

---

# Language Outcomes of 7-Year-Old Children With or Without a History of Late Language Emergence at 24 Months

**Mabel L. Rice**

University of Kansas, Lawrence

**Catherine L. Taylor**

**Stephen R. Zubrick**

Curtin University, Perth, Australia

---

**Purpose:** The aim of this study was to investigate the language outcomes of 7-year-old children with and without a history of late language emergence at 24 months.

**Method:** One hundred twenty-eight children with a history of late language emergence (LLE) at 24 months and 109 children with a history of normal language emergence (NLE) at 24 months participated in direct behavioral assessment of multiple dimensions of language at 7 years. The children were recruited from a prospective cohort study of 1,766 epidemiologically ascertained 24-month-old singleton children.

**Results:** The group mean for the LLE children was within the typical range on an omnibus measure of general language ability and measures of specific dimensions of language. However, a greater percentage of LLE children, relative to NLE children, performed below normative expectations on a measure of general language ability (20% versus 11%), speech (7% versus 2%), syntax (18% versus 8%), and morphosyntax (9%–23% versus 2%–14%), but not vocabulary or semantics.

**Conclusion:** The results provide support for growth models of language impairment that predict that late onset of language foretells a protracted growth difference for some LLE children relative to NLE children, particularly for syntax and morphosyntax.

**KEY WORDS:** late talking outcomes, specific language impairment, longitudinal study

---

The emergence of language during the toddler period is one of the most striking accomplishments in young children's development. Children are known to show considerable variation in onset times, with some children beginning to talk much later than others. Previous studies have benchmarked 24 months as a time when delayed language emergence can be reliably ascertained. Among the questions of interest is whether children with late language emergence (LLE), also known as "late talking," recover from the late start-up or whether a late start is prognostic of a longer lasting risk for language acquisition. A number of relatively small-scale studies have compared children with a positive history with control groups of children matched for gender, socioeconomic status (SES), and nonverbal intelligence. The outcomes suggest long-term risk, controlling for other factors, although the picture is complicated by indications that language outcomes are differentiated by linguistic dimensions and by other methodical limitations, such as small samples and differences across studies in measurement methods.

More recently, growth modeling studies of preschool children with specific language impairment (SLI) have documented prolonged delays and disruptions in linguistic growth over a period of years (Rice, Redmond, & Hoffman, 2006; Rice, Tomblin, Hoffman, Richman, & Marquis, 2004;

Rice & Wexler, 1996b; Rice, Wexler, & Hershberger, 1998). Importantly, the models project a delayed onset of language for affected children, relative to control children, corroborating clinical reports of delayed language emergence as an integral part of the SLI phenotype. What is needed is more information to bridge the gap between language onset and later language acquisition, to shed light on linguistic differences in risk for persistent delay, and to clarify the growth trajectories and related growth mechanisms from onset to later language milestones, including precise indicators of linguistic outcomes indexed to clause structure.

This study documents the language outcomes at 7 years of two groups of children for whom there is concurrent report of language emergence or LLE at 24 months of age. The participants were drawn from a larger population-based sample of children. One group was identified as LLE at 24 months; a control group, which was equivalent to the LLE group on a number of potential covariates, was ascertained. Thus, this study provides information from larger and more diverse samples than have been reported previously. In addition, this study compares two methods of identifying LLE and includes outcome measures across several language domains and language sample data as well as test response data.

### ***Previous Studies of Late Language Emergence: Measurement, Samples, and Outcomes***

Previous studies of late talkers have benchmarked delayed onset according to two indices drawn from children's language production at age 24 months. One is the total vocabulary size reported by parents. Two widely used parent-report measures of total vocabulary size are the Language Development Survey (LDS; Rescorla, 1989) and the MacArthur Communicative Development Inventories: Words and Sentences (CDI:WS; Fenson et al., 1993). Two approaches have been used to differentiate children with language delay from children with typical language development on the basis of expressive vocabulary. The first is to use vocabulary cut-points derived from other studies with established concurrent validity (cf. Klee et al., 1998; Rescorla, 1989). This criterion has yielded estimates of 10%–19% of children as late talkers (Klee et al., 1998; Rescorla, 1989, 2002; Rescorla, Hadick-Wiley, & Escarce, 1993). The second approach is to identify children whose vocabulary score places them in the bottom of the distribution for vocabulary development in a particular sample. Fenson et al. (1994), Dale et al. (1998), and Ellis Weismer (2007) identified children with language delay at 24 months on the basis of a vocabulary score on the CDI:WS at or below the 10th percentile.

Rescorla and Achenbach (2002) recommended using a vocabulary score on the LDS at or below the 15th percentile to identify children with language delay in the 18- to 23-month age range. Feldman, Campbell, Kurs-Lasky, and Rockette (2005) used five levels to define language delay, the 5th, 10th, 15th, 20th, and 25th percentiles. The vocabulary score that places children in the low range of the distribution varies across samples and according to the parent-report measure used (cf. Rescorla, Ratner, & Jusczyk, 2005). For example, in Klee et al.'s (1998) sample, a vocabulary score of 62 words on the LDS placed children at the 10th percentile. In Fenson et al.'s (1993) sample, a vocabulary score of 79 words for boys and 119 words for girls on the CDI:WS placed them at the 10th percentile.

An alternative index is the presence of 2–3 word combinations, also as reported by parents. Children with an expressive vocabulary of fewer than 50 words or who were not combining words by 24 months represented the bottom 10% of the normative 24-month sample for the Communicative Development Inventories (Fenson, 1993). An important limitation of these studies is that the participants were drawn from convenience samples of primarily middle- or upper class children.

A recent study provides the first population-based estimate of LLE. Zubrick, Taylor, Rice, and Slegers (2007) investigated language emergence in a sample of 1,766 children at age 24 months that was drawn from an epidemiologically ascertained sample of infants who participated in an ongoing study of health outcomes. The language index was based on the six-item Communication Scale of the Ages and Stages Questionnaire (ASQ; Bricker & Squires, 1999), a maternal report instrument that assesses early comprehension as well as production abilities. Item response analyses yielded a composite index that differentiated children with LLE from unaffected children and that served as the outcome variable. The obtained percentage of LLE children was 13%, an estimate congruent with those reported in other studies (Fenson et al., 1994; Klee et al., 1998; Rescorla & Achenbach, 2002). One item asked whether a child used 2–3 word combinations. Failure on this item yielded an estimate of 19% as LLE children. This was identical to the estimates from large-scale studies in the United States (Fenson et al., 1994) and the United Kingdom (Roulstone, Loader, Northstone, Beveridge, & the ALSPAC Team, 2002).

Several longitudinal outcome studies have followed late talkers to observe language skills in later childhood. Estimates vary for the number of late talkers whose language performance is below normative expectations. In a study of the first year out, between 24 and 36 months, Hadley and Short (2005) reported that at 36 months, 57% of the LLE children were in the low range (i.e., below the 16th percentile) for mean length of utterance

(MLU) and 79% were in the low range for emergent grammar, as measured by the Index of Productive Syntax (Scarborough, 1990). They also noted that when a more comprehensive criterion of affectedness was used, including other indicators of low performance, 13 of 14 children (93%) showed continued language risk on at least one indicator. Hadley and Short's study (2005) was the first to include grammatical tense marking (also referred to as "finiteness marking"; cf. Rice, Wexler, & Cleave, 1995; Wexler, 1994) in measures of children with a positive history. They reported that 83% of the positive history group showed low levels of performance, defined as below the 16th percentile.

Other studies measured outcomes at 4 and 5 years of age. Paul (1993) reported that 47% of the positive history group showed low levels of performance at age 4; Rescorla, Dahlsgaard, and Roberts (2000) reported 71% at age 4; Whitehurst and Fischel (1994) reported 7% at age 5; and Ellis Weismer (2007) reported 8% at age 5. A large-scale twin study (Dale, Price, Bishop, & Plomin, 2003), using parent-based assessments of vocabulary, grammar, and use of abstract language, reported 44% at age 3 and 40% at age 4. Determination of a single best estimate is complicated by sampling differences across studies as well as some differences in the criteria used for diagnosis and the outcome measures used. A further potential complication is that in some of the studies it is not clear whether the sample was limited to monolingual children not exposed to other languages.

The outcome studies of single-born children measured via direct assessment are limited by small sample sizes with poor statistical power: Hadley and Short (2005) followed 14 children with LLE and 6 controls; Paul (1993), 37 late talkers and 32 controls; Rescorla et al. (2000), 34 late talkers and 16 controls; Whitehurst and Fischel (1994), 37 late talkers; and Ellis Weismer (2007), 53 late talkers and 53 controls.

A further limitation is that information is sparse for outcomes beyond 5 years. Two studies have reported school-age outcomes: Rescorla (2002) followed 34 late talkers and 25 controls to age 9; and Paul, Murray, Clancy, and Andrews (1997) followed 32 late talkers and 27 controls to age 8. Rescorla (2002) reported that late talkers had lower scores on most language measures through age 8 compared with control children matched initially on SES and nonverbal ability and equivalent on gender, although the mean scores were typically within average range. Clinical impression based on conversational speech suggested that 6 of the late talker children (17%) were affected. Rescorla's findings were consistent with those of Paul et al. (1997), who reported on outcomes at 8 years for 27 control children matched initially on SES, gender, birth order, and nonverbal ability and 32 late talkers subdivided into two groups: 5 children

(16%) with concurrent low Developmental Sentence Scoring (DSS; Lee, 1974) scores (considered to have "chronic expressive language delay") and 27 children from the original late talker group who scored above the 10th percentile on the DSS. At group mean levels of performance, the positive history group performed below the controls on an omnibus expressive language scale, although most were in the typical range of performance. Paul and Fountain (1999) reported that for the late talker group, SES and expressive language skills predicted language outcome measured as DSS scores in second grade; gender was not a predictor.

Although the general outcomes suggest that, on average, children with a history of LLE are likely to perform within typical range by school age, there are suggestions that the risk for continued weaknesses in language may be more apparent in some dimensions of language than in others. Rescorla (2002) reported group differences on vocabulary, grammar, phonology, and sentence imitation at age 6; vocabulary only at age 7; and vocabulary and grammar at age 8. The picture is complicated by measurement issues, in that different assessments were used at different age levels. Paul et al. (1997) noted that the children with "chronic expressive language delay" showed morphological and syntactic limitations consistent with their lower DSS scores. Paul et al. (1997) also noted persistent minor articulation errors in the positive history group.

Suggestions of risk for grammatical, morphological, or syntactic limitations at later ages are consistent with recent studies documenting the morphosyntactic limitations of children with SLI, limitations that exceed the other dimensions of language delay. Children with language impairments take longer than unaffected children to master the use of obligatory grammatical tense or finiteness markers, a deficit documented in 8-year-old children in simple (Rice & Wexler, 1996a, 1996b) and complex (Owen & Leonard, 2006) clause constructions, and in 10- to 12-year-old children in written language (Windsor, Scott, & Street, 2000). What is needed are more detailed assessments in this part of the grammar to investigate whether they will yield greater sensitivity to long-term risk for language acquisition in children with a history of LLE.

## ***Growth Perspectives: LLE Outcomes as Clues to the Mechanisms of Language Growth***

From a growth perspective, the phenomenon of LLE represents a delayed onset of language. This is captured in formal growth analyses as the projected intercept defining the initial point of the subsequent growth-curve trajectory. Rice (cf. 2003, 2004, 2007) has reported longitudinal data comparing growth trajectories from 3 to

9 years of age in children with SLI, age control children, and younger children at equivalent levels of MLU at the outset, across multiple dimensions of language (MLU, semantics, grammatical tense marking). A replicated finding is that the affected children differ from age controls in the intercept, that is, the initial stages, of vocabulary, MLU, and grammatical tense marking (cf. Rice et al., 2006). Affected children are older at the initial stages—that is, language onset is projected to appear at later ages. Furthermore, the affected children differ from MLU-equivalent children in the intercept for grammatical tense marking, suggesting that this property of language emerges even later than vocabulary and expansion in clause length (indexed by MLU). At the same time, once the growth trajectory is under way, the affected children do not differ from the control groups—in other words, even with such a late start, the affected children’s linguistic systems follow the same growth trajectories as other children, showing linear change of even acceleration over time as well as points of acceleration change for nonlinear effects. Interestingly, the points of change follow the same amount of elapsed time for affected children as for the control groups.

Rice (2003, 2004, 2007; Rice et al., 2006) argues that it is as if the timing mechanisms are benchmarked to onset. In the case of SLI, a delayed onset is followed by a growth pattern that is highly similar to the expected pattern, although it plays out as if the children were several years younger. Furthermore, even though grammatical tense marking is not synchronized with vocabulary and MLU, because it takes even longer to start up, it nevertheless follows the same growth path once started. This means that more time is required for affected children to reach a given level of performance. (In fact, it is an open question as to whether affected children ever fully reach adult expectations on grammatical tense marking.) Finally, the predictors of growth differ across the linguistic dimensions. Nonverbal IQ predicts growth in MLU and vocabulary for affected and unaffected children (Rice et al., 2006) but does not predict growth in grammatical tense marking (Rice, Tomblin, et al., 2004; Rice et al., 1998; Rice, Wexler, Marquis, & Hershberger, 2000). Mother’s education does not predict growth in any of the three dimensions. MLU predicts growth in grammatical tense marking, but vocabulary does not (Rice et al., 1998, 2000), although vocabulary predicts acquisition of irregular past tense forms (Rice et al., 2000).

A recent growth study by Hadley and Holt (2006) provides growth trajectories for grammatical tense marking from 24 to 36 months for 22 children: 16 with LLE and 6 with low normal language levels. Because the children were considered as one group, there is no direct comparison between growth of affected and unaffected children. Linear and quadratic components were detected;

MLU predicted growth in tense marking, but significant linear growth remained in a model controlling for MLU, suggesting that independent factors contribute to growth; mother’s education did not predict growth, although a positive family history of affectedness was a significant predictor. Overall, the findings are highly consistent with the growth modeling investigations with older children.

The implication here is that a growth perspective sets up an interpretively rich view of the potential import of longitudinal outcome studies of LLE. Information about outcomes can help clarify the extent to which an LLE period leads to a protracted period of language delay and the extent to which the areas of language delay are selective or more general. Furthermore, it also bears on the growth resources available to children who overcome an initial delay. Rice (2003, 2004, 2007) notes that recovery from delayed onset would require accelerated growth rates in order to “catch up” with other children. In effect, the growth rates of compensated LLE children would be more complex than that of other children. Given the full range of available evidence, Rice and colleagues propose a strong role for growth-timing factors in the acquisition of grammatical tense marking, with a delayed onset as part of the selective growth-timing impairments in the linguistic systems of children with SLI. Under this perspective, children who overcome the initial delayed onset and “reset” their growth trajectories are of considerable interest. It will be important to sort out whether they recover equally in all dimensions of language acquisition, whether the causal factors in their initial delay may be different from children with persistent language impairments, and/or whether other mediating influences, either in the home or in the child’s other developing abilities, contribute to a later shift in acceleration of language. To further understand these growth mechanisms and the source of the persistent language impairments of affected children, further study of the language characteristics of children’s early language abilities and later language outcomes are needed to reveal the ways in which language growth unfolds.

To recapitulate, the available evidence on long-term outcomes of LLE is hampered by (a) a relatively small evidence base; (b) measurement limitations, including no reports of grammatical tense marking at school age even though there are multiple indicators that morphology and syntax are likely to be at risk for a protracted time and there is a large literature indicating that this is a robust clinical marker for affected children in this age range; (c) samples restricted to a relatively narrow range of parental education and familial SES levels; (d) group comparisons controlled by gender so that gender outcomes cannot be examined; and (e) limited means to control possible parental or familial effects.

## Purpose of this Study

In this study, we follow up on the children reported in Zubrick et al. (2007). These children comprise a cohort randomly selected from a known total population frame and longitudinally followed. Of interest to us were the following questions: (a) Do children with a positive history of LLE, drawn from a population sample of families, differ from control children in language outcomes? If so, what proportion of children in the LLE group show language impairment at 7 years of age? (b) Do the methods of measuring LLE at 24 months yield differences in long-term outcomes? (c) Do negative outcomes appear evenly across multiple dimensions of language, or are there selective dimensions of greater risk? (d) Are there long-term gender effects?

We describe the outcomes at 7 years of 128 children identified as LLE at 24 months and compare their status with that of 109 children with normal language emergence (NLE) selected at random who are equivalent on maternal and family characteristics. The study yields estimated proportions of children in the LLE and the control groups who are in the range of language impairment at 7 years of age as a function of two different ways of measuring LLE at 24 months, compares group performance on multiple dimensions of speech and language at 7 years, and examines possible gender and nonverbal IQ effects in 7-year outcomes.

## Method

### Participants

Children with and without a history of LLE at 24 months were recruited from the Randomly Ascertained Sample of Children born in Australia's Largest State (RASCALS) longitudinal study of children's health and development from birth to 8 years (S. R. Zubrick, S. R. Silburn, J. J. Kurinczuk, and P. R. Burton, coprincipal investigators). The original RASCALS cohort was drawn at random from statutory notifications of birth in Western Australia (WA) in 1995 and 1996 and recruited into a study of maternal and child health in pregnancy and early infancy (J. J. Kurinczuk, V. P. Dawes, and P. R. Burton, coprincipal investigators). According to current census data, WA is demographically similar to some states in the midwestern United States. For example, the population of the state of Kansas is 2.7 million, the population of WA is 1.8 million and in each state most of the population is in urban areas. The states are predominantly Caucasian (86% for Kansas; 96% for WA) and the majority of the population are native speakers of English, well educated (86% of the population has completed high school in each state), and family oriented (in Kansas, 55% of all families are couple families with

children, and 9% are sole-parent families; in WA, these rates are 49% and 15%, respectively). On a wide variety of behavioral and biological assessments of children and adults, distributional outcomes conform to normative expectations for instruments normed in the United States or the United Kingdom.

Following the 3-month postpartum response, the RASCALS study was converted to a longitudinal study from birth to 4 years (J. J. Kurinczuk, S. R. Zubrick, A. J. Plant, and S. R. Silburn, coprincipal investigators), and for resource reasons just less than a 70% random sample of mothers of singletons was drawn from the initial 4,007 respondents. However, to ensure that "hard-to-reach" groups remained in the RASCALS study in sufficiently informative proportions, all mothers who were either unmarried or not cohabiting, those who had an annual household income of A\$16,000 or less, and those women whose partner was absent from the household were included in the sample. One hundred mothers were added to the sample on this basis. Thus, a total sample of 2,837 mothers and their singleton infants were selected for longitudinal follow-up, of whom 2,224 (78%) agreed, when their infant was 1 year old, to participate. Of the 2,224 women who agreed to participate, 1,880 (85%) returned a completed questionnaire when their child was 2 years old (see Zubrick et al., 2007, for a detailed description of the characteristics of this sample at 2 years). The potential effects of sample attrition were examined by comparing a range of early life characteristics present at 3 months of age between the respondents at 2 years of age and the respondents at 3 months. No significant differences were observed for mother's place of birth, number of children or adults in the household, or father absence. Furthermore, there was no loss of families who received government benefits.

Some small and statistically significant shifts in the characteristics of the sample at 3 months and 2 years were observed. The sample at 2 years had a slightly higher percentage of families who were earning more than A\$25,000 per year (74.5% vs. 70.7%),  $\chi^2(2) = 10.4, p < .01$ ; who were living in married households (79.3% vs. 75.6%),  $\chi^2(4) = 11.7, p < .03$ ; and who had obtained post-high school technical qualifications or a university degree (27.8% vs. 23.3%),  $\chi^2(4) = 23.6, p < .001$ . Although the differences were significant, they were small and yielded a 2-year-old follow-up sample that was reasonably representative of 2-year-old non-Aboriginal WA singleton children. Note that the percentage of mothers with post-secondary degrees is much lower in this sample than is typical in the literature.

Information about the children's language development was collected in the 2-year questionnaire. All 1,766 children, born in 1995 and 1996, were administered the ASQ at 24 months. The children born in 1996

( $n = 902$ ) were also administered the LDS (Rescorla, 1989). These instruments yielded two criteria for identifying LLE, the ASQ criterion and the Expressive Language criterion (ELANG), which are described below in the *Language Measures* section.

Only children with and without a history of LLE who lived in homes where only English was spoken and who did not have exclusionary conditions for SLI (e.g., deafness, intellectual disability, autism spectrum disorder, Down syndrome, cerebral palsy) were eligible to participate in this study. Additional requirements were that the children were current participants in the RASCALS study who lived in WA and were less than 8 years of age at the time of the face-to-face behavioral assessment. At 8 years, 1,240 (66%) of the 1,880 respondents to the 2-year questionnaire were still participating in the RASCALS study. There was no significant change in the percentage of children with LLE in the sample over time (13.4% at 2 years vs. 14.2% at 8 years).

We identified 185 children with a history of LLE at 24 months who met our inclusion criteria. We successfully contacted 169 of the families of these children, and 128 (76%) consented to participate in this study. We identified 765 children with a history of NLE at 24 months who met our inclusionary criteria. A random sample of 170 of these children was created. We successfully contacted 131 of the families of these children, and 109 (83%) agreed to participate in this study.

## Language Measures

The study evaluated two criteria for the initial classification of LLE. The first was ELANG, drawn from the previous literature and based on two expressive language indices, vocabulary size from the LDS and use of word combinations from the ASQ. As described earlier, the ASQ “combining words” item was available for all children in the sample and the LDS was available for half the sample. Children who met the ELANG criterion for LLE were children with a vocabulary of fewer than 70 words on the LDS or no word combinations on the ASQ. A vocabulary score of 70 words or fewer on the LDS placed children below the 15th percentile in this sample of 902 children. The second criterion identified children who were not combining words at 24 months. This was defined as not saying “2 or 3 words together,” an item from the ASQ Communication Scale, which was administered to all children. The ASQ item “Does your child say 2 or 3 words together?” and the LDS item “Does your child combine 2 or more words into phrases?” ask parents to report whether the child produces word combinations. Data were available on both of these items for 896 of the children born in 1996. Frequency distributions were obtained on both the LDS and ASQ items. Ninety percent of children were reported on the LDS to

be combining two or more words into phrases, and 89% of children were reported on the ASQ to be saying two or three words together. Cross-tabulation of these items indicated complete correspondence of these items for 860 of these cases,  $\chi^2(1) = 547.9, p < .001$ .

The second criterion was developed by Zubrick et al. (2007). Children were classified as LLE on the basis of a composite measure developed from the six-item ASQ Communication Scale (Bricker & Squires, 1999) administered at 24 months. The six items include questions about children’s receptive as well as expressive language ability. Mothers were asked to report whether their child could (a) point to pictures on request; (b) use two- or three-word phrases; (c) carry out simple directions on request; (d) name simple objects; (e) point to body parts on request; and (f) use personal pronouns such as “me,” “I,” and “you.” Zubrick et al. (2007) developed a composite measure derived from item-response analyses that shows good psychometric properties for the full sample for validity and reliability estimates and for differentiating affected from unaffected children, using a cutoff of 1 *SD* below the mean to identify children with LLE. We will refer to this measure as the ASQ criterion for the grouping variable of LLE versus NLE.

Outcome language measures were obtained in face-to-face behavioral assessments when the children were 7 years of age. We selected instruments to evaluate speech; general language ability; and semantic, syntactic, morphosyntactic, and nonverbal intellectual development. In addition, the children’s hearing was screened; all had typical hearing levels at 25 dB HL at 1000, 2000, and 4000 Hz.

The speech measures were the Goldman Fristoe Test of Articulation—Second Edition (GFTA-2; Goldman & Fristoe, 2000) and the Phonological Probe from the Rice/Wexler Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001). The Phonological Probe assesses children’s ability to produce final *-t*, *-d*, *-s*, and *-z*, the phonemes necessary to mark third-person singular and past tense *-ed*. The language measures were the Test of Language Development—Primary, Third Edition (TOLD-P:3; Newcomer & Hammill, 1997); the Peabody Picture Vocabulary Test—III (PPVT-III; Dunn, Dunn, & Williams, 1997); and the TEGI (Rice & Wexler, 2001). The TOLD-P:3 served as an omnibus language assessment with important precedents in the literature (cf. Tomblin et al., 1997) with the further advantage that it provides standard score quotients for syntax versus semantics. As with other omnibus tests, the subtests of TOLD do not lend themselves to a descriptive linguistic analysis in that many dimensions of syntax and semantics are intermingled in the item selection.

TEGI outcomes provide more closely grained analyses of a linguistically coherent set of morphemes that

comprise the set of morphemes that are obligatory finiteness markers: third-person singular *-s*, past tense, copula and auxiliary *BE*, and auxiliary *DO*. In addition, we included the grammatical composite score, which is the composite of the four target morphemes, and the screener composite, an average of the third-person singular *-s* and past tense. Three measures from the grammatical judgment tasks on TEGI were included: dropped marker, agreement (AGR), and dropped *-ing*. The dropped marker task evaluates judgments of clauses with or without omitted third-person *-s*, past tense, or *BE* and is thought to be a comprehension analog of the pattern of omissions characteristic of children's productions. AGR evaluates judgments of clauses with or without errors of subject-verb agreement (such as "he am coming back"). Dropped *-ing* evaluates judgments of clauses with or without a dropped *-ing* progressive morpheme (such as "he is cough"). It is expected that the dropped marker will be more sensitive to grammatical tense deficits at 7 years than will the AGR judgments or dropped *-ing*, because, following the predictions of the extended optional infinitive model, children with language impairments are more likely to judge omitted finiteness forms as acceptable than overt errors of agreement or omitted *-ing* (Rice & Wexler, 2001; Rice, Wexler, & Redmond, 1999).

The Systematic Analysis of Language Transcripts program (Miller & Chapman, 1991) was used to generate values for MLU in morphemes (MLU-M) from spontaneous samples of adult-child conversation. MLU was selected because it is thought to be an index of clausal expansions that are general indicators of language growth (cf. Rice et al., 2006) and because of its use in previous studies of late talkers (cf. Hadley & Short, 2005). The spontaneous samples were collected following the procedures described in Rice et al. (2006) using a standard set of age-appropriate toys. These included toy people, a toy house or garage and furniture, and toy animals. The toys were chosen to elicit a wide variety of grammatical forms and sentence types. The samples were audio-recorded using a dual-microphone setup. The aim was for a minimum of 200 child utterances ( $M = 243.10$ ,  $SD = 84.16$ , for the LLE group;  $M = 239.97$ ,  $SD = 76.26$ , for the NLE group). Typically, 200 child utterances were collected in 20–30 min. The samples were transcribed by the examiners and coded for grammatical morphemes following the conventions of the Kansas Language Transcript Database (Rice, Ash, et al., 2004). Utterance segmentation followed Miller (1981, p. 14): terminal intonation contour, pauses of 2–3 s, a limit of two independent clauses conjoined by a coordinating conjunction (e.g., "and," "but") in one utterance, and clauses joined by subordinating conjunctions (e.g., "because," "when") included in single utterances.

Overall, the protocol documents speech outcomes (GFTA-2 and TEGI Phonological Probe) and a range of

language outcomes, from omnibus measures of language (TOLD-P:3 Spoken Language Quotient) to indicators of general semantics (TOLD-P:3 Semantics Quotient) and general syntax (TOLD-P:3 Syntax Quotient), to more specific measures of vocabulary growth (PPVT-III), clause expansion (MLU-M), and morphosyntax (TEGI). In addition, nonverbal intelligence was assessed with the Columbia Mental Maturity Scale (CMMS; Burgemeister, Blum, & Lorge, 1972). The measures were administered over an average of three sessions per child.

## Training and Reliability

The language measures were administered by 11 graduate research assistants with qualifications in speech-language pathology, psychology, or education. The research assistants were trained to administer the language measures, and systematic reliability and validity checks were completed throughout data collection. Training involved tutorials for each measure, directed practice, observed practice assessment sessions, and observed field assessment of study children. Routine checks of examiner response coding during data collection sessions yielded point-to-point agreements of 95%–100% for all language measures.

The research assistants were trained to follow best practice guidelines regarding sample collection. This included following the children's conversational lead, engaging in parallel talk and parallel play focusing on everyday event schemas such as household activities, sharing personal anecdotes and experiences, and introducing topics related to past and ongoing events during their conversational interactions. Research assistants were also trained to keep the use of yes/no and *wh*-questions to a minimum. Transcription and coding followed a written protocol (Rice, Ash, et al., 2004). Research assistants had to meet a criterion of interrater reliability on training transcripts at a level of 90% on three samples prior to independent transcription and coding. Ongoing monitoring of transcription and coding accuracy was provided by Rice's lab, via error checking of each transcript and regular feedback to the transcribers. Pairwise reliability assignments and calculations were carried out in Rice's lab. Independent interrater reliability for transcription and coding were calculated as point-to-point percentage agreement across 65 samplings with an overall mean of 90%, collapsed across the levels of utterances, words, morphemes, and codes.

## Results

Employing the ELANG criterion, we formed two groups of children. The positive history LLE group was comprised of children who met the ELANG criterion at

24 months. This yielded a group of 128 children available at follow-up. The mean age for this group was 7;3 (years; months;  $SD = 0.6$ ). A comparison group of 109 NLE children was randomly drawn, with a mean age of 7;3 ( $SD = 0.2$ ), from the full group of 765 NLE children. A  $t$  test was calculated to confirm that the groups did not differ on age. A series of chi-square analyses found that the groups were equivalent for the following variables: levels of maternal education, maternal mental health, maternal hours in paid employment, family income, SES area resources 1  $SD$  below the population mean, SES area disadvantage 1  $SD$  below the population mean, family type (nuclear, blended, single parent), family functioning, and parenting style (see Zubrick et al., 2007, for further description of the variables). The groups differed in gender composition,  $\chi^2 = 10.6$ ,  $p < .001$ , and family size,  $\chi^2 = 3.69$ ,  $p < .05$ . There were more males (72%) than females in our LLE group than in our NLE group, consistent with other studies (Klee et al., 1998; Rescorla, 1989; Roulstone et al., 2002). There were roughly equivalent numbers of males (51%) and females in the NLE group. This male-to-female sex ratio was the same as the sex ratio in the population-based sample from which the LLE and the NLE groups were drawn originally (Zubrick et al., 2007), allowing for estimation of affectedness

among males in the NLE group. In previous follow-up studies, the control groups have been predominantly male (Paul et al., 1997; Rescorla, 2002). Children with LLE were more likely to have one or more siblings (58%) than were children with NLE, meaning that there were more only (i.e., first-born) children in the NLE group than in the LLE group. We know of one other study that compared family size between groups and reported a similar result (Whitehurst, Fischel, Arnold, & Lonigan, 1992).

The first question of interest is whether the positive history group differs from the control group. We report the group differences in tandem with a second question: whether there are differences in outcomes dependent on two different ways of measuring LLE at 24 months. Table 1 summarizes the outcomes. Variables are reported as standard scores in order to control for age. Standard scores for variables derived from the TEGI (Rice & Wexler, 2001) are calculated from the means and standard deviations reported in the manual for each age level. For the MLU-M standard scores, we drew on the archival sample in Rice's lab of monolingual control children in the Midwest whose samples were collected following the same procedures used in this study (cf. Rice et al., 2006). The sample consisted of 29 children, mean age of

**Table 1.** Means, standard deviations, and effect sizes for ELANG and ASQ grouping criteria.

Measure	ELANG					ASQ				
	LLE (n = 128)		NLE (n = 109)		Effect size (Cohen's <i>d</i> )	LLE (n = 88)		NLE (n = 144)		Effect size (Cohen's <i>d</i> )
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
PPVT-III SS	101.23	11.06**	105.01	11.71	0.32	100.94	10.38*	104.4	12.09	0.29
GFTA SS	101.53	10.28**	104.53	5.65	0.53	101.25	1.11*	104.1	6.59	0.44
CMMS SS	107.81	14.43	109.14	13.47	0.10	107.99	14.72	108.9	13.74	0.07
TOLD SpQ	95.52	13.21*	99.51	11.95	0.33	95.01	13.57*	98.97	12.22	0.32
TOLD SynQ	94.30	12.45***	101.42	14.28	0.50	93.02	14.67***	100.46	12.59	0.59
TOLD SemQ	97.73	14.93	96.86	12.18	-0.07	97.50	15.44	97.45	12.74	0.00
MLU-M_Z	-0.76	1.37	-0.79	1.09	0.03	-0.77	1.45	-0.78	1.11	0.07
3S_Z	-0.27	2.70**	0.47	0.38	1.93	-0.38	2.99*	0.34	1.07	0.67
Past_Z	-0.17	2.85*	0.56	1.46	0.50	-0.35	3.15*	0.48	1.65	0.51
Be_Z	-0.48	2.59**	0.13	0.97	0.63	-0.69	2.56**	0.09	1.60	0.49
Do_Z	-0.11	1.60*	0.24	0.71	0.49	-0.18	1.72	0.19	0.91	0.40
GrmComp_Z	-0.17	2.64**	0.48	0.69	0.94	-0.29	2.78*	0.38	1.35	0.50
Screen_Z	-0.04	2.21**	0.51	0.68	0.81	-0.10	2.35*	0.41	1.16	0.44
DrpMrk_Z	-0.57	1.65**	-0.01	1.18	0.47	-0.73	1.89**	-0.05	1.10	0.62
AGR_Z	-0.84	2.36	-0.50	1.87	0.18	-1.07	2.58*	-0.44	1.85	0.34
DrpING_Z	-0.25	1.59	0.03	1.29	0.21	-0.36	1.77	0.02	1.25	0.30

Note. ELANG = Expressive Language Criterion; ASQ = Ages and Stages Questionnaire; LLE = late language emergence; NLE = normal language emergence; PPVT-III = Peabody Picture Vocabulary Test—III; SS = standard score; GFTA = Goldman Frisbie Test of Articulation; CMMS = Columbia Mental Maturity Scale; TOLD = Test of Language Development; SpQ = Spoken Language Quotient; SynQ = Syntax Quotient; SemQ = Semantics Quotient; MLU-M = mean length utterance in morphemes; Z = Z score; 3S = third-person singular; Past = past tense; Be = copula and auxiliary production; Do = auxiliary production; GrmComp = composite of -s, past tense, Be and Do; DrpMrk = drop marker; Screen = composite of -s and past tense; AGR = agreement; DrpING = dropped -ing.

\* $p \leq .05$ . \*\* $p \leq .01$ . \*\*\* $p \leq .001$ .



6;11 ( $SD = 0.4$ ), who scored within the typical range on standardized language assessments; had typical hearing; had no history of social, behavioral, or neurological disorders; and whose nonverbal IQ was above 85. The mean MLU-M was 5.04 ( $SD = 0.55$ ). The means for MLU-M for the LLE group for ELANG were 5.19 ( $SD = 1.05$ ) and for ASQ were 5.15 ( $SD = 1.08$ ); for the NLE group the mean for ELANG was 5.21 ( $SD = 0.99$ ), and for ASQ the mean was 5.22 ( $SD = 0.99$ ).

In Table 1, standard score measures are reported within group, as group means and standard deviations. The asterisks indicate obtained  $p$  levels from a series of univariate  $t$  calculations comparing group means per measure, within ELANG or ASQ criteria. We set alpha at .05 and do not adjust for potential experiment-wise error rates because in this initial descriptive study, we prefer error on the side of sensitivity to potential group differences. Also reported are effect sizes calculated as Cohen's  $d$  (Cohen, 1988, p. 41). Following Rosenthal (1994, p. 232),  $d$  was calculated as the mean of NLE minus the mean of LLE divided by the standard deviation of the control group (NLE). Finally, note that the number of participants identified as LLE varies across the criteria, such that the ELANG criterion yields more children with LLE (128 for LLE versus 88 for ASQ; there was a total sample of 237 children for whom ELANG scores were available and a total sample of 232 children for whom ASQ scores were available).

An initial observation is that, as expected, the NLE children's mean levels of performance were congruent with the expected levels of performance on the normative assessments. At the same time, even though the number of participants per group varies across the grouping criteria, the general patterns of group differences hold.

The next question addresses possible group differences on multiple dimensions of speech and language. As shown in Table 1, the following measures yielded significantly lower performance for the LLE group: (a) PPVT-III: ELANG,  $t(1, 235) = 2.55, p = .01$ ; ASQ,  $t(1, 230) = 2.23, p = .027$ ; (b) GFTA-2: ELANG,  $t(1, 235) = 2.72, p = .007$ ; ASQ,  $t(1, 230) = 2.59, p = .01$ ; (c) TOLD-P:3 Spoken Quotient: ELANG,  $t(1, 235) = 2.42, p = .016$ ; ASQ,  $t(1, 230) = 2.29, p = .023$ ; and (d) TOLD-P:3 Syntactic Quotient: ELANG,  $t(1, 235) = 4.10, p = .000$ ; ASQ,  $t(1, 230) = 4.10, p = .000$ . From TEGI the following measures were significant: (a) third-person singular  $-s$ : ELANG,  $t(1, 234) = 2.81, p = .005$ ; ASQ,  $t(1, 229) = 2.63, p = .009$ ; (b) past tense: ELANG,  $t(1, 234) = 2.41, p = .017$ ; ASQ,  $t(1, 229) = 2.64, p = .009$ ; (c) *BE*: ELANG,  $t(1, 234) = 2.34, p = .02$ ; ASQ,  $t(1, 229) = 2.84, p = .005$ ; (d) *DO*: ELANG,  $t(1, 232) = 2.11, p = .036$ ; ASQ,  $t(1, 227) = 2.12, p = .035$ ; (e) grammatical composite score: ELANG,  $t(1, 234) = 2.51, p = .013$ ; ASQ,  $t(1, 229) = 2.48, p = .014$ ; (f) screener composite: ELANG,  $t(1, 234) = 2.51, p = .013$ ; ASQ,  $t(1, 229) = 2.23,$

$p = .027$ ; (g) dropped marker: ELANG,  $t(1, 235) = 2.95, p = .004$ ; ASQ,  $t(1, 230) = 3.47, p = .001$ ; and (h) AGR for ASQ criterion only: ASQ,  $t(1, 230) = 2.15, p = .03$ . As expected, dropped *-ing* did not differentiate the groups. MLU-M also did not differentiate the groups.

Comparison across measures is facilitated by inspection of the effect sizes. According to Cohen (1988), effect sizes of .20 can be considered small; .50, medium; and .80, large. Across the grouping criteria, measures of syntax and morphosyntax consistently yielded large effects, whereas vocabulary measures yielded small (PPVT-III) or no effects (TOLD-P:3 Semantics Quotient). Note that the TOLD-P:3 Spoken Quotient has a medium effect size, but this seems to be entirely attributable to the Syntax Quotient. Speech development differentiates the group means, with small-to-medium effect size (depending on the criterion), but the performance levels are solidly in the typical range for the LLE group.

The percentages of children in the LLE and the NLE groups who were in the range of language impairment at 7 years of age are reported in Table 2 for ELANG. The results are essentially the same for ASQ, so only ELANG is described here. The cutoff for affectedness per variable is defined as 1  $SD$  or more below the mean. This cutoff follows the criterion used for the definition of LLE at 24 months in this study (cf. Zubrick et al., 2007, who report distributional criteria used in the decision to adopt

**Table 2.** Percentage of language impairments at 7 years for LLE and NLE (ELANG criterion,  $-1 SD$ ).

Measure	LLE		NLE	
	n/N	%	n/N	%
PPVT-3 SS	6/128	5	4/109	4
GFTA SS	9/128	7*	2/109	2
CMMS SS	8/127	6	4/109	4
TOLD SpoQ	26/128	20*	12/109	11
TOLD SynQ	23/128	18*	9/109	8
TOLD SemQ	24/128	19	17/109	16
MLUM_Z	28/125	22	25/109	23
3 S_Z	13/127	10**	1/109	1
Past_Z	16/127	13**	4/109	4
Be_Z	24/127	19	14/109	13
Do_Z	19/126	15	11/108	9
GrmComp_Z	14/127	11	6/109	10
Screen_Z	11/127	9*	2/109	2
DrpMrk_Z	29/128	23*	15/109	14
AGR_Z	25/128	20	14/109	13
DrpING_Z	15/128	12	9/109	8
Phono probe	5/122	4	1/108	1

Note. Chi-square is one-sided, Fisher's exact test where appropriate.

\* $p \leq .05$ . \*\* $p \leq .01$ .

this level) and in other studies (cf. Hadley & Holt, 2006; Hadley & Short, 2005). One *SD* below the mean is the criterion in use in the studies of Rice and colleagues for differentiating between affected and unaffected preschool- and elementary school-aged children. These studies demonstrate replicated group differences for affected and unaffected children on finiteness marking, a recognized phenotype of SLI (cf. Tager-Flusberg & Cooper, 1999). Although this is a somewhat more liberal definition of affectedness than the oft-cited criterion of  $-1.25$  of Tomblin et al. (1997), it is quite close to the actual level of  $-1.14$  reported by Tomblin et al. (1997) for their final definition collapsed across subtests of TOLD-P:2.

Chi-square calculations were carried out to determine the statistical significance of observed proportions of affected children across the two groups, per variable, estimated by the Fisher's exact test when appropriate for low expected frequencies and one-sided predictions given the expected greater proportion of children with positive history of LLE at 24 months in the low end of the distribution. As indicated by the asterisks in Table 2, statistical significance was apparent for the following variables: GFTA-2,  $\chi^2 = 3.93$ ,  $p = .05$ ; TOLD-P:3 Spoken,  $\chi^2 = 3.88$ ,  $p = .037$ ; and TOLD-P:3 Syntax,  $\chi^2 = 4.93$ ,  $p = .022$ ; and for the following TEGI variables: third-person present *-s*,  $\chi^2 = 10.99$ ,  $p = .002$ ; past tense,  $\chi^2 = 6.50$ ,  $p = .011$ ; screener,  $\chi^2 = 5.85$ ,  $p = .019$ ; and drop marker,  $\chi^2 = 3.138$ ,  $p = .055$ .

The outcomes in Table 2 show that the vocabulary indices drop out of group differentiation in the low end of the distribution. With regard to syntax, the TOLD-P:3 Syntax Quotient placed 18% of the positive history children in the low end of the distribution. On the morphosyntactic level, as expected the judgment of omitted finiteness marking (dropped marker) was sensitive to persistent weaknesses in morphosyntax, picking up 23% of the positive history group, at the highest levels of sensitivity for the variables yielding significantly higher numbers of affected children in the positive history group. Conversely, of the significant variables, third-person singular present tense *-s* had the highest percentage of the NLE group (99%) performing above the cutoff level, compared with 90% of the LLE group. These values are in line with the criterion scores provided from the normative sample of TEGI (Rice & Wexler, 2001), which show ceiling effects by 7 years for control children.

Another perspective on outcome risk for LLE is the percentage of all children who scored below the cutoff at 7 years of age who were LLE. For the variables that differentiated the groups, the percentages were as follows: GFTA, 82%; TOLD-P:3 Spoken Quotient, 68%; TOLD-P:3 Syntactic Quotient, 72%; third-person present *-s*, 93%; past tense, 80%; screener, 85%; and drop marker, 66%. Thus, on these variables, a relatively high

percentage of children who scored below the cutoff had a positive history of LLE.

Although the GFTA-2 speech outcomes differentiated the groups, and the percentage of LLE children in the low range at outcome was high, at the same time the overall percentage of children in the low range was small—that is, 7% in the LLE group versus 2% in the comparison group (cf. Table 2). This could be because low performance on the GFTA-2 at 7 years is likely to be due to minor misarticulations of late-developing sounds. If so, the likelihood of clinically significant speech problems in the LLE group is small. A related observation is that only 4% of the children with a positive history failed the phonological probe of the TEGI, an index of final consonant use necessary for morphological use. This is further indication of generally robust speech development at 7 years. Preliminary analyses found that TEGI outcomes were not different for the children who had failed the phonological probe, so they were included in the group means reported here.

The issue of gender risk for language impairments was motivated by the strong predictive role of gender for LLE at 24 months. With regard to the gender composition of the groups defined by ELANG versus ASQ criteria at 7 years, as expected we found that a higher percentage of males were in the positive history group: 92 males (72%) for ELANG; 65 males (74%) for ASQ. In order to evaluate whether a greater proportion of males was apparent at the low end of the distribution at 7 years, we carried out chi-square calculations for each of the criteria. There was only one instance of a statistically significant gender effect, for the *DO* variable from TEGI. Interestingly, the effect was in the opposite direction; more girls than boys were in the low end of the distribution. Overall, there was no reason to believe that the strong gender effects evident in the 24-month sample persist at 7 years of age (cf. Tomblin et al., 1997).

Following the interest in previous studies with nonverbal intelligence as a predictor of language outcomes, we investigated whether the children who showed low language performance at 7 years were likely to be drawn from children with low nonverbal intelligence performance levels. To explore this possibility, we carried out chi-square analyses with the speech and language variables that differentiated the two groups—that is, those indicated by an asterisk in Table 2. Children were categorized according to performance above standard scores of 85 versus performance below these scores on the CMMS, cross-tabulated with above-standard scores of 85 versus below-standard scores on each of the speech and language outcomes. Inspection of the cross-tabulation tables showed very low numbers of children scoring below 85 on the CMMS (8 for the LLE group; 3 for the NLE group). For the LLE group, the split of children above

and below the language cutoff for the significant variables was 3 and 5. This poses a strong caveat for the significant chi-square outcomes, given that they are attributable to the placement of 1 or 2 children. The speech measure, GFTA-2, yielded a nonsignificant outcome. Of the language measures, the following variables yielded significant chi-square values: TOLD-P:3 Spoken Quotient: LLE,  $\chi^2 = 9.26, p < .1$ ; NLE,  $\chi^2 = 24.94, p = .001$ ; TOLD-P:3 Syntax Quotient: LLE,  $\chi^2 = 11.34, p < .01$ ; NLE,  $\chi^2 = 13.89, p < .05$ ; dropped marker: NLE,  $\chi^2 = 7.28, p < .05$ . Note that the following TEGI variables did not yield significant differences for either group: third-person singular -s, past tense, and screener score. All things considered, it is not the case that children with low language levels at 7 years are likely to have low levels of nonverbal intelligence. Instead, the large majority of them are within typical range.

## Discussion

This study compared language outcomes of 7-year-old children with a positive history of LLE at 24 months with a comparison group of children without a positive history, controlled on multiple indicators of maternal characteristics and family resources. The short answers to our initial questions are as follows: Yes, there are group differences for long-term language outcomes; the methods of measurement at 24 months vary in sensitivity to initial language delays (i.e., in the proportion of children identified as LLE) but yield the same proportion of affectedness and patterns of group differences at later outcomes; risk for persistent language impairment differs across linguistic dimensions, with syntax and, more specifically, grammatical tense marking as yielding a higher proportion of LLE children at low levels of performance at 7 years; and the pronounced gender effects for LLE at 24 months are no longer evident at 7 years, when males are no more likely than females to score at the low end of the distribution across multiple measures. Additionally, nonverbal intelligence levels are not implicated in long-term risk within this sample.

It is very interesting that there are group differences in language outcomes between the positive history group and the no-history group on multiple dimensions of language, given the equivalence of the groups on an unprecedented range of maternal and family variables, in an epidemiologically diverse sample randomly drawn from a sample frame generalizable to a known population. This is a strong indication that long-term risk for later language outcomes for children with LLE is robust even when group equivalence is present for any possible mediating effects contributed by maternal and family advantages. These findings are consistent with the lack of positive prediction by these variables in the concurrent

assessment of predictors of LLE at 24 months (Zubrick et al., 2007) and with the lack of prediction of maternal education in short-term outcomes reported by Hadley and Holt (2006) and in the modeling studies of older children by Rice and colleagues summarized above. Conversely, Paul and Fountain (1999) reported SES effects on DSS outcomes in second grade. At the same time, it must be noted that groupwise equivalency on home and environmental variables does not rule out the possibility that one or more of these variables could predict some of the variance in individual outcomes, either singly or as mediating variables, within the full range of performance or in the clinical range. Future studies need to carry out multivariate analyses with maternal and family variables as predictors to complement the group design of this study.

Measurement issues are clearly of great importance. At 24 months, ELANG identified more children with LLE than did the ASQ method. Without an independent gold standard of language impairment at this early stage of development, it is difficult to know whether the larger estimate of ELANG is better than the smaller estimate obtained with the six-item ASQ. If we use outcome estimates as a criterion for gold standard at 24 months, the two measures are equivalent in detecting group differences in the full samples and in the proportion of children identified as low performing at 7 years, equivalencies that appear across a wide range of outcome measures. Thus, the operative difference bears on the sensitivity to LLE at 24 months. A case could be made that ELANG is preferred for clinical assessment, in that more children will be identified for consideration for early language intervention.

The findings are consistent with previous studies cautioning that generalizations about linguistic outcomes differ when considering risk for the full range of children in the groups versus risk at the lower levels of performance. As a group, relative to children without a history of LLE, the positive history children were more likely to score lower on a wide range of speech and language indicators, for either of the measures used to define LLE: Speech, vocabulary, syntax, and grammatical tense marking are each likely to be at lower levels of outcome than for the control group. The effect sizes are highest for the syntax and grammatical tense marking measures as compared with semantics. This differentiation is more apparent at the low end of performance, where the significant linguistic differences between groups appear more narrowly on syntax and morphosyntax, adding support to suggestions from earlier studies that syntax is likely to be affected over the long-term.

The protocol for this study clarified further the ways in which different dimensions of speech and language influence long-term outcomes. Speech development was relatively robust, yielding estimates of 2% in the NLE

group and 7% in the LLE group. This suggests that in the general population of children with speech disorders, children with a positive history of LLE are three times more likely to have immature speech relative to age norms than are children without such a history. Caveats apply as to whether these speech differences warrant designation as “speech impairment” of clinical import. Depending on the precise profile of clinical symptoms, children who score at the low end of performance on an articulation assessment may or may not be regarded as in need of speech therapy. These outcomes are consistent with earlier studies of late talkers that found immature speech to be the first linguistic dimension to normalize (Paul, 1996; Whitehurst & Fischel, 1994).

Within the language domain, the protocol further clarifies that syntactic and morphosyntactic development are more vulnerable than semantic development to long-term language impairment. Although the positive history group did not perform as well on average as the comparison group on the PPVT-III estimate of vocabulary development, performance levels in the clinical range do not differentiate the groups on either this index or the semantics subtests of the TOLD-P:3.

In contrast, multiple indicators of syntax show a marked disadvantage for the positive history group. A general estimate of affectedness in this domain is available from the TOLD-P:3 Syntax Quotient, which yields 18% of the LLE group as affected at 7 years. At the morphosyntactic level of finiteness marking, obligatory use of third-person singular present tense *-s* and past tense morphemes (and their composite, the screener score) and grammatical judgments of omissions of these morphemes (dropped marker) show clinical levels of performance for 10%–13% of the LLE group for the production tasks and 23% for the judgment task. This is especially noteworthy because 7 years is at the upper range of sensitivity for the third-person and past tense morphosyntax, because unaffected children are likely to be at ceiling levels of performance. The judgment task for omitted finiteness marking shows higher levels of sensitivity (cf. TEGI norms) because affected children still lag behind controls on that task and because this task taps into the later manifestations of weakness in this grammatical property. At the same time, the difference between the two groups for the dropped marker is about 10%, which is about the same difference as for the third-person and past tense morphology.

The morphosyntactic outcomes are consistent with earlier reports of children with SLI at 7 years, showing that children identified as SLI at 5 years of age (and followed longitudinally for years) are at lower levels of performance on the TEGI third-person singular *-s* and past tense probes (cf. Rice et al., 1998) and on the judgment task (Rice et al., 1999). The import of the findings here is that they provide a partial map of the potential link

between risk for LLE and a risk for subsequent acquisition of morphosyntax.

Overall, this study further clarifies the links between LLE and later language outcomes. The growth perspective positions the observations of LLE as intricately linked with the growing base of growth curve data from older children with and without language impairments. This study contributes by comparing two different ways to measure LLE at the outset and by providing multiple measures of language at 7 years, ranging from general to specific linguistic dimensions, at a time when key properties of morphosyntax involving grammatical tense marking should be at adult levels. The emerging picture implies complex factors at work in the growth trajectories of children with LLE. This study corroborates the emerging conclusions from earlier studies: Most children with LLE are able to compensate for this initial delay in language onset. Most strikingly, young boys are apparently able to overcome their greater likelihood of early delay to close the gap with girls by age 7. This means that boys must accelerate their acquisition rates faster than girls in order to catch up. The mechanisms or concomitant conditions contributing to this are not known but certainly warrant further attention. The next phases of research with this sample will pursue multivariate analyses to investigate predictors of individual outcomes and the extent to which variance is affected by family and child characteristics that may mediate long-term outcomes, for semantic, syntactic, and morphosyntactic outcomes. Identification of mediating variables would help with the clinical issue of how to identify, out of the LLE children, those most at risk for long-term language weaknesses in need of intervention.

## Acknowledgments

Approval to conduct the follow-up study was obtained from the Curtin University of Technology Human Research Ethics Committee; King Edward Memorial Hospital and Princess Margaret Hospital Ethics Committee; and the Confidentiality of Health Information Committee of Western Australia. This study was supported by National Institutes of Health Grant R01 DC005226 and grants from the Health Promotion Foundation of Western Australia. Further support was provided by grants from the National Institutes of Health to the University of Kansas through the Center for Biobehavioral Neurosciences in Communication Disorders (Grant P30DC005803) and through the Mental Retardation and Developmental Disabilities Research Center (Grants P30HD002528 and R01 DC01803). We thank Janet Marquis for statistical consultation; Denise Perpich for assistance with data management; and Sarah Beveridge-Pearce, Bradley Calamel, Piers Dawes, Melanie Epstein, Erika Hagemann, Antonietta Italiano, Stephanie McBeath, Megan McClurg, Kirsty Mackenzie, Lucy Masterson, Virginia Muniandy, Eva Muir, Elke Scheepers, and Michaela Stone for data collection. We

especially thank the children and families who participated in this study.

## References

- Bricker, D., & Squires, J.** (1999). *Ages and stages questionnaires: A parent completed, child-monitoring system* (2nd ed.). Baltimore: Brookes.
- Burgemeister, B. B., Blum, L. H., & Lorge, I.** (1972). *Columbia Mental Maturity Scale*. San Antonio, TX: Psychological Corporation.
- Cohen, J.** (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- 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.
- Dale, P. S., Simonoff, E., Bishop, D. V. M., Eley, T. C., Oliver, B., Price, T. S., et al.** (1998). Genetic influence on language delay in two-year-old children. *Nature Neuroscience, 1*, 324–328.
- Dunn, L. M., Dunn, L. M., & Williams, K. T.** (1997). *Peabody Picture Vocabulary Test—III*. Circle Pines, MN: American Guidance Service.
- Ellis Weismer, S.** (2007). Typical talkers, late talkers, and children with specific language impairment: A language endowment spectrum? In R. Paul (Ed.), *The influence of developmental perspectives on research and practice in communication disorders: A festschrift for Robin S. Chapman* (pp. 83–102). Mahwah, NJ: Erlbaum.
- Feldman, H. M., Campbell, T. F., Kurs-Lasky, M., & Rockette, H. E.** (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*, 856–868.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J.** (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*, 1–173.
- Fenson, L., Dale, P. S., Reznick, S., Thal, D., Bates, E., Hartung, J., et al.** (1993). *MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular.
- Goldman, R., & Fristoe, M.** (2000). *Goldman-Fristoe Test of Articulation* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Hadley, P. A., & Holt, J. K.** (2006). Individual differences in the onset of tense marking: A growth-curve analysis. *Journal of Speech, Language, and Hearing Research, 49*, 984–1000.
- Hadley, P. A., & Short, H.** (2005). The onset of tense marking in children at-risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*, 1344–1362.
- Klee, T., Carson, D., Gavin, W., Hall, L., Kent, A., & Reece, S.** (1998). Concurrent and predictive validity of an early language screening program. *Journal of Speech, Language, and Hearing Research, 41*, 627–641.
- Lee, L.** (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.
- Miller, J.** (1981). *Assessing language production in children*. Baltimore: University Park Press.
- Miller, J., & Chapman, R. S.** (1991). *Systematic analysis of language transcripts*. Madison: University of Wisconsin Language Analysis Lab.
- Newcomer, P. L., & Hammill, D. D.** (1997). *Test of Language Development—Primary* (3rd ed.). Austin, TX: Pro-Ed.
- Owen, A. J., & Leonard, L. B.** (2006). The production of finite and nonfinite complement clauses by children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research, 49*, 548–571.
- Paul, R.** (1993). Patterns of development in late talkers: Preschool years. *Journal of Childhood Communication Disorders, 15*, 7–14.
- Paul, R.** (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5*, 5–21.
- Paul, R., & Fountain, R.** (1999). Predicting outcomes of early expressive language delay. *Infant-Toddler Intervention: The Transdisciplinary Journal, 9*, 123–136.
- Paul, R., Murray, C., Clancy, K., & Andrews, D.** (1997). Reading and metaphonological outcomes in late talkers. *Journal of Speech, Language, and Hearing Research, 40*, 1037–1047.
- Rescorla, L.** (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*, 587–599.
- Rescorla, L., Hadick-Wiley, M., & Escarce, E.** (1993). Epidemiological investigation of expressive language delay at age two. *First Language, 13*, 5–22.
- Rescorla, L.** (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45*, 360–371.
- Rescorla, L., & Achenbach, T. M.** (2002). Use of the Language Development Survey (LDS) in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language, and Hearing Research, 45*, 733–743.
- Rescorla, L., Dahlsgaard, K., & Roberts, J.** (2000). Late talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27*, 643–664.
- Rescorla, L., Ratner, N. B., & Jusczyk, P.** (2005). Concurrent validity of the Language Development Survey: Associations with the MacArthur-Bates Communicative Development Inventories: Words and sentences. *American Journal of Speech-Language Pathology, 14*, 156–163.
- Rice, M. L.** (2003). A unified model of specific and general language delay: Grammatical tense as a clinical marker of unexpected variation. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment* (pp. 63–95). Mahwah, NJ: Erlbaum.
- Rice, M. L.** (2004). Growth models of developmental language disorders. In M. L. Rice & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207–240). Mahwah, NJ: Erlbaum.
- Rice, M. L.** (2007). Children with specific language impairment: Bridging the genetic and developmental perspectives. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 411–431). Malden, MA: Blackwell.

- Rice, M. L., Ash, A., Betz, S., Francois, J., Kepler, A., Klager, E., & Smolik, F.** (2004). *Transcription and coding manual for analysis of language samples*. Lawrence: University of Kansas, Language Acquisition Studies Lab.
- Rice, M. L., Redmond, S. M., & Hoffman, L.** (2006). Mean length of utterance in children with SLI and younger control children shows concurrent validity, stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research, 49*, 793–808.
- Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J.** (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research, 47*, 816–834.
- Rice, M. L., & Wexler, K.** (1996a). A phenotype of specific language impairment: Extended optional infinitives. In M. L. Rice (Ed.), *Toward a genetics of language* (pp. 215–237). Mahwah, NJ: Erlbaum.
- Rice, M. L., & Wexler, K.** (1996b). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39*, 1239–1257.
- Rice, M. L., & Wexler, K.** (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: Psychological Corporation.
- Rice, M. L., Wexler, K., & Cleave, P. L.** (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research, 38*, 850–863.
- 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., Marquis, J., & Hershberger, S.** (2000). Acquisition of irregular past tense by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 43*, 1126–1145.
- Rice, M. L., Wexler, K., & Redmond, S. M.** (1999). Grammaticality judgements 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.
- Rosenthal, R.** (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231–244). New York: Sage.
- Roulstone, S., Loader, S., Northstone, K., Beveridge, M., the ALSPAC Team.** (2002). The speech and language of children aged 25 months: Descriptive data from the Avon Longitudinal Study of Parents and Children. *Early Childhood Development and Care, 172*, 259–268.
- Scarborough, H. S.** (1990). Index of productive syntax. *Applied Psycholinguistics, 11*, 1–22.
- Tager-Flusberg, H., & Cooper, J.** (1999). Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1275–1278.
- Tomblin, J. B., Records, N. L., Buckwater, P., Zhang, X., Smith, E., & O'Brien, M.** (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*, 1245–1260.
- Wexler, K.** (1994). Optional infinitives, head movement, and the economy of derivations. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (pp. 305–350). Cambridge, United Kingdom: Cambridge University Press.
- Whitehurst, G. J., & Fischel, J.** (1994). Early developmental language delay: What, if anything, should the clinician do about it? *Journal of Child Psychology and Psychiatry, 35*, 613–648.
- Whitehurst, G. J., Fischel, J., Arnold, D. S., & Lonigan, C. J.** (1992). Evaluating outcomes with children with expressive language delay. In S. F. Warren & J. Reichle (Eds.), *Causes and effects in communication and language intervention* (Vol. 1, pp. 227–313). Baltimore: Brookes.
- Windsor, J., Scott, C. M., & Street, C. K.** (2000). Verb and noun morphology in the spoken and written language of children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*, 1322–1336.
- Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W.** (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.

---

Received January 26, 2007

Accepted July 17, 2007

DOI: 10.1044/1092-4388(2008/029)

Contact author: Mabel L. Rice, 3031 Dole Human Development Center, University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045. E-mail: mabel@ku.edu.