Early Gestures score was computed as the number of times respondents used *often* or *sometimes* on the 18 questions. Total Early Gestures scores could range from 0 to 18. The EHSRE dataset did not include standardized scores for the CDI-W&G, and raw scores on Vocabulary Comprehension, Production, and Early Gestures scales were used in analysis.

**25 months**—During the home visit, primary caregivers (98.3% mothers) were asked to report on children's language using the short-form of the MacArthur Communicative Development Inventory-Words and Sentences (CDI-W&S; Fenson et al., 1994, 2000). The CDI-W&S was designed for children 16 to 30 months as a measure of expressive vocabulary and emerging grammatical complexity. First, respondents were asked about children's vocabulary production from a 100-word list. The production score was a count of words produced. Then, respondents were asked 36 questions about children's use of word combinations and closed-class morphemes. For each question, the experimenter read two phrases, one simple and one more complex, and asked respondents to choose one that more resembled their children's speech. The complexity score was re-coded by the EHSRE team so that "0" represented those children for whom respondents answered *not yet* on the

question “Has your child begun to combine words yet”; for all other cases, the complexity score was computed as the number of times respondents chose the more complex example plus 1. Total Sentence Complexity scores might range from 0 to 37. The EHSRE dataset did not include standardized scores for the CDI-W&S, and raw scores on Vocabulary Production and Sentence Complexity were used in analysis.

The Bayley Scales of Infant Development, Second Edition (BSID-II; Bayley, 1993) was administered. Boller (U.S. Department of Health and Human Services, 2001) conducted a factor analysis using the 42 BSID items appropriate for children ages 23 to 28 months on responses from 1,739 children participating in the EHSRE study and yielded a language factor made up of 12 items. Six of the 12 require the child to understand or produce lexical items, and the remaining six require syntactic and/or conversational skills. This language factor score was used in analysis.

**5 years**—The Peabody Picture Vocabulary Test, Third Edition (PPVT–III; Dunn & Dunn, 1997), which measures receptive vocabulary of spoken words, was administered. The experimenter presented a series of pictures to the children, and asked them to point to the picture that the word described. Raw scores were converted to age-adjusted, standardized scores based on the published norms. The standardized score was used in analysis. The Letter–Word Identification subtest in the Woodcock–Johnson III tests of cognitive academic competence (Woodcock, McGrew, & Mather, 2001) was also administered. On this subtest, children are asked to identify letters and read words out of context. The age normed standardized score was used in analysis.

**11 years**—The PPVT–III (Dunn & Dunn, 1997) was administered, and the age-standardized score was used in analysis. The fifth grade language and literacy assessments from the Early Childhood Longitudinal Study, Kindergarten Cohort study (ECLS-K) were also administered. Children’s proficiency in the following areas were evaluated: making inferences using cues that were directly stated with key words in text (literal inference); identifying clues used to make inferences (extrapolation), and using personal background knowledge combined with cues in a sentence to understand use of homonyms; demonstrating appreciation of author’s craft and making connections between a problem in the narrative and similar life problems (evaluation) and comprehension of biographical and expository text (evaluating nonfiction). The language/literacy item response theory (IRT) scale scores, which represented estimated numbers of items children would have answered correctly on the whole set of test items used in kindergarten through 11 years (probabilities of corrected answers summed over all items in the pool), were used in analysis (Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009).

## Covariates

Based on the extensive body of research on factors associated with child language, and to guard against threats to validity, we controlled for five kinds of covariates that might affect or underlie stability of children’s language development: children’s positive social interactions and nonverbal intelligence, their mothers’ education and language, and their family home environment.

**Child positive social interaction**—Ratings of child behavior during parent-child interaction at the 15-month home visit were obtained from videorecords of a semi-structured free-play task adapted from the NICHD Study of Early Child Care’s Three Box coding scales (NICHD Early Child Care Research Network, 1992, 1999). The three child scales rated children’s engagement of parent (extent to which child initiates and/or maintains interaction with parent); sustained attention with objects (degree of child’s involvement with toys in the three bags); and negativity toward parent (degree to which child shows anger or hostility toward parent), each on a 7-point scale with higher scores representing greater amounts. In the current sample, one principal component accounted for 60.2% of the variance in these 3 scales with un-rotated loadings ranging from .72 to .84. A mean score from ratings of child engagement of parent, sustained attention with objects, and positivity toward parent (reverse coded negativity scale) was computed to represent a measure of child positive social interaction with mother during free play.

**Child nonverbal intelligence**—At age 25 months, we used Bayley visual/spatial factor scores in the EHSRE dataset that were composed of 15 items from the BSID-II (Bayley, 1993) to represent a measure of child nonverbal intelligence. The sum of these 15 items (United States Department of Health and Human Services, 2001) was computed and used in analysis. At age 11 years, child nonverbal intelligence was represented by the Matrix Reasoning subtest of the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2003). The Matrix Reasoning subtest is one of three core subtests on the Perceptual Reasoning Index of the WISC-IV. The child was presented with a series of incomplete matrices, each of which is a series of abstract patterns and designs, and the child is directed to select the best from among several answer choices to complete the matrix. Standard scores were used in analysis.

**Maternal education and language**—Maternal education (in years) was obtained at study entry. At the 25-month visit, the Picture Vocabulary subtest of the Woodcock–Johnson Tests of Achievement (WJ; Woodcock & Johnson, 1990) was administered to mothers to assess their lexical knowledge, and the age-standardized score was used in analysis.

**Family home environment**—At the 15-month visit, the Parent Interview included the Infant/Toddler version of the Home Observation for Measurement of the Environment (HOME; Caldwell & Bradley, 2003) and additional items from National Longitudinal Survey of Youth that assess the quality of stimulation and support available to a child in the home environment. Information needed to score the inventory was obtained through a combination of interview and observation conducted in the home with the child’s parent while the child was present. Aspects of the assessed home environment included the extent of responsiveness of the parent to the child, support of cognitive, language, and literacy environment, parental lack of hostility/nonpunitive towards suboptimal behavior, and parental verbal skills. Higher total HOME scores indicate a more enriched home environment. Total HOME scores were used in analysis.



## Results

### Preliminary Analyses and Analytic Plan

First, variable distributions were examined for univariate normality (Tabachnick & Fidell, 2012) and transformations were applied to improve distributions. Transformed variables were used in analyses; for clarity, untransformed data are presented in reports of descriptive statistics. Because there was a range of child age at each assessment wave, we explored concurrent correlations of child age with test scores that were not age-standardized to determine if age adjustment was warranted. Age-adjusted scores were computed for all 15- and 25-month language measures (with the exception of Sentence Complexity which did not correlate with child age,  $r = .07$ ) and 11-year language/literacy scores and were used in structural equation models (SEM).

Language stability was evaluated using SEMs fit with Maximum Likelihood Functions (MLF) and following the mathematical models of Bentler and Weeks (1980) as implemented in EQS 6.1 (Bentler, 2006). Missing data points (17.3% of the total data were missing completely at random; Little's MCAR tests  $\chi^2(df = 624, N = 201) = 619.87$  and  $\chi^2(df = 714, N = 220) = 745.93$  for the low and average-to-high skill groups, respectively, both were *ns*) were handled in EQS using full information maximum likelihood (FIML) with a two-stage Expectation-Maximization (EM) estimation of the structured model and the MLF (Jamshidian & Bentler, 1999). Monte Carlo studies have demonstrated the general superiority of the structured-model EM method implemented in EQS 6.1 compared to other techniques to recover missing data, especially in MCAR normal or slightly nonnormal data (Gold & Bentler, 2000; Yuan & Bentler, 2000). In the course of fitting SEMs, we evaluated Mardia (1970) coefficients of multivariate kurtosis and the cases that contributed disproportionately to parameter estimates. No significant problems of nonnormality or influential cases emerged.

The fit of SEMs was assessed using the robust Yuan-Bentler (Y-B) scaled  $\chi^2$  statistic, robust comparative fit index (CFI), standardized root mean squared residual (SRMR; Browne & Cudeck, 1993), and root mean square error of approximation (RMSEA). Cutoff values  $\approx .95$ ,  $\approx .08$ , and  $\approx .06$  for CFI, SRMR and RMSEA, respectively, are indicative of a relatively good fit between the hypothesized model and observed data (Hu & Bentler, 1999). We gave greater weight to the alternative fit indices than to  $\chi^2$  because the  $\chi^2$  value is sensitive to sample size (Cheung & Rensvold, 2002). Standardized path coefficients are presented.

To obtain stability estimates across ages, an *a priori* model in which language indicators at 15 and 25 months and 5 and 11 years loaded on their respective latent variables, and each language latent variable was a function of the immediately preceding language variable, was hypothesized and tested in the total sample first and then separately on samples of low and average-to-high skill groups. After fitting the stability models, we re-evaluated the stability estimates controlling for child positive social interaction and nonverbal intelligence, maternal education and language, and family home environment. For both stability and covariate models, we performed multiple-group analysis to assess differences in stability estimates between children of low and of average-to-high language skill. We report the difference in  $\chi^2$  statistics and CFI values (Cheung & Rensvold, 2002) for nested models

between the unconstrained and constrained models (Vandenberg & Lance, 2000). If the  $\chi^2$  between the unconstrained and constrained models was nonsignificant ( $p > .05$ ) and the CFI  $\geq .91$  (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000), the model was deemed to fit equally well in both skill groups.

Sample size consideration prohibited further investigation of child gender and program enrollment as potential moderators within the framework of differential language skill groups. However, we found that child gender and program enrollment status were equally distributed across the 2 language skill groups:  $\chi^2 (df=1, N=421) = 0.04, p = .85$  for child gender (% female was 46.8 in low skill group, and 47.7 in average-to-high skill group), and  $\chi^2 (df=1, N=421) = 0.17, p = .70$  for program enrollment status (% program enrollment was 54.7 in low skill group, and 52.7 in average-to-high skill group).

### Language Stability from 15 Months to 11 Years

**Descriptive statistics**—Table 2 shows the *M*s, *SD*s, and ranges of language measures by skill groups. The *SD* and ranges on all measures indicate considerable variation in child language, as is commonly found in the literature. On average, children in the low skill group scored in the  $-1.04$  *SD* to  $-1.52$  *SD* range on standardized tests at earlier ages, and children in the average-to-high skill group scored within the  $M \pm 0.5$  *SD* range on standardized tests at ages younger than 11 years.

Also shown in Table 2 are effect sizes for group mean comparisons on language measures. Children who were of average-to-high language skill at 11 years scored higher on all language measures collected across previous data collection waves than did children of low language skill at 11 years, with only one exception: the 15-month CDI Early Gestures did not differ between the two groups. Twenty-six out of the 201 children in the low skill group (13%) scored more than 2 *SD*s below the normed averages on both PPVT-III and ECLS-K at 11 years. Scoring more than 2 *SD*s below the normed averages (i.e., below the third percentile) is one of the specific requirements in the International Classification of Diseases (ICD-10; World Health Organization 1993) code to receive a diagnosis of language disorder.

**Language stability**—The *a priori* language stability model fit the data for the total sample: robust Y-B scaled  $\chi^2(32) = 74.83, p < .001$ , Robust CFI = .99, SRMR = .07, RMSEA = .04, 90% CI = [.01, .05]. All indicators of child language loaded significantly (all  $p$ s  $< .001$ ) on their factors at each age, which indicated that various measures of language formed stable, single factors of core language skill at each age. Language stability was large (all  $p$ s  $< .001$ ) between each succeeding time point across the ages from 15 months to 11 years: .64 from 15 to 25 months, .62 from 25 months to 5 years, and .90 from 5 to 11 years.

The *a priori* language stability models also fit the data of low and average-to-high skill groups: robust Y-B scaled  $\chi^2(32) = 53.50, p < .01$ , Robust CFI = 1.00, SRMR = .07, RMSEA = .00, 90% CI = [.00, .05] for the low skill group, and robust Y-B scaled  $\chi^2(32) = 45.59, p = .06$ , Robust CFI = 1.00, SRMR = .07, RMSEA = .00, 90% CI = [.00, .04] for the average-to-high skill group. Figure 1 presents the standardized solution of these stability models. For both low and average-to-high skill groups, all indicators of child language loaded significantly on their factors at each age. Language stability was large between each

succeeding time point across the ages from 15 months to 11 years for both language skill groups (with the exception of stability between 25 months and 5 years which was of medium effect size in the average-to-high skill group). Table 2 displays the pair-wise variance covariance matrix of the 10 language measures by skill groups.

**Group comparison of stability estimates**—Testing differences on stability estimates between the low and average-to-high groups is a 3-step process. First, we established that the same model form applies to the two groups (i.e., configural invariance: no constraints across groups). Next, we demonstrated that the four language latent variables in the SEM were similar constructs for children of low and average-to-high language skill (i.e., metric invariance: constraining the factor loadings across groups). Finally, if metric invariance was confirmed, we then constrained the stability estimates across groups (Bollen, 1989; Steenkamp & Baumgartner, 1998; Taris, Bok, & Meijer, 1998).

A preliminary configural invariance multiple-group model, in which no parameter estimates were constrained to be equal between children of low and average-to-high skill, fit the data, robust Y-B scaled  $\chi^2(64) = 99.36, p < .01$ , Robust CFI = 1.00, SRMR = .07, RMSEA = .00, 90% CI = [.00, .04], suggesting that the same “model form” (Bollen, 1989) could be applied to both skill groups and more restrictive tests were appropriate.

The metric invariance model, testing the meaning of the latent factors across groups, suggested that one or more factor loadings differed between the groups,  $\chi^2(6) = 16.80, p < .05$ , CFI = .00. The Lagrange Multiplier test was significant for the factor loading of CDI Production on the 15-month language factor,  $\chi^2(1) = 6.43, p < .05$ . When this constraint was released, the difference in  $\chi^2$  statistics between the constrained model with invariance constraints on factor loadings and the unconstrained model with no invariance constraints was no longer significant,  $\chi^2(5) = 7.63, ns$ , CFI = .00. Full metric invariance was established for the 25-month, 5-year, and 11-year language factors, whereas partial metric invariance (Byrne, Shavelson, & Muthén, 1989) was indicated for the 15-month language factor. These results suggested that all four language factors were similar constructs for children in the two skill groups.

The final multiple-group analysis, conducted to test any difference in stability estimates, suggested that one or more stability coefficients differed between the groups,  $\chi^2(3) = 8.46, p = .04$ , CFI = .00. One Lagrange Multiplier univariate test statistic was significant,  $\chi^2(1) = 4.75$ , for the equality constraint on stability estimates from 25 months to 5 years. Releasing this constraint, the difference in  $\chi^2$  statistics was no longer significant,  $\chi^2(2) = 2.96, ns$ , suggesting that stability coefficients did not differ between children of low and average-to-high language skill from 15 to 25 months or from 5 to 11 years, but stability was higher in the low skill group than the average-to-high skill group from 25 months to 5 years (Figure 1).

A separate follow-up analysis assessed stability models and multiple-sample analysis on datasets removing 114 children from the study sample (60 from the low and 54 from the average-to-high skill groups) who were identified as being at risk of adverse developmental outcomes at birth (i.e., children who had one or multiple established biomedical and/or

environmental risks; see Table 1 note<sup>b</sup>). Stability estimates for children not identified as being at risk of adverse developmental outcomes were .68 from 15 to 25 months, .48 from 25 months to 5 years, and .80 from 5 to 11 years for the low skill group ( $n = 141$ ) and .69 from 15 to 25 months, .21 from 25 months to 5 years, and .68 from 5 to 11 years for the average-to-high skill group ( $n = 166$ ); all parameter estimates remained significant. Removing children identified as being at risk for adverse developmental outcomes from the study sample resulted in no differences in the stability estimates between the low and average-to-high language skill groups,  $\chi^2(3) = 6.95$ , *ns*, CFI = .00.

### Language Stability Controlling for Covariates

**Language stability with controls**—As a check against threats to validity, we re-evaluated the stability models controlling for child positive social interaction and nonverbal intelligence, maternal education and language, and the family home environment. As shown in Table 1, children who were of average-to-high language skill at 11 years had higher scores on positive social interaction and nonverbal intelligence and had mothers of higher education who also scored higher on the Picture Vocabulary test and provided a better home environment in support of their children's language development than did children of low language skill. Direct paths from child positive interaction, maternal education and language, and the HOME total score to all four core language variables, and from 25-month nonverbal intelligence to 25-month and 5- and 11-year language, from 11-year nonverbal intelligence to 11-year language, as well as a stability path between the two child nonverbal intelligence measures, and covariances among the covariates were added to the stability model. Table 3 shows zero-order correlations between these covariates and the language measures by skill group.

The covariate models fit the data: robust Y-B scaled  $\chi^2(81) = 100.48$ , *ns*, Robust CFI = 1.00, SRMR = .07, RMSEA = .00, for the low skill group and robust Y-B scaled  $\chi^2(80) = 113.91$ ,  $p < .01$ , Robust CFI = 1.00, SRMR = .08, RMSEA = .00, for the average-to-high skill group. Controlling for child positive social interaction and nonverbal intelligence, maternal education and language, and family home environment, the stability estimates were still large between successive waves in the low skill group: .61 from 15 to 25 months, .65 from 25 months to 5 years, and .78 from 5 to 11 years. In the average-to-high skill group, the stability estimates were large from 15 to 25 months and from 5 to 11 years, .70 and .49, respectively; but stability was small, .17, from 25 months to 5 years.

**Group comparison of stability estimates with controls**—Because configural and metric invariance were already confirmed in the previous multiple-sample analysis across skill groups, we retained the constrained metric invariance model and examined model fit with constrained stability coefficients across groups. The difference in  $\chi^2$  was significant,  $\chi^2(3) = 13.93$ ,  $p < .01$ , CFI = .00, suggesting that one or more stability coefficients differed between the groups. Controlling for child positive social interaction and nonverbal intelligence, maternal education and language, and family home environment, language stability estimates were greater in children of low language skill from 15 to 25 months,  $\chi^2(1) = 5.19$ ,  $p < .05$ , and from 25 months to 5 years,  $\chi^2(1) = 7.45$ ,  $p < .01$ , than were those in children of average-to-high skill. Stability did not differ between the two groups from 5 to

11 years controlling for child positive social interaction and nonverbal intelligence, maternal education and language, and family home environment,  $\chi^2(1) = 0.15$ , *ns*.

## Discussion

This study addressed two underresearched but intertwined issues directly related to child development and language, *viz.* the long-term stability of language and moderation of the long-term stability of language by language skill. We first identified a core language skill from several measures of language taken at 15 and 25 months and 5 and 11 years. Next, we estimated the comparative stability of that core language skill in children with low and average-to-high language skills. Last, we tested whether a diverse set of controls for background characteristics accounted for stability in the child core language skill over the first 11 years of life. Clear evidence emerged for individual variation, long-term stability, and relative robustness across childhood in the two contrasting language skill samples, even controlling for child positive social interaction and nonverbal intelligence, maternal education and language, and the family home environment.

These results prompt several considerations. First, a corollary of the prevailing multidimensional and componential conceptualization of language might be that phenotypically distinct language domains are independent of one another. Because no single approach to measuring development of a characteristic in the child is best, no one representation of any characteristic predominates. Here, we found that diverse concurrent indices of language based on different language domains, measures, methods, sources, and contexts, each of which showed individual variation, were moderately to strongly positively associated, even at different ages, and on this basis we computed single latent variables of a child core language skill at each age. At each age tested, significant amounts of variance were accounted for by each latent variable, results that add to the validity of the stability model. Latent variables rely on empirical covariation among their indicators and capitalize on the unique variance that those indicators share. Latent variables are purer representations of the underlying construct because they remove unshared variance of the indicators. That is, latent variables relegate variance uniquely associated with rater bias, random measurement error, or specific error (error variance arising from some characteristic unique to a particular indicator that is not accounted for by the factor) to an error term (Kline, 2015). Our latent variable strategy has the value that it overcomes shortfalls associated with reliance on any individual measure. Converging operations are necessary to evaluate whether the child's behaviors reveal an inferred capacity and to demonstrate that apparent performance is not simply an artifact of a given procedure. Contemporary developmental science advocates applying multiple assessments and employing converging operations to target a given construct. Different approaches to studying child language -- seeking out those people closest to children to report about them or testing children directly -- are each valid, but each suffers certain limitations with implications for assessing stability. Multiple assessments, such as used here, take more aspects of the child into account and represent the child better than do single assessments. Overall, these results point to the value of measuring diverse components of language in different ways and contexts using multiple informants to obtain a picture of children's core language skill at different ages.

Second, it is noteworthy that the results of the present study replicate and extend (Bonett, 2012; Duncan et al., 2014) those of previous studies with single language measures taken over shorter periods of time (reviewed in the Introduction) as well as those in more recent reports taken with multiple measures over longer durations (Bornstein, Hahn, Putnick, & Suwalsky, 2014). Extending those results to low and high skill language groups is novel. The fact that language stability coefficients across studies are comparable is heartening because different assessment contents, procedures, and times normally contribute to different stability estimates of child language.

These stability coefficients seem to suggest that children's level of core language skill gels as early as 15 months. This is not necessarily the case. Focusing solely on stability runs the risk of overlooking or minimizing complementary changes in mean level. Stability of individual differences is mathematically independent of group mean-level consistency or change, so all children in a group may increase in their language (as they normally do) even as they remain stable relative to one another. Moreover, the language abilities of individual children relative to their peers still change across time. Even large relative stability leaves substantial common variance unaccounted for. For example, our largest stability estimate (.85 in the low skill group) from 5 to 11 years leaves nearly one-third of the variance ( $1 - .85^2$ ) in the 11-year core language skill unexplained by the 5-year core language skill. To be stable does not mean to be immutable to change, experience, or intervention, and language is ultimately modifiable and plastic. Children change in their mean level as they grow just as they do in their relative standing. Development in language acquisition balances the advantages of stability with the adaptive value of early susceptibility to modification and long-term growth. Maximizing the influence of factors that motivate language development early in life may therefore be advantageous for optimizing child language development.

Third, although our study design does not directly address the question, our results implicitly ask what the sources of stability of individual differences in core language skill might be. As is commonly acknowledged, development in children is governed by genetic and biological factors in combination with environmental influences and experiences (Overton, 2015). It is likely that variation and stability of core language skill are ascribable to individual factors (genetics, gender, sociability, and maturation, for example; Dale et al., 2010) and variation and stability emerge and maintain through the child's transactions with a stable environment that supports language (as in maternal language addressed to the child; Bornstein, Tamis-LeMonda, & Haynes, 1999). Thus, unsurprisingly, Rowe, Jacobson, and Van den Oord (1999) found that genetic *and* environmental factors explained similar proportions of variance in adolescent sibling pairs' verbal IQ. Our low skill children had somewhat stronger stability from 25 months to 5 years (or from 15 months to 5 years in the covariate controlled model). Perhaps the low skill group is more influenced by individual than by environmental factors based on their higher stability at the younger ages, the lack (or smaller) relations with maternal age, education, and home environment (Table 3), and the fact that skill group differences in stability attenuate when biological risk is removed (even if strong stability still obtained in children who were not at risk). Furthermore, the extent to which stability of core language skill in children reflects aspects of the child as well as circumstances that envelop the child is somewhat clarified by the multiple-covariate follow-up analyses. Significant long-term stability obtained in both skill groups separate and apart from multiple

endogenous and exogenous covariates, including child positive social interaction and nonverbal intelligence, maternal education and language, and the family home environment.

Fourth and theoretically, the strong stability in lower and higher language skill groups has implications for understanding underlying processes. Similar stabilities across different levels of language skill (as between 5 and 11 years) suggest that similar processes maintain core language skill despite different skill levels. Had children with lesser skill been more stable in their language between all time points, it might imply different underlying processes and a possible inherent limit to their ability to improve lagging language skills.

Finally, the relative ontogenetic stability we observed in child language may have clinical implications. Our low skill group scored on average  $>1$  *SD* below the mean on standardized tests at the early ages. In decision-making standards used by practicing clinicians to diagnose children as language impaired,  $1.0$  *SD* below the language composite *z* score represents a cutoff point (Records & Tomblin, 1994). Although our approach to child language assessment (via latent variables and defining the skills groups at 11 years) does not readily lend itself to early screening or diagnosis, and, practically speaking, clinicians do not normally have access to the measures, sample sizes, or technical support required to estimate latent variables, the present findings can inform clinical practice. This kind of multidimensional approach to child language has been applied productively in the past to predicting language delay (Olswang, Rodriguez, & Timler, 1998; Thal & Katich, 1996). Early screening, monitoring, and intervention might be improved if guided by findings from large-scale studies that identify factors associated with early and meaningful differences in children's vocabularies. However, we found that 15 months was too early to form reliable skill groups that predicted later outcomes. The scale means of the low and average-to-high skill groups (that were formed based on 11-year scores) differed at all time points except 15 months. Still, there was predictive validity of later skill from this early time point. Moreover, following the logic above, the  $.66$  coefficient between 25 months and 5 years for low-skill children implies that 56% of the variance in 5-year core language skill was not explained by 25-month core language skill. By contrast, the  $.85$  stability coefficient from 5 to 11 years implies that changing core language skill later in development might be more challenging. Targeting multiple aspects of the language environment early in life may represent a fruitful means to supporting the development of young children with low language skill. For example, vocabulary is malleable and contributes to increased academic and behavioral functioning, and so might be targeted in early interventions that manipulate aspects of the environment and so close the vocabulary gap between children in low- versus high-SES families (e.g., Dickinson, Golinkoff, & Hirsh-Pasek, 2010; Perfetti & Stafura, 2014).

### Strengths and Limitations

This study included varied measures of language skill -- communicative gestures, vocabulary comprehension and production, sentence complexity, syntactics, conversation, literary inference, extrapolation, and evaluation, and comprehension of homonyms. However, the measures available at the 5-year assessment only directly measured vocabulary. The study would have benefitted from even more varied measures of age-appropriate language skill. We defined the groups at 11 years, and so language stability may

be stronger at older ages. Unfortunately, we were not able to identify mechanism(s) underlying stability, although we found that stability obtained separate from some prominent child, maternal, and family factors. Future empirical steps ought to be designed to pinpoint such mechanisms as well as explore the role of stability in later cognitive achievement and socioemotional adjustment.

## Conclusions

Stability and change are prominent issues in the history of developmental science. The variance that is shared in multiple indices of language at several ages is stable across time in children of low as well as average-to-high language skill. This study adds to the developmental and language literatures by showing that children with low and average-to-high language skills share a core language skill at different ages, one that is composed of different language domains and measures, collected by different methods from different sources in different contexts, that that core language skill is distributed in children at different ages, and that some long-term stability of that core language skill begins very early in life, extends to early adolescence, and transcends methodological variance and multiple conservative controls.

## Acknowledgments

This research was supported by the Intramural Research Program of the NIH, NICHD.

## References

- Administration for Children and Families. Making a difference in the lives of children and families: The impacts of Early Head Start programs on infants and toddlers and their families. Washington, DC: U. S. Department of Health and Human Services; 2002a.
- Administration for Children and Families. Pathways to quality and full implementation in Early Head Start Programs. Washington, DC: U. S. Department of Health and Human Services; 2002b.
- Bayley, N. Bayley Scales of Infant Development, Second Edition, Manual. New York: The Psychological Corporation, Harcourt Brace & Company; 1993.
- Bentler, PM. EQS 6 structural equations program manual. Encino, CA: Multivariate Software; 2006.
- Bentler PM, Weeks DG. Linear structural equations with latent variables. *Psychometrika*. 1980; 45:289–308.10.1007/BF02293905
- Blake J, Quartaro G, Onorati S. Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*. 1993; 20:139–152.10.1017/S0305000900009168 [PubMed: 8454679]
- Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989.
- Bonett DG. Replication-extension studies. *Current Directions in Psychological Science*. 2012; 21:409–412.10.1177/0963721412459512
- Bornstein MH, Hahn CS, Haynes OM. Specific and general language performance across early childhood: Stability and gender considerations. *First Language*. 2004; 24:267–304.10.1177/0142723704045681
- Bornstein MH, Hahn CS, Putnick DL, Suwalsky JTD. Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development*. 2014; 85:1346–1356.10.1111/cdev.12192 [PubMed: 25165797]
- Bornstein MH, Hahn C-S, Suwalsky JTD. Language and behavioral adjustment: Developmental pathways from childhood to adolescence. *Development and Psychopathology*. 2013; 25:857–878.10.1017/S0954579413000217 [PubMed: 23880396]

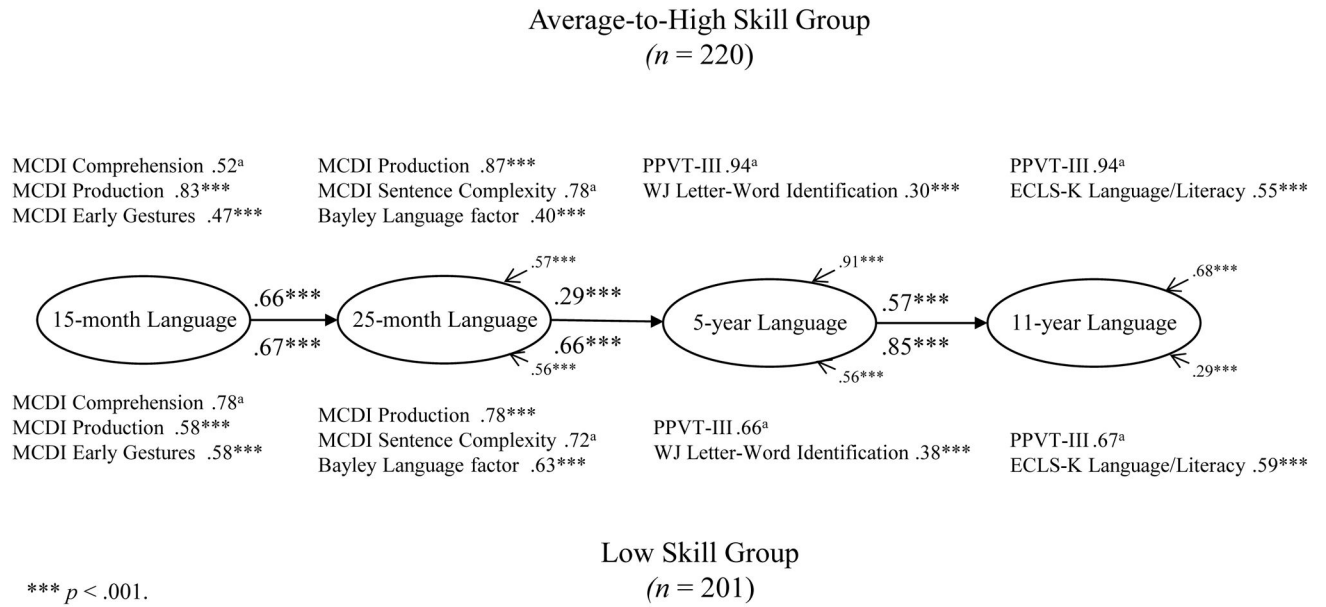


- Bornstein MH, Haynes OM. Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*. 1998; 69:654–671.10.2307/1132196 [PubMed: 9680678]
- Bornstein MH, Jager J, Putnick DL. Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review*. 2013; 33:357–370.10.1016/j.dr.2013.08.003 [PubMed: 25580049]
- 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 and Development*. 1999; 22:65–85.10.1016/S0163-6383(99)80006-X
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Beverly Hills, CA: Sage; 1993. p. 136-162.
- 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; 127–128:191–198.10.1080/0300443971270116
- Byrne BM, Shavelson RJ, Muthén B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*. 1989; 105:456–466.10.1037/0033-2909.105.3.456
- Caldwell, BM.; Bradley, RH. *Administration manual: Home observation for measurement of the environment*. Little Rock, AR: University of Arkansas at Little Rock; 2003.
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255.10.1207/S15328007SEM0902\_5
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Erlbaum; 1988.
- 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; 38:749–757.10.1037/0012-1649.38.5.749 [PubMed: 12220052]
- Dale PS, Harlaar N, Hayiou-Thomas ME, Plomin R. The etiology of diverse receptive language skills at 12 years. *Journal of Speech, Language, and Hearing Research*. 2010; 53:982–992.10.1044/1092-4388(2009)09-0108
- Dickinson DK, Golinkoff RM, Hirsh-Pasek K. Speaking out for language: Why language is central to reading development. *Educational Researcher*. 2010; 39:305–310.10.3102/0013189X10370204
- Duncan GJ, Engel M, Claessens A, Dowsett CJ. Replication and robustness in developmental research. *Developmental Psychology*. 2014; 50:2417–2425.10.1037/a0037996 [PubMed: 25243330]
- Dunn, LM.; Dunn, LM. *Peabody Picture Vocabulary Test*. 3. Circle Pines, MN: American Guidance Service; 1997.
- 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; 71:310–322.10.1111/1467-8624.00146 [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; 71:323–328.10.1111/1467-8624.00147 [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; 59:1–173. Serial No. 242. 10.2307/1166093 [PubMed: 7845413]
- Fernald A, Marchman VA, Weisleder A. SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*. 2013; 16:234–248.10.1111/desc.12019 [PubMed: 23432833]
- Gavin WJ, Giles L. Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research*. 1996; 39:1258–1262.10.1044/jshr.3906.1258 [PubMed: 8959610]
- Gold MS, Bentler PM. Treatment of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*. 2000; 7:319–355.10.1207/S15328007SEM0703\_1

- Harlaar N, Hayiou-Thomas ME, Dale PS, Plomin R. Why do preschool language abilities correlate with later reading? A twin study. *Journal of Speech, Language, and Hearing Research*. 2008; 51:688–705.10.1044/1092-4388
- Hart, B.; Risley, TR. Meaningful differences in the everyday experience of young American children. Baltimore, MD: Brookes; 1995.
- 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; 6:1–55.10.1080/10705519909540118
- Jamshidian M, Bentler PM. ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics*. 1999; 24:21–24.10.2307/1165260
- Johnson CJ, Beitchman JH, Young A, Escobar M, Atkinson L, Wilson B, Brownlie EB, Douglas L, Taback N, Lam I, Wang M. Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research*. 1999; 42:744–760.10.1044/jslhr.4203.744
- Kline, RB. Principles and practice of structural equation modeling. 3. New York: Guilford Press; 2015.
- Love JM, Kisker EE, Ross C, Raikes H, Constantine J, Boller K, Brooks-Gunn J, Chazan-Cohen R, Tarullo LB, Brady-Smith C, Fuligni AS, Schochet PZ, Paulsell D, Vogel C. The effectiveness of Early Head Start for 3-year-old children and their parents: lessons for policy and programs. *Developmental Psychology*. 2005; 41:885–901.10.1037/0012-1649.41.6.885 [PubMed: 16351335]
- Marchman VA, Fernald A. Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*. 2008; 11:F9–F16.10.1111/j.1467-7687.2008.00671.x [PubMed: 18466367]
- Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika*. 1970; 57:519–530.10.1093/biomet/57.3.519
- Morgan PL, Farkas G, Hillemeier MM, Hammer CS, Maczuga S. 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*. 2015; 86:1351–1370.10.1111/cdev.12398 [PubMed: 26283023]
- Najarian, M.; Pollack, JM.; Sorongon, AG.; Hausken, EG. Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K): Psychometric Report for the Eighth Grade. Washington, DC: National Center for Education Statistics; 2009.
- NICHD Study of Early Child Care. Procedures for videotaping mother-child interaction at 15 months. Chapter 15.3 in *The NICHD Study of early child care and youth development, phase I manuals 6*. 1992. Retrieved from <http://secc.rti.org/manuals.cfm>
- NICHD Early Child Care Research Network. Child care and mother-child interaction in the first three years of life. *Developmental Psychology*. 1999; 35:1399–1413.10.1037/0012-1649.35.6.1399 [PubMed: 10563730]
- Olswang LB, 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; 7:23–32.10.1044/1058-0360.0701.23
- Olszewski P. Individual differences in preschool children's production of verbal fantasy play. *Merrill-Palmer Quarterly*. 1987; 33:69–86. Stable : <http://www.jstor.org/stable/23086147>.
- Overton, WF. Processes, relations, and relational-developmental-systems. In: Bornstein, MH.; Leventhal, T.; Lerner, RM., Editor-in-chief, editors. *Theory and method* Volume 1 of the *Handbook of child psychology and developmental science*. 7. Hoboken, NJ: Wiley; 2015. p. 9-62.
- Pan BA, Rowe ML, Singer JD, Snow CE. Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*. 2005; 76:763–782.10.1111/1467-8624.00498-i1 [PubMed: 16026495]
- Paulsell D, Kisker EE, Love JM, Raikes HH. Understanding implementation in Early Head Start programs: Implications for policy and practice. *Infant Mental Health Journal*. 2002; 23:14–35.10.1002/imhj.10001
- Perfetti C, Stafura J. Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*. 2014; 18:22–37.10.1080/10888438.2013.827687

- Petersen IT, Bates JE, D'Onofrio BM, Coyne CA, Lansford JE, Dodge KA, Pettit GS, Van Hulle CA. Language ability predicts the development of behavior problems in children. *Journal of Abnormal Psychology*. 2013; 122:542–557.10.1037/a0031963 [PubMed: 23713507]
- Pine JM, Lieven EV, Rowland C. Observational and checklist measures of vocabulary composition: What do they mean? *Journal of Child Language*. 1996; 23:573–590.10.1017/S0305000900008953
- Records NL, Tomblin JB. Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *Journal of Speech, Language, and Hearing Research*. 1994 Feb.37:144–156.10.1044/jshr.3701.144
- Rowe DC, Jacobson KC, Van den Oord EJ. Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development*. 1999; 70:1151–1162.10.1111/1467-8624.00084 [PubMed: 10546338]
- Rowe ML, Raudenbush SW, Goldin-Meadow S. The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*. 2012; 83:508–525.10.1111/j.1467-8624.2011.01710.x [PubMed: 22235920]
- Sénéchal M, Ouellette G, Rodney D. The misunderstood giant: On the predictive role of early vocabulary to future reading. *Handbook of early literacy research*. 2006; 2:173–182.
- Snow, CE.; Burns, S.; Griffin, P. Preventing reading difficulties in young children. Washington, DC: National Academy Press; 1998.
- Sparrow, SS.; Balla, DA.; Cicchetti, DV. Vineland Adaptive Behavior Scales Survey Form Manual (Interview Edition). Circle Pines, MN: American Guidance Service; 1984.
- Steenkamp JEM, Baumgartner H. Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*. 1998; 25:78–90.10.1086/209528
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics. 6. Needham, MA: Allyn & Bacon; 2012.
- Taris TW, Bok IA, Meijer ZY. Assessing stability and change of psychometric properties of multi-item concepts across different situations: A general approach. *The Journal of Psychology*. 1998; 132:301–316.10.1080/00223989809599169
- Thal, DJ.; Katich, J. Predicaments in early identification of specific language impairment: Does the early bird always catch the worm?. In: Cole, KN.; Dale, PS.; Thal, DJ., editors. *Assessment of communication and language*. Vol. 6. Baltimore, MD: Paul H Brookes Publishing; 1996. p. 1-28.
- Tomblin JB, Zhang X. The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*. 2006; 49:1193–1208.10.1044/1092-4388(2006/086)
- Tourangeau, K.; Nord, C.; Lê, T.; Sorongon, AG.; Najarian, M. Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Combined User's Manual for the ECLS-K Eighth-Grade and K–8 Full Sample Data Files and Electronic Codebooks (NCES 2009–004). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; Washington, DC: 2009.
- Trouton A, Spinath FM, Plomin R. Twins early development study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research: The Official Journal of the International Society for Twin Studies*. 2002; 5:444–448.10.1375/136905202320906255 [PubMed: 12537874]
- U.S. Department of Health and Human Services Administration for Children Families. Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. 2001. Retrieved October 19, 2015 <http://www.mathematica-mpr.com/PDFs/buildingvoll.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; 3:4–69.10.1177/109442810031002
- Wechsler, D. Wechsler Intelligence Scale for Children-WISC-IV. Psychological Corporation; 2003.
- 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; 30:583–608.10.1017/S0305000903005671 [PubMed: 14513469]
- Woodcock, RW.; Johnson, MB. Woodcock-Johnson Revised Tests of Achievement. Itasca, IL: Riverside Publishing; 1990.

- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson tests of achievement. Itasca, IL: Riverside Publishing; 2001.
- World Health Organization. The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines. 1993. Available from: <http://www.who.int/classifications/icd/en/bluebook.pdf>
- Yuan KH, Bentler PM. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*. 2000; 30:165–200.10.1111/0081-1750.00078



**Figure 1.** Models of language stability from 15 months to 11-years. Parameter estimates shown on the upper half are from the average-to-high skill group, those shown in the lower half are from the low skill group. Numbers associated with single-headed arrows are standardized path coefficients. Indicators of each language latent variable are listed with their factor loadings. Not shown in the figure but estimated in the model are error variances of indicators, the amount of variance not accounted for by paths in the model.

**Table 1**

Child and Family Characteristics by Language Skill Group

Demographic Variable	Low Skill	Average-to-High Skill	Test Statistics/Effect Size
Child age, <i>M</i> ( <i>SD</i> )			
Wave 1 (15 months)	15.16 (1.80)	15.01 (1.66)	$t(349.26)^d = 0.83, ns/.00$
Wave 2 (25 months)	25.25 (2.02)	25.11 (1.89)	$t(360) = 0.64, ns/.00$
Wave 3 (5 years)	5.27 (0.34)	5.27 (0.33)	$t(325) = -0.04, ns/.00$
Wave 4 (11 years)	11.09 (0.30)	11.09 (0.32)	$t(416) = 0.03, ns/.00$
Child gender (% female)	46.8	47.7	$\chi^2(1, N = 421) = 0.04, ns/.01$
Firstborn child (%)	66.2	66.7	$\chi^2(1, N = 420) = 0.01, ns/.01$
Ethnicity (% African American)	82.1	24.5	$\chi^2(1, N = 421) = 139.35, p < .001/.58$
Child has one or multiple risks listed below (%)	29.9	24.5	$\chi^2(1, N = 421) = 1.50, ns/.06$
Has established risks <sup>b</sup>	15.7	7.8	$\chi^2(1, N = 288) = 4.39, p < .05/.12$
Has biological or medical risks <sup>b</sup>	22.4	13.0	$\chi^2(1, N = 288) = 4.41, p < .05/.12$
Has environmental risks <sup>b</sup>	33.6	27.3	$\chi^2(1, N = 288) = 1.35, ns/.07$
Maternal age at birth of child, <i>M</i> ( <i>SD</i> )	21.02 (5.15)	23.18 (5.94)	$t(417.88)^d = -4.00, p < .001/.04$
Living arrangements (%)			$\chi^2(2, N = 420) = 45.52, p < .001/.33$
Living with a spouse	8.4	35.2	
Living with other adults	48.8	28.8	
Living alone with child	42.8	36.0	
Highest education obtained <sup>c</sup> (%)			$\chi^2(2, N = 415) = 65.67, p < .001/.40$
Less than 12 <sup>th</sup> grade	53.2	22.0	
12 <sup>th</sup> grade or earned a GED	33.8	31.8	
More than 12 <sup>th</sup> grade	13.0	46.3	
Primary caregiver's occupation (%)			$\chi^2(2, N = 417) = 9.68, p < .01/.15$
Employed	23.0	31.3	
In school or a training program	31.5	18.9	
Neither employed nor in school or training	45.5	49.8	
Family income below poverty line (%)	89.4	82.9	$\chi^2(1, N = 353) = 3.02, ns/.09$
Welfare recipient (%)	69.2	41.6	$\chi^2(1, N = 404) = 31.05, p < .001/.28$

Demographic Variable	Low Skill	Average-to-High Skill	Test Statistics/Effect Size
Receiving Early Head Start services (%)	54.7	52.7	$\chi^2(1, N = 421) = 0.17, ns/.02$
<i>Covariate</i>			
Maternal Education in years	8.59 (1.97)	10.03 (1.74)	$\kappa(399.16)^a = -7.90, p < .001/.13$
Maternal Language	84.60 (7.91)	98.88 (10.76)	$\kappa(278.44)^a = -12.93, p < .001/.36$
HOME total scores	24.58 (3.68)	28.24 (2.34)	$\kappa(233.51)^a = -10.45, p < .001/.27$
Child Social Interaction	4.53 (0.90)	5.22 (0.78)	$\kappa(305) = -7.19, p < .001/.15$
<i>Demographic Variable</i>			
Bayley Visual/Spatial factor	Low Skill 3.62 (2.64)	Average-to-High Skill 6.89 (2.78)	Test Statistics/Effect Size $\kappa(294) = -10.22, p < .001/.26$
WISC Matrix Reasoning	6.26 (2.93)	10.57 (2.94)	$\kappa(415.96)^a = -15.07, p < .001/.35$

Note. Reported effect size for a *t*-test is partial eta squared and for a Chi-Square test is the *phi* coefficient (for a 2 by 2 table) or equivalent (for a 2 by 3 table). Eta squared values of .01, .06, and .14 are considered small, medium, and large effect sizes, respectively (Cohen, 1988). For a test of only one predictor variable, partial eta squared is equivalent to eta squared. *Phi* values of .1, .3, and .5 are considered small, medium, and large effect sizes, respectively (Cohen, 1988).

<sup>a</sup>Modified degrees of freedom are reported for the separate-variance *t*-test.

<sup>b</sup>Variable value was coded as "1" if child had risks or data were ambiguous, "0" if child had no risks, otherwise variable was set to missing. These categories of risks are used by most states to identify young children at risk for adverse developmental outcomes. Examples of established risks are a chromosomal abnormality, a congenital birth defect, a sensory impairment, or HIV/AIDS. Examples of biological or medical risks are congenital heart disease, diabetes, low birth weight, or a severe chronic illness. Examples of environmental risks are parental substance abuse, low maternal education, suspected child abuse or neglect, family social disorganization, or homelessness.

<sup>c</sup>Recoded from years of education. Years of education was used in SEM.

**Table 2**  
 Child Language Measures by Skill Group: Descriptive Statistics and Pair-wise Variance Covariance Matrix

	1	2	3	4	5	6	7	8	9	10
15 months										
1. CDI Comprehension	1.00/1.00	.43	.35	.29	.19	.09	.00	-.02	-.01	.11
2. CDI Production	.46	1.02/1.02	.38	.47	.48	.17	-.05	-.06	-.08	.03
3. CDI Early Gestures	.46	.30	1.00/1.00	.23	.11	.11	-.01	-.04	-.12	.04
25 months										
4. CDI Production	.37	.31	.38	1.00/1.00	.68	.33	.30	.13	.05	.11
5. CDI Sentence Complexity	.41	.25	.34	.57	1.00/1.00	.31	.24	.05	.01	-.01
6. Bayley Language Factor	.28	.30	.21	.45	.43	1.01/1.00	.09	-.01	.02	.03
5 Years										
7. PPVT-III	.06	.12	.10	.33	.19	.45	1.01/1.00	.27	.48	.27
8. WJ Letter-Word Identification	.09	.05	-.02	.16	.06	.08	.32	1.01/1.00	.24	.25
11 Years										
9. PPVT-III	.22	.05	.16	.46	.30	.25	.31	.08	1.01/1.00	.52
10. ECLS-K Language/Literacy	.09	.01	.04	.26	.31	.10	.29	.32	.40	1.00/1.00
<i>M</i>										
Low skill	47.13	10.81	14.22	48.45	6.86	6.01	77.23	84.40	74.76	94.01
Average-to-high skill	51.18	14.32	14.48	66.40	12.38	9.62	105.88	98.44	113.27	160.81
<i>SD</i>										
Low skill	18.83	11.48	2.01	21.96	8.37	3.38	11.65	13.22	7.26	21.95
Average-to-high skill	17.01	12.37	1.92	19.93	9.00	2.17	11.65	12.43	8.36	8.58
<i>Range</i>										
Low skill group	2-89	0-57	8-18	1-100	0-37	0-12	40-103	46-114	40-84	31.51-122.38
Average-to-high skill	14-89	0-64	8-18	13-100	0-37	0-12	73-135	68-133	100-138	143.48-180.65
Effect size for mean comparisons	.01	.04	.01	.17	.12	.32	.60	.23	.86	.88

*Note.* Covariances shown below the diagonal were from children of low language skill; those above the diagonal were from children of average-to-high language skill. Variances shown before the slash are from children of low language skill; those after the slash are from children of average-to-high language skill. All variables are scaled by constants so that variables' variances were  $\approx 1$ . Reported effect size is partial eta squared; values of .01, .06, and .14 are considered small, medium, and large effect sizes, respectively (Cohen, 1988).



**Table 3**  
Zero-order Correlations between Covariates and Language Measures by Skill Group

	Child Social Interaction	Bayley Visual/Spatial factor	WISC Matrix Reasoning	Maternal Education	Maternal Language	HOME total scores
15 months						
CDI Comprehension	.01/-.03	--	--	-.09/-.13	.03/.02	.10/.10
CDI Production	.02/.05	--	--	-.04/-.16*	-.03/-.13	.10/.09
CDI Early Gestures	.14/.01	--	--	-.03/-.09	.08/-.13	.23**/.12
25 months						
CDI Production	.23*/.13	.21*/.28***	--	-.15/.00	-.02/.08	.00/.07
CDI Sentence Complexity	.20*/.09	.26**/.21**	--	-.08/-.02	.15/.10	.12/.08
Bayley Language Factor	.18/.24**	.39***/.30***	--	-.21*/.11	-.00/.09	.05/.21**
5 years						
PPVT-III	.07/.24**	.18/.39***	--	-.02/.32***	.05/.40***	.10/.28***
WJ Letter-Word Identification	-.01/.08	.16/.07	--	.14/.22**	.00/.15	.12/.08
11 years						
PPVT-III	.23*/.07	.11/.25***	.36***/.17**	-.05/.25***	.08/.48***	-.04/.22**
ECLS-K Language/Literacy	.18*/.14	.13/.19*	.43***/.18**	.14*/.12	.12/.21**	.10/.18*

Note. Correlations shown before the slash are from children of low language skill (*d*f ranged from 87 to 199); those after the slash are from children of average-to-high language skill (*d*f ranged from 126 to 218). Bayley Visual/Spatial factor was measured at 25 months, WISC Matrix Reasoning at 11 years, only concurrent and predictive relations of child nonverbal intelligence on language measures are presented.

\*  $p < .05$ ;  
 \*\*  $p < .01$ ;  
 \*\*\*  $p < .001$ .