JSLHR

Research Article

Predicting Vocabulary Growth in Children With and Without Specific Language Impairment: A Longitudinal Study From 2;6 to 21 Years of Age

Mabel L. Rice[a] and Lesa Hoffman[a]

Purpose: Children with specific language impairment (SLI) often have vocabulary impairments. This study evaluates longitudinal growth in a latent trait of receptive vocabulary in affected and unaffected children ages 2;6 (years;months) to 21 years and evaluates as possible predictors maternal education, child gender, and nonverbal IQ.

Method: A sample of 519 participants (240 with SLI; 279 unaffected) received an average of 7 annual assessments for a total of 3,012 latent trait Peabody Picture Vocabulary Test (PPVT) observations. Unconditional and conditional multilevel growth models were estimated to evaluate growth trajectories and predictor relationships over time.

Results: Children with SLI had lower levels of receptive vocabulary throughout the age range assessed. They did not close the gap with age peers. Children with higher nonverbal IQs had better PPVT performance, as did children of mothers with higher education. Child gender showed an advantage for young girls that leveled out with age and then became an advantage for boys from ages 10 to 21 years. All children's rate of vocabulary acquisition slowed around 12 years of age.

Conclusions: The outcomes of the study have implications for hypothesized causal pathways for individual differences; predictions differ for children under 5 years, 6–10 years, and later ages.

Children's vocabulary acquisition is widely recognized as a core component of their emerging linguistic abilities, a component with ties to general cognitive abilities such as reading and school success. Standardized vocabulary assessments are often included in assessments of children's language and cognitive development. These assessments reveal that children of the same age vary in vocabulary size. Children with specific language impairment (SLI) often demonstrate vocabulary delays compared with their age peers (Gray, Plante, Vance, & Henrichsen, 1999; Rice et al., 2010), and experimental studies of word learning show consistent deficits for children with SLI compared with unaffected children (Ellis Weismer & Hesketh, 1993, 1998; Gray, 2004, 2005; Oetting, Rice, & Swank, 1995; Rice, Buhr, & Nemeth, 1990; Rice, Buhr, & Oetting, 1992; Rice, Oetting, Marquis, Bode, & Pae, 1994). The causes of this variation from age expectations are not known, although increasing evidence suggests that scores adjusted for age expectations tend to be stable over time. Effects of environment and genetics are thought to contribute to the individual differences, and recent studies have explored associations between reading and language across repeated times of measurement in investigations of causal pathways.

An important gap in our knowledge of children's vocabulary acquisition is a lack of long-term longitudinal studies comparing the vocabulary growth of children with and without SLI. The available studies using cross-sectional methods are informative but cannot address the actual path of individual change over time, the underlying dynamics of the observation that children maintain their relative rank within their age peer group, and the transitions from childhood to adolescence to adulthood. Individual growth data would inform studies of causal pathways, allowing for examination of individual differences in vocabulary change over time as well as the relationship of putative predictors over time comparing SLI-affected and unaffected groups. The 20-year study reported here aims to move the field forward by investigating a trait-based vocabulary metric suitable for comparing the performance of children from preschool into young adulthood, describing patterns of

---

[a]University of Kansas, Lawrence

Correspondence to Mabel L. Rice: mabel@ku.edu

Editor: Rhea Paul
Associate Editor: Shelley Gray

Received June 2, 2014
Revision received September 14, 2014
Accepted November 12, 2014
DOI: 10.1044/2015_JSLHR-L-14-0150

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

individual change, evaluating predictors of growth, examining ways in which children with SLI grow similarly or differently from unaffected age peers, and enhancing hypotheses of possible causal pathways.

Individual differences in vocabulary acquisition, SLI, heritability, associated variables, and possible causal pathways. Studies of twin children find significant heritability for vocabulary acquisition in early childhood, such that heritability estimates are higher for children with SLI as compared with estimates from a full population-based sample, suggesting that the etiology of limited vocabulary learning is likely to involve biological as well as environmental factors (Bishop, Price, Dale, & Plomin, 2003; Dale, Price, Bishop, & Plomin, 2003; DeThorne, Petrill, Hayiou-Thomas, & Plomin, 2005; Rice, Zubrick, Taylor, Gayán, & Bontempo, 2014). Although it is known that language impairments are likely to persist in children with SLI during elementary school, even into adolescence (Conti-Ramsden, St. Clair, Pickles, & Durkin, 2012; Johnson et al., 1999), little information is available about long-term vocabulary growth trajectories for children with SLI compared with unaffected children, limiting our understanding about when and how genetic influences operate or how genetic and environmental influences could change over time.

Understanding the sources of variability in vocabulary acquisition is important given that individual differences in vocabulary acquisition predict children's subsequent academic achievement. Early and robust vocabulary growth is associated with higher reading levels in elementary school (Adlof & Perfetti, 2013; Catts, Adlof, & Weismer, 2006; Catts & Kamhi, 2005). Vocabulary plays a central role in reading comprehension; correlations between receptive vocabulary scores and reading scores from kindergarten to 10th grade are in the range of $r = .45$ to $r = .77$, according to the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4) manual (L. M. Dunn & Dunn, 2007). The direction of influence is bidirectional: Vocabulary contributes to reading comprehension, and, after children learn to read, reading adds to vocabulary (Nagy & Anderson, 1984; Nagy, Herman, & Anderson, 1985), a relationship extending across the life span (Adlof & Perfetti, 2013; Cunningham & Stanovich, 1998; Mol & Bus, 2011). Children with SLI are at high risk for reading impairments in middle childhood. About 50% of young children with SLI have subsequent reading impairments, which are associated with their earlier and concurrent language impairments (Catts, 2004). Presumably, children with SLI who are poor readers are at long-term risk for vocabulary development, although little is known about their long-term vocabulary growth patterns.

Twin studies aim to differentiate environmental and genetic effects on reading abilities with recent introductions of repeated measurements of reading to evaluate effects over time. A recent twin study (Astrom, Wadsworth, Olson, Willcutt, & DeFries, 2011) used longitudinal measures of reading (at 10;6 [years;months] and 15;5 years of age) to assess heritability, reporting that nearly 60% of the target child's reading deficit at follow-up was due to genetic factors that influenced reading difficulties at the initial assessment. It is possible that the inherited effects on reading are related to the positive association of reading impairments with language impairments. Another twin study of children ages 8–18 years (Olson et al., 2013) examined heritability of language skills (including vocabulary), three different reading skills, and writing abilities. Substantial genetic influence was found on all of the language and reading measures, whereas environmental influences were significant for vocabulary but not for the other measures. A recent candidate gene linkage and association study of the families of children with SLI (Rice, Smith, & Gayán, 2009) is the only candidate gene study of SLI Probands known to the authors that uses a vocabulary phenotype (PPVT) as well as omnibus speech, language, and reading phenotypes. The study of 322 participants reported a high correlation between the vocabulary and reading comprehension scores ($r = .718$, $p \leq .01$), but the genetic outcomes did not reveal shared genetic sources of variance. There was suggestive association of general omnibus language ability, speech impairments, and text comprehension with single nucleotide polymorphisms on the *KIAA0319* gene, although no effects were evident for the vocabulary phenotype.

Environmental variables are thought to be associated with higher levels of vocabulary, although the evidence is mixed. Of interest here is maternal education, which can be considered an index of familial resources related to child rearing (Entwisle & Astone, 1994). Higher levels of maternal education are reported to be associated with higher levels of children's vocabulary development (Reilly et al., 2010). Yet in a large population-based longitudinal study, the maternal education effect was only apparent at the first assessment at 4 years of age and was negligible for producing onward change in subsequent years (Taylor, Christensen, Lawrence, Mitrou, & Zubrick, 2013).

Early on, gender differences are striking. Boys' vocabulary size is smaller than that of girls, requiring gender-differentiated norms during the early period of word acquisition (Fenson et al., 2007); boys are three times more likely than girls to be late in acquiring their first words (Zubrick, Taylor, Rice, & Slegers, 2007). The extent to which this strong gender difference persists into later childhood is not well documented, although a recent study of 4,332 children from 4 to 8 years of age reported negligible effects of gender on vocabulary growth indexed by intercept and slope during this early childhood period (Taylor et al., 2013). A previous sample of 329 children compiled across four longitudinal studies of language acquisition collectively bridging ages 1;1 to 6;10 (and using mixed methods of measurements) reported that in the second through fifth years, but not before or after, girls consistently outperformed boys in multiple measures of language (Bornstein, Hahn, & Haynes, 2004). Although the recent findings suggest that gender differences would not appear beyond early childhood, the manual for the Peabody Picture Vocabulary Test–Revised (PPVT-R; A. Dunn & Dunn, 1981) notes that boys scored slightly higher on the original PPVT; in the standardization sample of the PPVT-R, boys continued to outperform girls by a slight margin.

In addition to gender differences, children with higher levels of nonverbal intelligence are likely to have larger vocabularies than children with lower levels. The manual of the Peabody Picture Vocabulary Test–III (PPVT-III; L. M. Dunn & Dunn, 1997) reports correlations of $r = .67$ (.82–.84 when corrected for variability of the norm group) for the PPVT scores with nonverbal IQ scores from the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991). Recent genetics studies link nonverbal IQ and gender on the causal pathways of disorders such as autism, a finding of interest because many children with autism also have language disorders. For example, one study (Banach et al., 2009) examined the hypothesis that girls at risk for autism are protected in some way, so that only those with the greatest genetic liability are affected. The prediction that follows is that affected male siblings of girls with autism should be more impaired than affected male siblings of male probands. The study found support for the hypothesis in differences between families with a single child with autism (simplex) compared with families with more than one child with autism (multiplex). In the simplex families, girls had lower IQs than boys, but no such differences were seen among multiplex families. It remains to be seen whether possible interactions exist between nonverbal IQ and gender in the vocabulary acquisition of children with SLI; if so, these interactions would be informative for causal models.

Measurement issues: The need for longitudinal evidence. The available research on vocabulary growth is limited by a heavy reliance on cross-sectional samples of children and standard scores adjusted for age expectations to provide indications of how vocabulary growth changes over time. A relevant example, found in the fitted age curve provided in the manual of the PPVT-III test, is based on the cross-sectional standardization sample comprising ages 2;6–90+ years. The curve shows a strong linear effect from age 2;6 years until about age 15 years, when it decelerates and levels off from ages 20 to 90+ years. A similar linear-looking outcome was evident in an earlier cross-sectional analysis of PPVT data from a subset of children in this study, with SLI-affected ($n = 170$) and unaffected ($n = 136$) groupings from ages 3 to 9 years (Rice et al., 2010); the group mean outcomes showed consistently lower performance by the SLI group over this time frame.

A major limitation of such cross-sectional designs is that they cannot address change within individual children. Two recent longitudinal growth studies are interesting exceptions. The first study (Rowe, Raudenbush, & Goldin-Meadow, 2012) developed growth estimates from a demographically diverse sample of 62 children, using their expressive vocabularies in spontaneous language samples at 30 months of age to predict their receptive vocabulary scores on the PPVT-III at 54 months of age. Velocity and acceleration in early vocabulary development predicted later vocabulary, particularly for children with low socioeconomic backgrounds, suggesting a strong relationship between earlier growth patterns and later outcomes during the preschool developmental period. Although informative, these findings may be limited by the relatively small sample, different measurements across the time span, and the lack of information about the children's language or cognitive status on conventional measurements.

The second study (Taylor et al., 2013) investigated receptive vocabulary development in a population-based sample of 4,332 children at four occasions of measurement from 4 to 8 years (at 50, 57, 82, and 105 months). The outcome measure was the raw score from an adapted short version of the PPVT-III. The strengths of the study were the population-based sample, a large number of predictor variables, and growth modeling in which variables that predicted status at the first occasion of measurement differed from those that predicted change over time. Putative risk variables at age 4 years predicted a higher, not lower, rate of growth from ages 4 to 8 years, presumably because the low-ability children had more "room to grow," although the elevated rate was not sufficient to close the receptive vocabulary gap for children with and without these risks at age 8 years. In order of descending strength of prediction, these risk variables were maternal non-English-speaking background, low child school readiness, and maternal mental health distress. Socioeconomic-area disadvantage was the only risk associated with a lower rate of growth in receptive vocabulary, although it was not a significant risk for low receptive vocabulary ability at age 4 years. Of interest here is that gender had negligible effects on the PPVT intercept and slope, suggesting that an advantage for girls was not noteworthy from ages 4 to 8 years. The limitations of this study are a lack of a standardized omnibus language assessment to allow for classification of children as SLI, a lack of direct assessment of nonverbal IQ, and the 4-year time window relatively early in child development.

Longitudinal growth studies of children with SLI and control children have examined the development of morphosyntax in the age range of 3–16 years (Rice, Hoffman, & Wexler, 2009; Rice, Wexler, & Hershberger, 1998; Rice, Wexler, & Redmond, 1999), which can provide informative comparisons with vocabulary growth. It is an open question as to whether all dimensions of language change in tandem over childhood or whether different dimensions, such as grammar and vocabulary, follow different trajectories of growth. Although the morphosyntactic measures differed across studies, the outcomes were similar, and change over time was nonlinear. Children with SLI started to acquire target morphemes later than unaffected children, but the growth trajectories did not differ. Once the affected children started to acquire the grammatical property of finiteness marking, the group followed the same growth trajectory as unaffected children, but in early adolescence, the performance of the SLI group leveled off below the level attained by the unaffected children, whose performance also leveled off but with higher scores. One consequence is that the SLI group did not "catch up" to the unaffected children. In those studies, maternal education, gender, and nonverbal IQ did not predict growth, and interactions were not significant.

All things considered, there is a need for a longitudinal study with the following characteristics: (a) covers a longer time frame encompassing preschool into adulthood to see whether the strong linear growth followed by the adult asymptote shown in the PPVT-III standardization model holds at the level of individual children; (b) involves a well-documented sample that includes children identified as having SLI compared with children without SLI to evaluate possible similarities and differences in individual growth patterns as a step toward a better understanding of causal pathways; (c): evaluates the same robust index of vocabulary development across the full age range to avoid possible measurement confounds; and (d) evaluates children's gender, nonverbal IQ, and mother's education as predictors to add to the understanding of causal pathways.

Measurement Issues

Measurement challenges have presented significant barriers to longitudinal studies of vocabulary development over the full age range of childhood and the full range of variability across children. On the positive side, the PPVT-III is a psychometrically robust measure of receptive vocabulary that has low response demand (pointing to pictures) and age reference groups from age 2 years well into adulthood and also meets high standards for reliability and validity. In spite of the obvious advantages, challenges remain. New editions of the test appearing during the time of a longitudinal study may have psychometric differences that result in noncomparable estimates of ability that complicate the measurement of growth over time. Within the time of the longitudinal study reported here, three different editions of the PPVT appeared: PPVT-R, PPVT-III, and PPVT-4. The PPVT-R differed from the PPVT-III in outcomes for the same children, complicating the comparisons of children's earlier performance with their later performance as assessed by the PPVT-III (Pankratz, Morrison, & Plante, 2004; Pena, Spaulding, & Plante, 2006; Ukrainetz & Duncan, 2000). Another measurement issue is the widespread use of standard scores in archival databases as a measure that provides an estimate of how a child's performance compares with the distribution of children in an age reference group. Because standard scores are adjusted for change over age levels, they cannot be used to index absolute amounts of change—growth trajectories using standard scores show generally flat trajectories if children's language grows as expected per age level and if children with SLI do not improve their rank among age peers over time (Conti-Ramsden et al., 2012).

Fortunately, these measurement challenges are not unique to the assessment of vocabulary ability and can be well addressed by a field of statistics known as item response theory (IRT), a set of analytic techniques widely used in educational assessment and in many other fields. We conducted a preliminary study using IRT to develop a trait-level measure of receptive vocabulary ability by modeling item responses from two editions, the PPVT-R and the PPVT-III. The children's responses to each test were collected within days of each other using the same test protocols, such that a common-persons linking design was created in addition to the common items across tests (Hoffman, Templin, & Rice, 2012). This unusual—and time-consuming—method provided latent trait estimates that reliably assess the same underlying vocabulary acquisition ability over the full time window, from 2;6 to 21 years of age. This psychometric study also yielded a Monte Carlo algorithm for estimating latent trait estimates of vocabulary ability given responses to any of the PPVT-R and/or PPVT-III items. These ability estimates can be meaningfully compared across children at a single measurement occasion or across multiple measurement occasions regardless of which version of the test was administered and thus provide a robust method of modeling individual vocabulary growth.

Current Study: Hypotheses and Questions

Our main objective in the current study was to model the developmental outcomes of growth over time in a latent trait of receptive vocabulary in two groups of children: one ascertained with SLI and a comparison group without SLI. Our initial guiding hypotheses were (a) vocabulary growth would be strongly linear over time; (b) the SLI group would show a more shallow slope of vocabulary learning throughout the observed time; (c) maternal education, as an index of environmental parenting influences, would influence vocabulary acquisition for both groups, with a stronger influence on the SLI group; (d) girls would be better than boys at early vocabulary development, an advantage that would disappear by age 5 years; and (e) children with higher levels of nonverbal IQ would have an advantage throughout the age range; a Nonverbal IQ × Gender × Affectedness interaction would have implications for genetic causal models.

The analyses addressed these specific questions:

1. What are the trajectories of growth in a trait-based estimate of receptive vocabulary in children ages 2;6–21 years?
2. Do children with SLI demonstrate growth trajectories different from those of unaffected children?
3. Do growth trajectories differ between boys and girls affected with SLI and those unaffected with SLI?
4. Do growth trajectories differ by maternal education and nonverbal IQ after considering SLI affectedness and gender?

Method

Participants

Participants for this study were drawn from an archival database collected as part of a 20-year ongoing longitudinal study, approved by the University of Kansas Human Subjects Committee, of children with SLI and their siblings as well as control children and their siblings (Rice et al., 2009, 2010). The parent study is a family-based candidate gene study of SLI; one aim is the development of linguistic growth phenotypes for genetic investigations. Children with SLI and control children were recruited from speech pathologists in public schools. Over the course of the study, the children attended more than 100 schools and attendance centers in the midwestern United States. Siblings were recruited into the study following the enrollment of the target children. All children in this study had normal or above-normal intellectual functioning, defined as a standard score of 85 or above on an age-appropriate test of nonverbal IQ (see measures in the Measures section) at time of initial assessment; had no diagnosis of autism, intellectual, behavioral, or social impairments; and passed a hearing screening at 25 dB (30 dB in noisy environments) at 1000, 2000, and 4000 Hz.

A total of 519 persons participated (of whom 41% were girls and 59% were boys) from a total of 259 families, with a range of one to eight participants per nuclear family (M = 2.0, SD = 1.1). The race–ethnicity percentages were White, 82.7%; multiracial, 10.4%; American Indian, 4.6%; Black, 0.7%; Asian, <1%; and not reported, 1.5%. Hispanic ethnicity was reported by 6.5% of the sample. Participants ranged in age from 2;6 to 20 years at the first occasion (M = 6;11, SD = 3;8). The number of occasions of measurement per person ranged from 1 to 19 (M = 6.9, SD = 4.8), for a total of 3,012 observations analyzed.

Procedure

Data for this study were drawn from the standardized assessments in the archival longitudinal study. The base study uses an accelerated longitudinal design in which multiple age cohorts are sampled (i.e., such that children varied continuously in age at first assessment) with longitudinal data collected on members of each cohort. The age at first time of assessment varies across children; for example, in this study, siblings of children with SLI and controls could enter the study any time between 2;6 and 20;11. Data were collected by trained examiners in customized vans at the participants' homes or schools, thereby lessening the demands on the families for ongoing participation in the longitudinal study. An individual examiner assessed an individual participant in a session of about 1-hr duration, with multiple sessions as needed.

Measures

Exclusionary and inclusionary assessments for SLI. Nonverbal IQ was determined by age-appropriate measures at initial assessment: the Columbia Mental Maturity Scale (CMMS; Burgemeister, Blum, & Lorge, 1972) from ages 3;6–5;11 years (for children first seen at 2;6 years, their subsequent CMMS scores at 3;6 years were used for the nonverbal IQ estimate); the WISC-III from ages 6;0–16;11 years; and the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III; Wechsler, 1997) for participants ages 17–20;11. To be included in the study, participants needed scores of 85 or above.

Language grouping status was determined by participants' scores at first time of measurement on age-appropriate omnibus language tests: The Test of Early Language Development–Third Edition (TELD-3; Hresko, Reid, & Hammill, 1999) from 2 to 3;11 years of age; the Test of Language Development–Primary: Second Edition (TOLD P:2; Newcomer & Hammill, 1988) from 4 to 5;11, and the Clinical Evaluation of Language Fundamentals–Third Edition (CELF-3; Semel, Wiig, & Secord, 1995) from 6+ years. Standard scores were used to divide participants into an affected group on the basis of the criterion of scores ≤ 85 versus unaffected as children with scores ≥ 86 (with the exception of six children entered at 3 years as affected because their performance on the Test of Early Grammatical Impairment [Rice & Wexler, 2001] was more than two standard deviations below the mean, even though their omnibus language score on the TELD-3 was between 86 and 100). These criteria were applied across children initially recruited as SLI, unaffected controls, and their siblings. Thus, siblings would be differentiated as affected or unaffected, depending on their score on the initial omnibus language assessment. Of the full sample, 46% were ascertained as being affected in their language development; 54% were unaffected. Note that grouping effects over time could be weakened if omnibus tests varied in sensitivity for identifying children with SLI. It has been noted that the TELD-3 may be less sensitive and/or for children under 4 years may yield inaccurate groupings for identification of SLI. To evaluate this possibility, we examined the 106 participants who entered the study in this age range and were grouped according to TELD-3 outcomes. We compared their TELD-3 scores with their subsequent TOLD-P:2 scores. Of the 106 children, 11 (10%) were arguably "false negatives." They were assigned to the unaffected group at first testing, and on subsequent TOLD-P:2 assessment, they scored in the affected range. All things considered, the initial omnibus assessment was deemed most appropriate for grouping for the purpose of the study, though any errors in grouping would add error variance to the groups and act against detection of group differences over time.

Vocabulary outcomes. Each year, all participants were administered the PPVT-R form M and/or the PPVT-III Forms A and B. Latent trait estimates of receptive vocabulary were then created using the algorithm developed in the earlier study (Hoffman et al., 2012) in which all item responses—across all test forms, all participants, and all occasions of measurement—were analyzed using two-parameter logistic item response models. These models locate persons and items on a common latent trait metric (which was identified by setting person ability to $M = 0$, $SD = 1$) and provide direct estimates for the ability of each person (i.e., his or her location on the latent trait), the difficulty of each item (i.e., its location on the latent trait), and the discrimination of each item (i.e., its strength of relationship to the latent trait).

Given these calibrated item parameters, the most likely estimate of ability at each occasion for each person —whether part of the initial calibration sample or not— can then be found from his or her item responses. This ability estimate for each participant at each occasion was the outcome for which growth curve models were estimated in the current study. However, given that the metric of the latent trait of vocabulary is arbitrary, to facilitate the reporting of variance components for growth, ability estimates were multiplied by 100, such that the sample PPVT ability was $M = 1.88$, $SD = 88.20$. Note that multiplying a variable by a constant changes its variance but not the mean. Furthermore, additional observations were included relative to the calibration sample, such that the standard deviation for the original latent trait was no longer exactly 1.

Predictors of vocabulary outcome. Nonverbal IQ standard scores were used as a time-invariant continuous predictor, which was centered at 100 for analysis. Children's exact age at each occasion was a time-varying continuous predictor, which was converted to years and centered at age 10 years for analysis. In addition, children's age at the first time of assessment (centered near the sample mean of 7 years) was a time-invariant continuous predictor included to control for differential cross-sectional and longitudinal effects of age. Maternal education was an additional time-invariant continuous predictor, collected via questionnaire at the first time of assessment. Maternal education was coded as 1 (*some high school*), 2 (*high school graduate or general equivalency diploma* [*GED*]), 3 (*some college*), 4 (*bachelor's degree*), 5 (*some graduate school*), and 6 (*graduate degree*). Preliminary analyses revealed a nonsignificant improvement in fit from modeling mother's education as a categorical rather than a continuous predictor. Accordingly, mother's education was treated as a continuous predictor in the analyses reported below (centered such that the reference 0 group indicated high school graduates or GED).

Results

Baseline SLI Affectedness Comparisons

Table 1 reports the means, standard deviations, and confidence intervals for key variables for the unaffected ($n = 279$) and affected ($n = 240$) groups of participants at the time of initial measurement; the age of participants at first time of measurement ranged from 2;6 to 20;9 years. Recall that low performance on the PPVT was not a criterion for membership in the affected group. At the first time of measurement, 48% of the affected group had PPVT scores ≤ 85 plus an additional 19% had scores in the range 86–90; at 10 years of age, 40% of the affected group had PPVT scores ≤ 85 plus an additional 14% had scores in the range 86–90. Thus, 33% of the affected group had PPVT standard scores above 90 at entry and 46% exceeded 90 at 10 years. Descriptively, on average, the affected group, defined by omnibus language scores and screened for nonverbal IQ scores < 85, performed below the unaffected group on receptive vocabulary, nonverbal IQ, and mother's education, consistent with what is known about groups of children identified as having SLI as defined here. Precise estimates of group differences were obtained from the multilevel models of growth in PPVT that follow.

Sample Characteristics per Occasion of Measurement

Descriptive data on distribution of participants across age levels per time of PPVT measurement are provided in Table 2, which shows the sample characteristics per age level by affectedness group for the 3,012 occasions of measurement. It is evident that the fewest participants are at the youngest and oldest age levels, with more boys in the affected group. Note that the calculations are per age level (i.e., a child can enter the table more than once because PPVT assessments were repeated). A rough index of the relatedness of children within a cell is the number of unique families who had more than one child per cell. Note that given the long time frame of the study the siblings were unlikely to be in the same age cell at the same time but instead were typically separated by several or more calendar years for a given age cell. The estimates of relatedness are duplicated across cells (i.e., the calculations are not adjusted for repeated times of measurement) such that a given child and one or more siblings will generate a count for each age level assessed. There were also many

Table 1. Means (*M*), standard deviations (*SD*), and confidence intervals (CI) for key variables for unaffected and affected participants on initial measurement.

	Unaffected ($n = 279$)			**Affected ($n = 240$)**		
Variables	***M***	**(*SD*)**	**95% CI**	***M***	**(*SD*)**	**95% CI**
PPVT[a]	100.79	(12.71)	[97.2, 100.7]	84.99	(14.04)	[82.6, 86.4]
Omnibus language[b]	104.04	(12.87)	[100.5, 103.6]	77.98	(9.16)	[76.8, 79.3]
Nonverbal IQ[c]	103.78	(10.09)	[101.9, 104.6]	97.40	(9.13)	[96.1, 98.5]
Age, years;months[d]	7;1	(4;2)	[6;2, 7;3]	6;9	(2;11)	[6;1, 6;10]
Mother's education[e]	3.69	(1.36)	[3.5, 3.8]	2.93	(1.31)	[2.7, 3.1]

[a] Lowest of Peabody Picture Vocabulary Test (PPVT)–Revised or PPVT-III. [b] First available of Test of Early Language Development–Third Edition, Test of Language Development–Primary: Second Edition, or Clinical Evaluation of Language Fundamentals–Third Edition. [c] First available of Columbia Mental Maturity Scale or Wechsler Intelligence Scale for Children–Third Edition or Wechsler Adult Intelligence Scale–Third Edition. [d] Age at first time of measurement. [e] Coded where 1 = *some high school, no diploma*; 2 = *high school graduate, diploma, or general equivalency diploma*; 3 = *some college, no degree*; 4 = *bachelor's degree*; 5 = *some graduate work*; 6 = *graduate degree*.

Table 2. Sample characteristics per age level by affectedness group for Peabody Picture Vocabulary Test measurements.

	Measurements, *n*		Girls, *n*		Boys, *n*		
Age level	Unaffected (*n* = 1,484)	Affected (*n* = 1,528)	Unaffected (*n* = 771)	Affected (*n* = 513)	Unaffected (*n* = 713)	Affected (n = 1,015)	Unique families, *n*[a]
2–2;11	35	4	17	3	18	1	5
3–3;11	72	27	45	15	27	12	17
4–4;11	107	73	64	22	43	51	34
5–5;11	127	116	72	39	55	77	53
6–6;11	150	130	83	45	67	85	64
7–7;11	131	145	65	49	66	96	60
8–8;11	142	156	73	51	69	105	69
9–9;11	128	145	62	54	66	91	63
10–10;11	105	129	46	47	59	82	42
11–11;11	94	119	45	37	49	82	46
12–12;11	81	107	36	38	45	69	35
13-13;11	64	88	34	26	30	62	25
14–14;11	64	70	31	20	33	50	23
15–15;11	46	60	19	17	27	43	16
16–16;11	41	52	22	12	19	40	13
17–17;11	32	47	18	14	14	33	12
18–18;11	29	30	20	11	9	19	5
19–19;11	22	19	13	7	9	12	3
20–20;11	14	11	6	6	8	5	1

[a] Family relatedness among participants within the age level was determined by the number of unique families contributing two or more children to the age level. This includes siblings and extended family members such as cousins, nieces, and nephews. Individual family sizes range from two to nine people.

children per cell who did not have siblings in that particular cell and, therefore, would not be captured in this count.

Table 3 addresses the distribution of participants according to the age level at which a child entered the study, separated by affectedness group. It is clear that affected children were likely to enter the study when younger than 11 years, when they are more likely to be identified (although keep in mind that affected siblings who were not ascertained

Table 3. Sample characteristics per age level by first time of measurement and affectedness group for predictor variables.

	Number at 1st time of measurement[a]		Omnibus language, mean		Nonverbal IQ, mean		Mother's education, mean	
Age level	Unaffected	Affected	Unaffected	Affected	Unaffected	Affected	Unaffected	Affected
2–2;11	34	4	111.19	78.00	102.97	97.00	3.94	2.25
3–3;11	46	25	112.38	84.14[b]	101.61	100.68	3.83	3.08
4–4;11	36	49	105.14	79.14	105.31	95.86	3.97	3.16
5–5;11	29	51	102.86	79.41	106.24	97.00	3.90	2.94
6–6;11	29	28	96.69	74.29	106.85	97.76	3.55	2.61
7–7;11	15	19	100.13	76.58	104.91	100.30	3.53	2.63
8–8;11	15	13	98.80	73.15	103.11	96.92	3.07	2.62
9–9;11	12	18	102.67	77.11	103.03	96.02	4.00	3.33
10–10;11	12	10	102.75	73.60	102.08	96.50	3.17	3.10
11–11;11	12	6	95.58	71.17	98.94	97.17	3.50	2.50
12–12;11	10	7	100.70	74.57	96.90	95.19	3.40	2.71
13–13;11	6	3	97.67	84.33	103.83	97.67	4.00	3.33
14–14;11	8	2	101.88	84.00	108.88	106.33	3.63	2.00
15–15;11	2	2	104.50	78.50	96.67	104.00	3.00	4.50
16–16;11	6	2	99.00	70.00	104.83	89.00	3.00	2.50
17–17;11	—	1	—	75.00	—	90.00	—	2.00
18–18;11	3	—	94.33	—	106.00	—	2.33	—
19–19;11	3	—	104.00	—	104.33	—	4.67	—
20–20;11	1	—	103.00	—	125.00	—	2.00	—

Note. Em dashes indicate data not available.

[a] The null value here indicates all participants in this cell had previous times of measurement. [b] Six children were classified as affected on the basis of standard scores below 60 (more than 2 standard deviations below the age mean) on the Rice/Wexler Test of Early Grammatical Impairment (2001), although their omnibus overall standard scores were in the 86–100 range.

from caseloads are also in the affected group); older children at first time of measurement are likely to be siblings. Also reported are the mean scores for omnibus language, nonverbal IQ, and mother's education for the first time of measurement. It is evident that in addition to a lower language score, the affected group had a somewhat lower nonverbal IQ, although well within the normal range, as is often reported in studies of children with SLI. Mother's education for children in the affected group was generally lower than for those in the unaffected group; mothers of affected children were likely to be high school graduates, whereas mothers of unaffected children were likely to have some post–high school education, also consistent with other studies. Table 4 reports the number of children with three or more consecutive PPVT measures per age level, beginning at the first occasion of measurement, as a way of capturing the multiple measurements across the age range. Even at the lowest and highest age ranges, there are at least 23 children with at least 3 years of consecutive times of measurement; at 6 years, there are at least 229 children. Overall, the sample size and distribution across groups, ages and occasions of measurement, and familial relationships are robust for the methods of analyses.

Question 1: Unconditional Growth Models of Growth in PPVT

Overall, there were 519 unique participants with 3,012 PPVT latent trait outcomes, with an average of about seven occasions of measurement per participant. Growth in PPVT outcomes was assessed using multilevel models in which time was nested within persons, which was nested within families, creating a three-level model. All models were estimated using maximum likelihood and Satterthwaite denominator degrees of freedom within SAS Proc Mixed Version 9.3. The significance of individual fixed effects was evaluated via their Wald test *p* values; the significance of multiple fixed effects or of random effects variances and covariances was evaluated via $-2\Delta LL$ tests (i.e., likelihood ratio tests using degrees of freedom equal to the difference in the number of estimated parameters). Linear combinations of model fixed effects (i.e., to obtain simple effects within interaction terms) were obtained via estimate statements, which then provide an estimate, standard error, and corresponding *p* value for all requested model-implied fixed effects.

Given that standardized coefficients do not exist for three-level models such as ours, effect sizes are often created as the proportion of each variance component accounted for by the predictors (i.e., pseudo-R^2; Singer & Willett, 2003). However, because our conditional growth models resulted in the estimation of six separate variance components across levels, an alternative, more transparent measure of effect size was selected, which has been described as total R^2 in multilevel models for longitudinal data (Hoffman, 2014). Thus, total R^2 for overall PPVT variance explained was calculated as the square of the correlation between the actual PPVT outcomes and the PPVT outcomes predicted by the model fixed effects (i.e., analogous to a traditional single-level regression). This approach has the advantage of providing a single R^2 estimate for each model and will remain positive—unlike pseudo-R^2 estimates, which can become negative for a variety of reasons (Singer & Willett, 2003).

With respect to the modeling sequence, an empty means (i.e., random intercept only) model was first estimated to describe the variation in PPVT across levels of analysis.

Table 4. Number of children with three or more consecutive Peabody Picture Vocabulary Test (PPVT) measures.

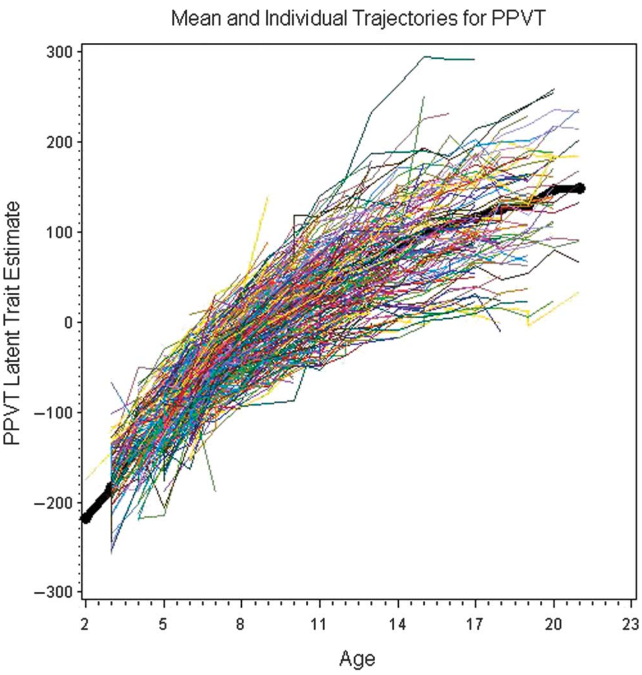
Age level, years;month	**Overall, *n***		**Girls, *n***		**Boys, *n***	
	Unaffected	**Affected**	**Unaffected**	**Affected**	**Unaffected**	**Affected**
2–2;11	28	3	16	2	12	1
3–3;11	64	23	43	13	21	10
4–4;11	89	65	53	19	36	46
5–5;11	98	101	53	35	45	66
6–6;11	118	111	64	40	54	71
7–7;11	101	115	53	39	48	76
8–8;11	101	111	51	36	50	75
9–9;11	74	102	37	37	37	65
10–10;11	51	89	25	32	26	57
11–11;11	49	83	27	24	22	59
12–12;11	50	57	23	15	27	42
13–13;11	36	57	17	13	19	44
14–14;11	40	51	21	14	19	37
15–15;11	25	41	12	15	13	26
16–16;11	24	27	15	7	9	20
17–17;11	18	17	9	7	9	10
18–18;11	13	10	7	7	6	3
19–19;11	—	—	—	—	—	—
20–20;11	—	—	—	—	—	—

Note. Em dashes indicate data not available. In these age levels, it was not possible for children to have three PPVT measures given that the PPVT was administered annually.

Of the total PPVT variance, 53% was over time within persons (Level 1), 34% was across persons within families (Level 2), and 13% was between families (Level 3). Said differently, the intraclass correlation reflecting the proportion of PPVT variance due to between-persons mean differences (at Levels 2 and 3) was .47; of that 47%, the intraclass correlation for the proportion actually due to between-families mean differences (at Level 3) was .28. Each of these intraclass correlations was significantly greater than 0, $-2\Delta LL(1) = 870.8$ and $-2\Delta LL(1) = 18.3$, *p*s < .001, indicating the need for a three-level model for time within person within family in modeling growth in PPVT latent trait outcomes (i.e., as opposed to a single-level or two-level model, respectively).

Individual PPVT growth trajectories (on a latent trait scale of $M = 1.88$, $SD = 88.20$; recall that the ability estimates were multiplied by 100 for ease of variance estimation) are shown in Figure 1, in which the heavy black represents the sample average. On the basis of these trends, polynomial random effects models were examined to describe quadratic growth over exact age centered at 10 years. Relative to a baseline model of fixed linear and quadratic effects of age with random intercepts for person and family, significant differences were found in linear growth across persons, $-2\Delta LL(2) = 998.9$, $p < .001$, and across families, $-2\Delta LL(2) = 20.9$, $p < .001$. Significant differences in quadratic growth were also found across persons, $-2\Delta LL(3) = 42.06$, $p < .001$, but not across families, $-2\Delta LL(3) = 1.57$, $p = .666$. Thus, although there were individual differences across persons in intercept, linear growth, and quadratic growth, there was only a shared family component for the intercept and rates of linear growth. The fixed effects of linear and quadratic age accounted for 79% of the total PPVT variance.

Figure 1. Mean and individual growth trajectories for PPVT latent trait estimates by age.



However, because participants ranged in age from 2;6 to 20 years at the first occasion, 40% of the variation in age was actually due to cross-sectional differences. Accordingly, a time-invariant continuous predictor of age cohort (created from the participant's age at the first measurement occasion, centered at 7 years) was used to represent the potential differential effects of this cross-sectional age variance (i.e., age contextual effects; Hoffman, Hofer, & Sliwinski, 2011; Sliwinski, Hoffman, & Hofer, 2010). Thus, the quadratic growth model was augmented to include fixed linear and quadratic effects of age cohort on the intercept, linear age slope, and quadratic age slope, which significantly improved model fit, $-2\Delta LL(6) = 37.7$, $p < .001$, and accounted for an additional 1% of the total PPVT variance. These age cohort effects are illustrated using prototypical age cohort values of the reference age 7 and ± 1 *SD* of age at entry in Figure 2, and the final unconditional growth model parameters are provided in the first set of columns in Table 5. As shown, a child who began the study at age 7 years (age cohort = 0) has an expected PPVT latent trait outcome of 15.13 with an instantaneous linear rate of growth per year of 20.05 at age 10 years. This linear rate of growth becomes less positive per year by twice the quadratic coefficient of −0.70, indicating a positive average function that decelerates at a constant rate over age. The age cohort effects then indicate that children who began the study at older ages have slightly (but significantly) lower predicted PPVT scores with more positive rates of growth at age 10 years but with slightly greater predicted rates of deceleration across ages. This unconditional model, including quadratic effects of age and age cohort, as reported in Table 5, was used as the baseline to which all subsequent time-invariant predictors were added, as described next. As derived using the formulas presented in Hoffman (2014, Chapter 6), model-based correlations of PPVT across ages 4–20 years as predicted by individual (Level 2) variance in the intercept,

Figure 2. Effects of age cohort on PPVT latent trait estimates from unconditional model.

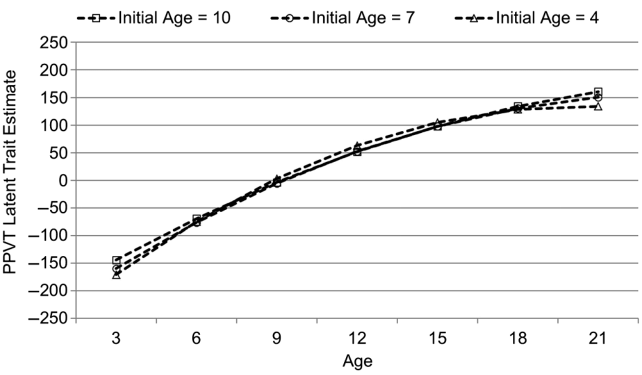


Table 5. Peabody Picture Vocabulary Test growth model parameters (bold values indicate $p < .05$).

	Unconditional Model			Final Model		
Model Parameters	**Est**	***SE***	***p* <**	**Est**	***SE***	***p* <**
Fixed effects:						
Intercept (at age 10 years)	**15.13**	2.33	.001	**27.33**	2.85	.001
Linear age slope	**20.05**	0.35	.001	**20.86**	0.54	.001
Quadratic age slope	**−0.70**	0.05	.001	**−0.72**	0.05	.001
Linear age cohort on intercept	**−1.55**	0.61	.011	**−1.24**	0.48	.011
Linear age cohort on linear slope	**−0.31**	0.15	.037	0.26	0.14	.075
Linear age cohort on quadratic slope	**0.08**	0.02	.001	−0.22	0.14	.131
Quadratic age cohort on intercept	**0.59**	0.16	.001	0.02	0.03	.444
Quadratic age cohort on linear slope	0.00	0.03	.956	**0.07**	0.02	.001
Quadratic age cohort on quadratic slope	**−0.01**	0.00	.007	**−0.01**	0.00	.001
SLI affectedness on intercept				**−30.00**	2.34	.001
SLI affectedness on linear slope				**−1.23**	0.43	.005
Boys versus girls on intercept				**−5.11**	2.25	.024
Boys versus girls on linear slope				**−1.75**	0.41	.001
Maternal education on intercept				**4.96**	1.02	.001
Maternal education on slope				**0.44**	0.18	.015
Nonverbal IQ on intercept				**0.86**	0.12	.001
Nonverbal IQ on slope				0.01	0.02	.566
Nonverbal IQ on quadratic slope				**0.01**	0.00	.001
Variance components:						
Residual variance	**169.71**	5.56	.001	**170.42**	5.57	.001
Level 2 intercept variance	**652.41**	71.10	.001	**364.92**	42.19	.001
Level 2 linear age slope variance	**6.44**	1.81	.001	**4.94**	1.52	.001
Level 2 quadratic age slope variance	**0.08**	0.03	.001	**0.07**	0.02	.001
Level 2 intercept–linear covariance	**43.77**	8.99	.001	**31.34**	6.39	.001
Level 2 intercept–quadratic covariance	−0.50	1.04	.630	−1.46	0.81	.072
Level 2 linear–quadratic covariance	−0.15	0.13	.255	−0.12	0.11	.269
Level 3 intercept variance	**484.10**	93.88	.001	**203.59**	41.75	.001
Level 3 linear age slope variance	**6.88**	1.79	.001	**6.34**	1.63	.001
Level 3 intercept–linear covariance	**42.92**	10.59	.001	**27.18**	6.69	.001
ML −2LL	26207			25825		
AIC	26245			25881		
BIC	26313			25980		
Total R^2	.798			.891		

ML = maximum likelihood; LL = log likelihood; AIC = Akaike information criterion; BIC = Bayesian information criterion.

linear age slope, and quadratic age slope ranged from $r = .46$ to $r = .92$ with an average of approximately $r = .70$, suggesting relative stability in individuals' rank order of PPVT across ages.

Questions 2–4: Conditional Models Predicting Growth in PPVT

Main effects and interactions with linear and quadratic age of the time-invariant predictors of interest (SLI affectedness, gender, maternal education, and nonverbal IQ) were added in sequential models. The final model, including only significant effects or necessary lower-order effects, is reported in the second set of columns in Table 5. The results from this full model that pertain to each research question are presented herein; the process of building it can be summarized as follows. First examined were effects of the SLI affectedness group (in which unaffected children served as the reference group), which accounted for another 6.5% of the total PPVT variance. Effects of gender (in which boys served as the reference group) were then added, which accounted for another 0.09% of the total PPVT variance. Finally, effects of maternal education were examined, which accounted for another 1.3% of the total PPVT variance, followed by nonverbal IQ, which accounted for another 1.6% of the total PPVT variance, resulting in a total $R^2 = .89$. Although tested, no interactions among SLI affectedness, gender, maternal education, and nonverbal IQ were significant, indicating only additive effects for these predictors. Therefore, in the figures that follow, all nonincluded predictors were held constant at their centered 0 values.

With respect to Questions 2 and 3, the model-predicted results of the additive effects of SLI affectedness and gender on growth in PPVT across ages are shown in Figure 3. Relative to the reference group of unaffected children, children affected by SLI scored significantly lower on the PPVT at age 10 years and had a significantly less positive linear rate of growth on the PPVT at age 10 years but did not differ

Figure 3. Effects of SLI affectedness and gender on PPVT latent trait estimates from final model.

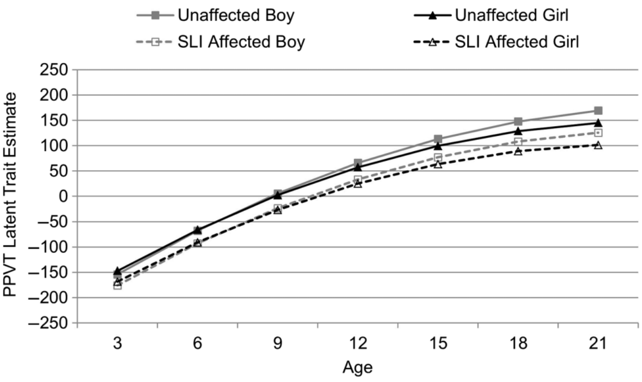
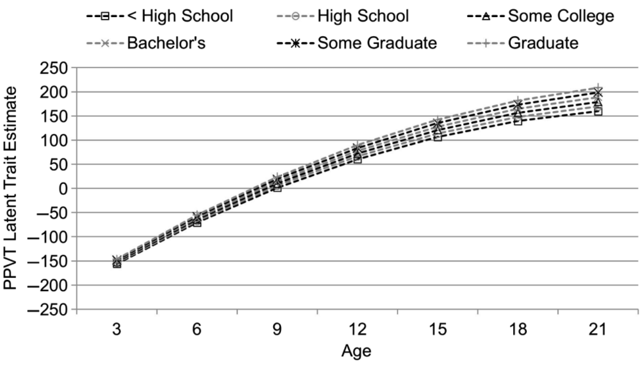


Figure 4. Effects of maternal education on PPVT latent trait estimates from final model.



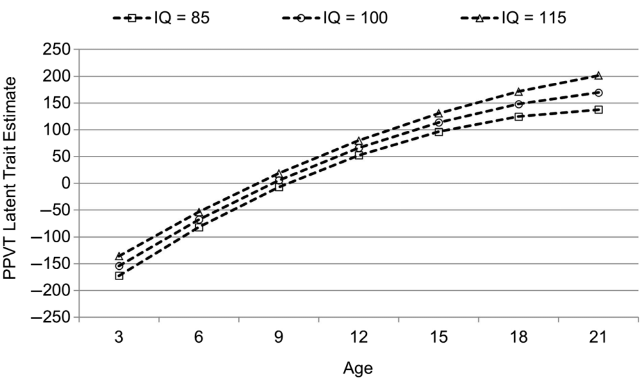
in their rates of deceleration across ages. Said differently, although model-implied simple effects of SLI affectedness were significant at all ages, the SLI effect became larger linearly with age. With respect to gender, relative to the reference group of boys, girls scored significantly lower on the PPVT at age 10 years and had a significantly less positive linear rate of growth in PPVT at that age but did not differ in their rates of deceleration across ages. Model-implied simple effects revealed that the gender difference in PPVT was predicted to significantly favor girls from ages 3 to 4 years and to be nonsignificant from ages 5 to 9 years but to significantly favor boys from ages 10 to 21 years. The result of this gender gap at older ages is that unaffected boys at age 18 years had a similar predicted vocabulary to that of unaffected girls at age 21 years (difference = 2.86, $p = .65$). However, the net result of the additive effects of both SLI and gender were evidenced by the relatively poor outcomes for affected girls at age 21 years, whose predicted vocabulary was on par with that of unaffected girls at age 15 years (difference = 1.86, $p = .80$). In comparison, predicted vocabulary for affected boys at age 21 years was on par with that of unaffected girls at age 18 years (difference = 3.02, $p = .70$).

With respect to Question 4, there were significant additive effects of maternal education and nonverbal IQ on growth in PPVT across ages. As shown via the model-predicted results in Figure 4, children whose mothers had higher levels of education were significantly higher on the PPVT at age 10 years and had a significantly more positive linear slope at age 10 years but did not differ in their rates of deceleration. As shown via the model-predicted results in Figure 5, children with higher nonverbal IQ were significantly higher at age 10 years but did not have a significantly different linear rate of growth as evaluated at age 10 years, although they did have a significantly less negative rate of deceleration. The model-implied simple effect of IQ was significant at all ages, although it had a smaller effect initially and then increased again after age 11 years (especially at older ages, as created by the IQ × Quadratic Age interaction).

Summary and Discussion

This study is the first longitudinal investigation to model the acquisition of receptive vocabulary in children with and without SLI over the nearly two decades in which the base of the adult vocabulary level is established, from age 2 to 21 years. The first research question addressed growth trajectories of receptive vocabulary across the age range of 2;6–21 years. Although our results documented strong growth within persons with significant individual differences, the rank ordering of individuals also showed strong stability across ages. Our study design also allowed examination of between-families sources of variance in addition to between-persons variance. Of the between-persons mean differences found in the empty model (47% of the total PPVT variance), almost half (28%) was due to between-families mean differences. Individual differences across persons were evident in the intercept, linear growth, and quadratic growth; their fixed effects of age accounted for 79% of the total PPVT variance. Yet the time of entry into the study also revealed important differences, such that children who were older at study entry had lower ability at

Figure 5. Effects of nonverbal IQ on PPVT latent trait estimates from final model.



age 10 years but more positive rates of linear change with greater deceleration over age (as shown in Figure 2).

The second research question addressed whether there were differences in growth between children with and without SLI. This group effect, examined in a model also including the effects of change over time, familial relationships, and age at entry, accounted for another 6.5% of the total PPVT variance. The overall shape of change over time was highly similar for the affected and unaffected groups, with strong linear effects in early childhood that decelerated in later childhood. However, the SLI group did not catch up with the unaffected group over time. Instead, the SLI group had a lower level of performance at the age 10 years intercept for receptive vocabulary and remained lower through a reduced rate of linear growth relative to unaffected children. The persistent differentiation of groups over time suggests that the study's operational definition of affectedness was appropriate for detecting systematic patterns of change over time.

Answers to Questions 3 and 4 revealed less robust, even minimal, but statistically significant effects for gender ($\Delta R^2 = 0.09\%$), maternal education ($\Delta \mathrm{R}^2 = 1.3\%$), and nonverbal IQ ($\Delta R^2 = 1.6\%$). These effects were additive across groups. Although the effects were small, the cumulative consequences are noteworthy. The model revealed that the vocabulary trait levels of affected girls at age 21 years were similar to those of unaffected girls at age 15 years; affected boys at age 21 years were similar to unaffected girls at age 18 years; and unaffected girls at age 21 years were similar to unaffected boys at age 18 years. These outcomes require revision of the original hypotheses. The first hypothesis predicted that vocabulary growth would be strongly linear over time, a prediction informed by the previous reports of cross-sectional modeling of PPVT scores over this age range that show strong linear effects and little evidence of quadratic trends or deceleration until around 20 years. Contrary to this prediction, the longitudinal vocabulary trait model reveals a never-before-reported quadratic trend whose deceleration effect was manifested at a much younger age for both unaffected and affected children. This age period is also associated with a reduction in second-language learning abilities (Genesee, Paradis, & Crago, 2004), suggesting influences tied to early adolescence or the pubescent period, a time of rapid change across biological, social, and cognitive abilities. These findings suggest that vocabulary acquisition of a native language seems to show the same age vulnerability around adolescence. An obvious conclusion is that it is inaccurate to project long-term outcomes in individual differences in vocabulary acquisition from longitudinal studies that end around 8–10 years of age.

The second hypothesis predicting a shallower slope of vocabulary acquisition for the SLI group throughout the time interval was supported. The SLI group generally mirrored the growth trajectory of the unaffected group albeit at a lower level of performance. Yet, as shown in Figure 3, the SLI deficit became larger linearly with age. So, though for much of the time period the way in which change occurs over time is highly similar for both groups, the ultimate effect of the linear increase in the effect of SLI with age is that the SLI group falls further behind.

As noted in the Introduction, it is possible that children with SLI may have a disadvantage for vocabulary acquisition in middle elementary grades if their reading abilities are not robust (Adlof & Perfetti, 2013). This could contribute to the significantly less positive linear rate of growth in PPVT for the SLI group relative to the unaffected group. Under a strong interpretation of the effect of poor reading on vocabulary acquisition, a deceleration in vocabulary growth could be expected for the SLI group. The lack of an SLI × Quadratic Age interaction is inconsistent with this prediction; that is, the rate of deceleration in vocabulary growth from 10 to 21 years for the SLI group was not greater than for the unaffected group. The relationship of reading and PPVT trait scores over time in the two groups warrants further study, which we plan to do in a follow-up study.

Our hypothesis that effects of maternal education would interact with group over time was also not upheld. There were modest, but additive, effects for maternal education, with a more positive linear slope corresponding to greater maternal education. The generalization is that there is a modest advantage in vocabulary acquisition for children of mothers with higher education, for unaffected as well as affected children, an outcome consistent with many reports in the literature. Interpretation is complicated by the fact that children share inherited as well as environmental influences with their parents and siblings. In this study, familial effects are treated as additional variance components in the model and maternal education as a predictor of outcomes. Maternal education is likely to contribute to familial relationships in ways not revealed in the models used here. Additional studies will be needed to differentiate the effects of shared inherited ability to acquire new words versus shared family environment.

Our hypothesis that there was no lasting advantage throughout childhood for girls' receptive vocabulary development was supported by a statistically small additive effect that played out in significant ways over time. The upshot is that girls with SLI are at the greatest disadvantage for accumulated receptive vocabulary acquisition at 21 years, as a consequence of a gender gap affecting both groups that appears to open up from age 12 years onward due to a significantly less positive linear slope for girls than for boys. This leads to a marked disparity for 21-year-old girls with SLI when compared with their age peers who had reached the same level of performance when they were age 15 years. Explanations for a higher risk of late vocabulary development of boys in infancy have focused on possible effects of testosterone levels in umbilical cord blood (Whitehouse et al., 2012). This suggests that hormonal influences could have a negative effect on vocabulary growth in preadolescent girls, although our literature review did not yield any relevant studies. It is also possible that the gender differences beginning in preadolescence that work against the girls with SLI may be related to reports from a large longitudinal study that girls with language impairments are at risk for adverse social outcomes in adolescence and early adulthood, suggesting that language impairments in adolescence expose girls disproportionately to social risk (Brownlie, Jabbar, Beitchman, Vida, & Atkinson, 2007; Johnson et al., 1999; Johnson, Beitchman, & Browlie, 2010). The social risk, in turn, could influence vocabulary acquisition if the girls were less likely to aim for higher education or social status.

The final hypothesis, that children with higher levels of nonverbal IQ would have a persistent advantage for receptive vocabulary growth, was supported. The effect of nonverbal IQ was significant at all ages, across both groups, with an advantage for children with higher levels. The new finding here is the outcome showing that the IQ effect increased across ages, especially at older ages, as reflected by the significant IQ × Quadratic Age interaction. This may be related to the development of executive function skills associated with nonverbal IQ that accumulate with age in ways that enhance receptive vocabulary acquisition.

Collectively, the outcomes of the study have implications for hypothesized causal pathways for individual differences in vocabulary acquisition. A growth perspective reveals age-related effects that could lead to misleading generalizations from an age-stratified sample. Overall, the full arc of receptive vocabulary acquisition over childhood suggests three phases of receptive vocabulary development: 5 years and under, 6–10 years, and adolescence into adulthood, with different gender effects over time. Girls, for example, may be more likely to be identified with low levels of vocabulary when they reach adolescence than when younger. Any gender effects on the causal pathways deduced from a young sample may not be the same as those from an older sample, and possible related socially mediated effects might be missed. Another caveat is that nonverbal IQ and maternal education do not seem to play a different role in vocabulary growth in children with SLI than in children without SLI, as demonstrated by minimal (at best) significant interactions with SLI in the present study. A further caveat provided by the current study is that the adolescent period shows gradual weakening of growth for the children with SLI relative to age peers, suggesting the need for a better understanding of any ways in which causal pathways operative early on may change with adolescence.

Another caution is that although vocabulary acquisition is a key language outcome, it is not a surrogate index because all dimensions of language change over time. An investigation of children with and without SLI documented the way the grammatical property of finiteness-marking changes from ages 6 to 16 years (Rice et al., 2009). Overall, the growth trajectory for the grammar marker was not similar to the generally linear vocabulary trajectories documented here. Furthermore, evaluation of PPVT, nonverbal intelligence (WISC or CMMS), and maternal education as predictors of the growth parameters in the models found no significant predictive relationships for either group for any of the grammar variables. Thus, the predictors and shape of change over time of vocabulary development should not be seen as operative across all dimensions of language acquisition. Even with these differences across the linguistic dimensions, several characteristics were shared by growth in vocabulary and finiteness marking: On average, for both linguistic dimensions, children with SLI demonstrated growth trajectories that paralleled their peers, although at a lower level of proficiency that did not resolve with age, followed by a deceleration that appeared around ages 10–12 years and left the children with SLI with lower proficiency than their age peers as they moved into adulthood. In the context of genetics investigations, Rice (2012, 2013) proposes inherited maturational effects that operate across dimensions of language that interact with predictors in different ways.

The limitations of the study are related to the means of vocabulary measurement, sampling, and modeling methods. The PPVT vocabulary test is one way to measure children's vocabulary growth. Other ways to measure the construct of vocabulary may or may not reveal the same growth patterns. A population-based sample of a larger number of participants would yield results generalizable to a broader population of children and greater power for detection of possible interactions of age, gender, and other predictors over time. The modeling methods, although highly informative, do not determine causal relationships that account for change over time.

The outcomes highlight the relative gap in our literature documenting the adolescent period for children with SLI and the need for longitudinal studies of this sort. It is clear that the onset of adolescence brings significant slowing of vocabulary acquisition for children with and without SLI, and gender plays an unexpected role, especially for girls with SLI. Given the great significance of vocabulary development for children's academic achievement as well as its likely contributing role to social relationships in adolescence and beyond, it is of high importance to learn more about the intrinsic mechanisms and external influences that drive longitudinal change to formulate effective clinical treatment programs. Ultimately, the development of etiological models of SLI must reckon with the shape and timing of change over the long term, similarities and differences across linguistic dimensions, and age-referenced dynamic shifts in predictor relationships. Formulation of accurate causal models of individual differences in language acquisition demand nothing less.

Acknowledgments

This work was made possible by National Institutes of Health Grants RO1DC001803, P30DC005803, and P30HD002528 awarded to Mabel L. Rice. We express special appreciation to Denise Perpich for her data management of the longitudinal study. We thank the children and families who participated in the study and appreciate their long-term commitment to the research project.

References

Adlof, S. M., & Perfetti, C. A. (2013). Individual differences in word learning and reading ability. In A. Stone, B. Ehren, E. Silliman, & G. Wallach (Eds.), *Handbook of language and* *literacy development and disorders* (2nd ed., pp. 246–264). New York, NY: Guilford.

Astrom, R. L., Wadsworth, S. J., Olson, R. K., Willcutt, E. G., & DeFries, J. C. (2011). DeFries–Fulker analysis of longitudinal reading performance data from twin pairs ascertained for reading difficulties and from their nontwin siblings. *Behavior Genetics, 41*(5), 660–667.

Banach, R., Thompson, A., Szatmari, P., Goldberg, J., Tuff, L., Zwaigenbaum, L., & Mahoney, W. (2009). Brief report: Relationship between non-verbal IQ and gender in autism. *Journal of Autism & Developmental Disorders, 39,* 188–193.

Bishop, D. V. M., Price, T. S., Dale, P. S., & Plomin, R. (2003). Outcomes of early language delay: II. Etiology of transient and persistent language difficulties. *Journal of Speech, Language, and Hearing Research, 46,* 561–575.

Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24,* 267–304.

Brownlie, E. B., Jabbar, A., Beitchman, J. H., Vida, R., & Atkinson, L. (2007). Language impairment and sexual assault of girls and women: Findings from a community sample. *Journal of Abnormal Child Psychology, 35,* 618–626.

Burgemeister, B. B., Blum, L. H., & Lorge, I. (1972). *The Columbia Mental Maturity Scale*. San Antonio, TX: The Psychological Corporation.

Catts, H. W. (2004). Language impairments and reading disabilities. In R. D. Kent (Ed.), *The MIT encyclopedia of communication disorders* (pp. 329–331). Cambridge, MA: MIT Press.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49,* 278–293.

Catts, H. W., & Kamhi, A. (2005). *The connections between language and reading disabilities*. Mahwah, NJ: Erlbaum.

Conti-Ramsden, G., St. Clair, M. C., Pickles, A., & Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: From childhood to adolescence. *Journal of Speech, Language, and Hearing Research, 55,* 1716–1735. doi:10.1044/1092-4388(2012/10-0182)

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22,* 8–15.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46,* 544–560.

DeThorne, L. S., Petrill, S. A., Hayiou-Thomas, M. E., & Plomin, R. (2005). Low expressive vocabulary: High heritability as a function of more severe cases. *Journal of Speech, Language, and Hearing Research, 48,* 792–804.

Dunn, A., & Dunn, A. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Third Edition*. Circle Pines, MN: American Guidance Service.

Ellis Weismer, S., & Hesketh, L. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research, 36,* 1013–1025.

Ellis Weismer, S., & Hesketh, L. (1998). The impact of emphatic stress on novel word learning by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1444–1458.

Entwisle, D. R., & Astone, N. M. (1994). Some practical guidelines for measuring youths' race/ethnicity and socioeconomic status. *Child Development, 65,* 1521–1540.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories*. Baltimore, MD: Brookes.

Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual language development & disorders: A handbook on bilingualism and second language learning* (Vol. 11). Baltimore, MD: Brookes.

Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47,* 1117–1132.

Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language, and Hearing Research, 48,* 1452–1467.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30,* 196–206.

Hoffman, L. (2014). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York, NY: Routledge Academic.

Hoffman, L., Hofer, S. M., & Sliwinski, M. (2011). On the confounds among retest gains and age-cohort differences in the estimation of within-person change in longitudinal studies: A simulation study. *Psychology and Aging, 26,* 778–791.

Hoffman, L., Templin, J., & Rice, M. L. (2012). Linking outcomes from *Peabody Picture Vocabulary Test* forms using item response modeling. *Journal of Speech, Language, and Hearing Research, 55,* 754–763. doi:10.1044/1092-4388(2011/10-0235)

Hresko, W. P., Reid, D. K., & Hammill, D. D. (1999). *Test of Early Language Development–Third Edition*. Austin, TX: Pro-Ed.

Johnson, C. J., Beitchman, J. H., & Browlie, E. G. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology, 19,* 51–65.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., . . . Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42,* 744–760.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137,* 267–296.

Nagy, W. E., & Anderson, R. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19,* 304–330.

Nagy, W. E., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly, 20,* 233–253.

Newcomer, P. L., & Hammill, D. D. (1988). *Test of Language Development–Primary: Second Edition*. Austin, TX: Pro-Ed.

Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick incidental learning (QUIL) of words by school-aged children with and without SLI. *Journal of Speech and Hearing Research, 38,* 434–445.

Olson, R. K., Hulslander, J., Christopher, M., Keenan, J. M., Wadsworth, S. J., Willcutt, E. G., . . . DeFries, J. C. (2013). Genetic and environmental influences on writing and their relations to language and reading. *Annals of Dyslexia, 63*(1), 25–43.

Pankratz, M., Morrison, A., & Plante, E. (2004). Difference in standard scores of adults on the Peabody Picture Vocabulary Test (revised and third edition). *Journal of Speech, Language, and Hearing Research, 47,* 714–718.

Pena, E. D., Spaulding, T. J., & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology, 15,* 247–254.

Reilly, S., Wake, M., Ukoumunne, O., Bavis, E., Prior, M., Cini, E., . . . Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from Early Language in Victoria Study. *Pediatrics, 126,* 1530–1537.

Rice, M. L. (2012). Toward epigenetic and gene regulation models of specific language impairment: Looking for links among growth, genes, and impairments. *Journal of Neurodevelopmental Disorders, 4*(27). doi:10.1186/1866–1955-4-27

Rice, M. L. (2013). Language growth and genetics of specific language impairment. *International Journal of Speech-Language Pathology, 15,* 223–233.

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word learning abilities of language delayed preschoolers. *Journal of Speech and Hearing Disorders, 55,* 33–42.

Rice, M. L., Buhr, J. C., & Oetting, J. B. (1992). Specific language impaired children's quick incidental learning of words: The effect of a pause. *Journal of Speech and Hearing Research, 35,* 1040–1048.

Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: A study of growth and asymptote. *Journal of Speech, Language, and Hearing Research, 52,* 1417–1433.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research, 37,* 106–122.

Rice, M. L., Smith, S. D., & Gayán, J. (2009). Convergent genetic linkage and associations to language, speech and reading measures in families of probands with specific language impairment. *Journal of Neurodevelopmental Disorders, 1,* 264–282.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53,* 1–17.

Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: The Psychological Corporation.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1412–1431.

Rice, M. L., Wexler, K., & Redmond, S. M. (1999). Grammaticality judgments of an extended optional infinitive grammar: Evidence from English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 943–961.

Rice, M. L., Zubrick, S. R., Taylor, C. L., Gayán, J., & Bontempo, D. E. (2014). Late language emergence in 24 month twins: Heritable and increased risk for twins. *Journal of Speech, Language, and Hearing Research, 57*(3), 917–928. doi:10.1044/1092-4388(2013/12-0350)

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83,* 508–525. doi:10.1111/j.1467-8624.2011.01710x

Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals–Third Edition*. San Antonio, TX: The Psychological Corporation.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Sliwinski, M. J., Hoffman, L., & Hofer, S. M. (2010). Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. *Research in Human Development, 7,* 45–60.

Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., & Zubrick, S. R. (2013). Risk factors for children's receptive vocabulary development from four to eight years in the Longitudinal Study of Australian Children. *PLoS ONE, 8*(9), e73046.

Ukrainetz, T. A., & Duncan, D. S. (2000). From old to new: Examining score increases on the *Peabody Picture Vocabulary Test–III. Language, Speech, and Hearing Services in Schools, 31,* 336–339.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–Third Edition*. San Antonio, TX: The Psychological Corporation.

Whitehouse, A. J. O., Mattes, E., Maybery, M. T., Sawyer, M. G., Jacoby, P., Keelan, J. A., & Hickey, M. (2012). Sex-specific associations between umbilical cord blood testosterone levels and language delay in early childhood. *Journal of Child Psychology and Psychiatry, 53,* 726–734.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50,* 1562–1592.