Research Article

Heritability of Specific Language Impairment and Nonspecific Language Impairment at Ages 4 and 6 Years Across Phenotypes of Speech, Language, and Nonverbal Cognition

Mabel L. Rice,[a] Catherine L. Taylor,[b] Stephen R. Zubrick,[b] Lesa Hoffman,[c] and Kathleen K. Earnest[d]

Purpose: Early language and speech acquisition can be delayed in twin children, a twinning effect that diminishes between 4 and 6 years of age in a population-based sample. The purposes of this study were to examine how twinning effects influence the identification of children with language impairments at 4 and 6 years of age, comparing children with specific language impairment (SLI) and nonspecific language impairment (NLI); the likelihood that affectedness will be shared within monozygotic versus dizygotic twin pairs; and estimated levels of heritability for SLI and NLI. Twinning effects are predicted to result in elevated rates of language impairments in twins.
Method: The population-based twin sample included 1,354 children from 677 twin pairs, 214 monozygotic and 463 dizygotic, enrolled in a longitudinal study. Nine phenotypes from the same comprehensive direct behavioral assessment protocol were investigated at 4 and 6 years of age. Twinning effects were estimated for each phenotype at each age using structural equation models estimated via diagonally weighted least squares. Heritabilities were calculated for SLI and NLI.
Results: As predicted, the twinning effect increased the percentage of affected children in both groups across multiple language phenotypes, an effect that diminished with age yet was still not aligned to singleton age peers. Substantial heritability estimates replicated across language phenotypes and increased with age, even with the most lenient definition of affectedness, at −1 SD. Patterns of outcomes differed between SLI and NLI groups.
Conclusions: Nonverbal IQ is not on the same causal pathway as language impairments. Twinning effects on language acquisition affect classification of 4- and 6-year-old children as SLI and NLI, and heritability is most consistent in the SLI group. Clinical practice requires monitoring language acquisition of twins to avoid misdiagnosis when young or a missed diagnosis of language impairments at school entry.

W orldwide, the proportion of babies born as twins is increasing. In the United States in 2009, one in every 30 babies (3.3%) born was a twin, compared to one in every 53 babies (1.8%) in 1980 (Martin et al., 2012). The increase is due to two factors: the tendency for women to delay having children until they are older and the increased use of fertility treatments, resulting in an increase in fraternal (dizygotic [DZ]) twins. Twins are at an elevated risk for prenatal, perinatal, and neonatal mortality and morbidity, and their developmental outcomes are of major interest to researchers, clinicians, educators, and parents (Blickstein & Keith, 2005). Early speech and language outcomes and possible impairments are the focus of this study.

The potential scientific informativeness of twins has long been recognized. For over a century, the most widely used methodology for evaluating the role of genetics and environment in language and cognition, including their development and impairments, has been the twin study, which uses—rather than studies—twins. There is a much smaller literature focused on twins, which has typically

[a]Child Language Doctoral Program, University of Kansas, Lawrence
[b]Telethon Kids Institute, The University of Western Australia, Nedlands, Australia
[c]College of Education, University of Iowa, Iowa City
[d]Language Acquisition Studies Lab, University of Kansas, Lawrence

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

documented delays, especially in language, known as a "twinning effect" on language acquisition (Hay et al., 1987; Taylor et al., 2018; Thorpe, 2006; Thorpe et al., 2003). The delays have seemed to diminish by school years, although empirical evidence is limited. There has been very little research connecting these foci, looking at the effect of using twins on the identification of language impairment and the heritability estimates from twin studies. This study addresses that gap.

Language Impairments as Phenotypes

Central to the study of heritability in twins and in molecular genetics studies is the notion of "phenotype" or the observable/measurable characteristic hypothesized to be the results of genetic and/or environmental influences. The more precise and replicable the measurement methods for phenotyping, the more accurate are the estimates of genetic and environmental influences. Toward this goal, the National Institutes of Health has created the PhenX Toolkit (consensus measures for Phenotypes and eXposures) to identify and promote the use of standard measurement protocols that allow for cross-study analyses and increased statistical significance. Measurements in the speech, language, and hearing domain were developed by a working group of experts in October 2010 and recently updated in 2019 (https://www.phenxtoolkit.org). See https://www.phenxtoolkit.org/domains/view/200000 for the roster of the expert panel that included the first author of this article. Replicated evidence of heritability of individual measures is the standard for likely reproducibility of outcomes across studies. Multiple phenotypes in the same sample of participants enhance comparisons across phenotypes. This study uses phenotypes listed in PhenX and evaluates replicability of outcomes at two age levels in early childhood across different criteria of affectedness.

The identification of speech and language impairments in twins involves consideration of multiple phenotypes across the dimensions of speech, language, and nonverbal cognitive abilities. There is widespread consensus that earlier attempts to identify "profiles" of relative strengths in speech and various dimensions of language proved to be unreliable or did not conform to assumptions (Bishop et al., 2017). For example, it is widely assumed that children with language impairments are very likely to have speech impairments, further implying a common causal pathway. This assumption is consistent with the distribution of children in clinical caseloads, presumably because speech impairment attracts attention due to limited intelligibility (Tomblin et al., 1997). There is only one population-based study, that is, with participants representative of the full population, with proper phenotyping methods for 6-year-old children (Shriberg et al., 1999). The outcomes were that the two forms of impairment were essentially orthogonal, with a very small likelihood of overlap in speech and language impairments.

Another common assumption is that language impairments overlap with nonverbal cognitive impairments.

That assumption has also been disproved by the outcomes of population-based samples of 6-year-old children. The classic design for examining this assumption is to cross-classify children with or without language impairments and with or without nonverbal cognitive impairments. This cross-classification requires definitions of language impairment. One classification is that of specific language impairment (SLI), defined on the National Institute on Deafness and Other Communication Disorders website as "a language disorder that delays the mastery of language skills in children who have no hearing loss, intellectual impairment, or other developmental delays" (https://www.nidcd.nih.gov/health/specific-language-impairment). For decades, research studies of SLI excluded children with nonverbal cognitive abilities below their typical age peers (Stark & Tallal, 1981), a definition that provides a phenotype not confounded with broader cognitive impairments. This group can be compared to children with impairment in both language and nonverbal cognition, labeled nonspecific language impairment (NLI) in previous population-based studies (Tomblin et al., 1997), which excluded children with related conditions such as attention deficit disorders, neurological disorders such as epilepsy, or syndromic conditions such as Down syndrome.

Recently, there has been controversial advocacy for removing the nonverbal criterion from the definition of "language impairments," thereby establishing a single criterion of lower than age expectations in language regardless of nonverbal cognitive levels and some of the other impairments as well (although not including autism), under the assertion that there is no qualitative (although acknowledging a quantitative) difference between the two groups (Bishop et al., 2017). This collapsed grouping is labeled "developmental language disorders" (DLDs). There has been a movement toward replacing the term "SLI" with "DLD" in research reports when referring to studies that in fact used the more precise definition of SLI (Rudolph et al., 2019). This expanded definition would make it difficult to follow a legacy of precedents in the literature and to evaluate possible differences in causal pathways for nonverbal cognitive impairments and language impairments. No previous studies have addressed the possible implications of twin studies using the SLI versus NLI versus DLD criteria for grouping by affectedness, although one study examined SLI versus NLI (Hayiou-Thomas et al., 2005). This study addresses the gap by calculating heritability outcomes for the full sample and each of the two groups.

Overall, the study reported here is the first to consider in the same study these foci: heritability estimates, possible twinning effects, multiple precise phenotypes with standard score outcomes for benchmarking age expectations, and evaluation of competing methods of defining the categorical phenotype of language impairment cross-referenced to nonverbal cognitive impairment. The condensed literature review that follows is organized in that order of topics.

Heritability of Speech and Language in Contemporary Twin Studies

The available behavioral genetics studies of twins over the past 30 years provide valuable estimates of the heritability of language, speech, and cognitive behavioral phenotypes. Across studies, there is replication of statistically significant estimates of the heritability of language, speech, and nonverbal cognitive phenotypes in young children, showing genetic influences on these phenotypes early in children's development. At the same time, genetic effects are not uniformly strong across age, phenotypes, or levels of performance relative to age expectations, suggesting a need for further investigation to clarify how phenotypes, developmental trajectories, and levels of ability influence heritability.

Across the full range of ability levels, heritability in population-based samples increases with age (Hayiou-Thomas et al., 2012; Rice et al., 2018), and phenotypes vary with greater heritability estimates for measures of grammar, as compared to omnibus language measures or vocabulary or speech or nonverbal cognitive measures (Dale et al., 2000; Olson et al., 2011; Rice et al., 2014, 2018). These differences replicate in studies of population-based twin samples grouped categorically as affected or unaffected defined by language impairment, with the additional replicated finding that heritability estimates are higher for children at lower levels of performance (DeThorne et al., 2006; Hayiou-Thomas et al., 2014). Speech impairments show higher heritability than language impairments in a study of children at 4.5 years of age (0.56 vs. 0.34) using a latent trait method of estimation (Hayiou-Thomas et al., 2006). A grammar phenotype yields substantial heritability at 16 years of age (Dale et al., 2018).

The available evidence points toward genetic influences early in children's language acquisition that shift over time as new dimensions of language, speech, and nonverbal cognitive phenotypes emerge over the preschool years. At the molecular level, an epigenetic/gene regulation model was proposed to account for possible differences in genetically controlled timing mechanisms for early language acquisition (Rice, 2012), such that necessary onset signals controlled by as yet unidentified genetic influences early in children's development do not engage at the expected time in children with language impairments. Replicated growth models across different phenotypes consistently reveal that, once language acquisition starts in affected children, the growth curves parallel those of unaffected children, with a persistently lower level of language attainment relative to age peers (Rice & Hoffman, 2015; Rice et al., 2009, 2006, 1998). Heritability estimates at subsequent ages in the same sample of twins are needed for further clarification of possible genetic influences as language advances with age.

Twinning Effects on Language Acquisition

Although twin studies are a strong experimental paradigm for the estimation of heritability, we cannot assume all twin outcomes parallel the developmental arc of singleton children. Twins' development differs from singleton children, with a delayed onset of language acquisition, a phenomenon known as a "twinning effect" (Hay et al., 1987; Rutter et al., 2003; Taylor et al., 2018; Thorpe, 2006; Thorpe et al., 2003). Environmental effects are commonly assumed to be the cause of the twinning delay, attributable to the additional caregiver time demands for raising two babies of the same age, thereby reducing essential face-to-face talking time with the infants, which could affect quality of input such as semantic contingency and joint attention (Thorpe, 2006; Thorpe et al., 2003).

A limitation of available heritability studies is that the measurement methods and/or sample size has not been sensitive to possible twinning effects. In earlier studies, the psychometric properties of the language phenotype did not meet contemporary standards for controlling for within-age variance. Calculation of "language age" measures is based on group means for singleton children, not controlling for wide differences in group variances around the mean from young to older ages. "Age" scores are unreliable estimates of children's rank within their age level across age levels (Rutter et al., 2003). In recent population-based studies of the language of twins, such as the important Twins' Early Development Study (TEDS; Oliver & Plomin, 2007), the phenotypes were not benchmarked to standardized normative performance expectations for the general population. Instead, the measures were short-form versions of standardized assessments that provided percentile ranks within the twin sample, which provided classification of low performance (Dale et al., 2018; Hayiou-Thomas et al., 2014). A further limitation of previous studies is possible zygosity effects in language acquisition were not noted. If the cause of the twinning effect is attributable to two infants of the same age competing for the attention of a caretaker, as hypothesized in early studies (Thorpe et al., 2003), then the effects should be equivalent across monozygotic (MZ)/identical pairs and DZ pairs of twins.

A recent study was the first to provide replicated empirical evidence of a twinning effect in a population-based sample of twins 2–6 years of age. Standard scores benchmarked to age means and variances were used as phenotypes to allow for comparison data across ages. Relative to singleton age peers, twins were delayed in their acquisition of language across multiple phenotypes. Yet, the twinning effect was not evident for speech or nonverbal cognitive phenotypes (Rice et al., 2014, 2018), suggesting a more language-localized effect instead of more limited speech delays or pervasive cognitive delays in twins. Furthermore, the twinning effect was more pronounced for MZ than DZ twin pairs, a zygosity effect that disappeared by 6 years of age (Rice et al., 2018). The MZ/DZ differences are inconsistent with a simple environmental/competition effect attributable to multiple infants (Thorpe et al., 2003), because such social interaction effects should affect both types of twin pairs.

Instead, the zygosity effects point toward biological differences between MZ and DZ twin types in the developmental

pathways for language acquisition. This could be related to the hypothesized delay of the signals necessary for onset of language acquisition hypothesized in an epigenetic account of SLI (Rice, 2012). The twinning effects lessened between 4 and 6 years of age, reflecting a dynamic "catch-up" period relative to singleton norms (Rice et al., 2018), although not fully resolved at 6 years of age. We note that any such increased acceleration clearly would have to be later reduced or the children would become much better than their age peers, which clearly is not the case.

Prenatal and perinatal risks for late language emergence at 2 years of age in a population-based sample of twins found the same risks as the risks evident for delayed development in singletons: gestational diabetes, prolonged time to spontaneous respiration, and fetal growth restriction (Taylor et al., 2018). These three risks are well-known complications of twin pregnancy. Sociodemographic risk factors (e.g., low maternal education, socioeconomic area disadvantage) were not associated with increased odds of late language emergence in twins (Rice et al., 2014; Taylor et al., 2018). So far, there are no documented unique-to-twins neurobiological risks associated with language acquisition that can account for the twinning effect for language acquisition that is spared for speech and cognitive measures. There are, however, findings of possible epigenetic modifications that may play a role in the developmental consequences of early life events (Bloomfield, 2011). The very earliest periods of pregnancy may be an important period determining the developmental trajectory of the fetus.

At the empirical level, the twinning effect for language acquisition warrants careful consideration in estimates of heritability of language impairment in young twins. Delays associated with twinning do not necessarily signal inherent individual differences in language acquisition that are likely to persist into adulthood as are evident in children with SLI (Rice & Hoffman, 2015; Rice et al., 1999; Tomblin & Nippold, 2014). Longitudinal follow-up studies are needed to determine whether twin–singleton differences in language and cognition further diminish or resolve over time. This has important ramifications for the identification of language and cognitive impairments in twins. In the classic logic for behavioral genetics methods of estimating heritability, the twinning effect would contribute to estimates of common (shared) environment, thereby reducing sensitivity to detection of heritability effects (Rice et al., 2018), or the twinning effect could contribute to error estimates in the case of categorical phenotypes of affected versus unaffected children, if measurement accuracy is reduced for estimates of low performance on language tasks during a period of dynamic resolution of the twinning effect. Twin children at the age of 4 years who score in a low range of language acquisition could move out of a low range by the age of 6 years. This proportion is not known, nor is the stability of performance on nonverbal intelligence assessments in twins at this age level. Yet, these data are needed to inform decisions about definitions for the criteria of affectedness in the calculation of heritability estimates.

SLI Versus NLI in Twins

Differences between children with SLI versus NLI are potentially informative for clarifying the extent to which language impairments and nonverbal cognitive impairments share the same causal pathways, that is, whether the causes of language impairment differ as a function of low or typical nonverbal ability. Two population-based studies of singleton children (Norbury et al., 2016; Tomblin & Nippold, 2014; Tomblin et al., 1997) provided generalizations about differences in language acquisition between NLI and SLI groups. Children in the NLI group exhibit lower performance levels than those in the SLI group on speech and language assessments, in speed of language processing, and on some measures of processing capacity; furthermore, they tend to have more diffuse impairments across speech, language, social, and cognitive tasks. Also, the long-term outcomes in language and literacy are worse (Catts et al., 2002). Differences in long-term outcomes in morphosyntax are documented in detail (Rice et al., 2004), indicating delayed acquisition by the NLI group relative to the SLI group and differing profiles of recovery from overgeneralization (learning) errors that persist in the NLI group through fourth grade (about 9–10 years of age).

A limitation of our current understanding of causal pathways is that the available studies focused on children with language impairments, that is, SLI or NLI, compared to typically developing children. Under this design, children with low nonverbal cognitive abilities who did not have language impairments were excluded because they are not "typical." That is, they did not enter a group defined by a language impairment, and they did not enter a group defined as "typical." This left unexamined a very interesting group of children with low nonverbal cognitive abilities whose language abilities are in typical range or above (Rice, 2020). This group is largely unreported, perhaps because they would be identified only in population-based studies. There are no longitudinal data available on this group of children. There was one report of this group from the population-based Iowa study (Shriberg et al., 1999), which reported 12% of the full sample in this group.

To the best of our knowledge, there is only one previous study of twins comparing SLI and NLI groups of children (Hayiou-Thomas et al., 2005). This is a study of three hundred fifty-six 4.5-year-old children with low language ability and their twin partners (total $N = 712$). The sample for analysis was ascertained as at least one affected twin per pair. The children were assessed at home on multiple language and cognitive phenotypes analyzed as composite scores for language and nonverbal cognitive abilities. For the NLI group, genetic influence on language impairment was moderate (0.52); for the SLI group, it was 0.18 and not statistically different from 0. We note that a heritability level of around 0.50 is common for language measures in the TEDS. The 0.52 heritability estimate for NLI was not statistically different from the SLI group due to overlapping confidence intervals. Shared environmental effects were substantial for both groups. The conclusions

were that the findings pointed toward different causal pathways for language versus nonverbal cognitive impairments, perhaps due to a "double-hit" effect in the conjoined definition of affectedness.

Previous studies of singleton children suggest potential empirical challenges and pitfalls to avoid or minimize. Research definitions of affectedness are important elements of design. Classic empirical research definitions of SLI and NLI use inclusionary and exclusionary criteria based on norm-referenced standardized assessments (Norbury et al., 2016; Stark & Tallal, 1981; Tomblin et al., 1997). The inclusionary criterion requires performance on language assessments below typical range, usually defined as 1 *SD* or more below the age mean. The exclusionary criteria include nonverbal cognitive performance below typical range, hearing loss, and neurodevelopmental disorders. Two possible pitfalls are related to measurement issues. One is that it is important to maintain consistent definitions across groups. For example, in a study (Reilly et al., 2010) predicting SLI versus "low language" outcomes at 4 years of age, the inclusionary criteria for the "low language" group included children from non–English-speaking backgrounds, whereas these children were excluded from the SLI group. Furthermore, the methods for estimating low levels of performance per child differed for language versus nonverbal IQ variables: Normative population data were the basis for the language measures, whereas within-sample levels were the basis for the nonverbal IQ variables. Such inconsistencies work against straightforward interpretation of the outcomes. Another unavoidable potential empirical challenge in investigating possible differences between SLI and NLI groups is that the expected proportion of children who meet the NLI definition (low language + low nonverbal cognitive performance) is smaller than the SLI group (low language only), a likelihood based on the distributional properties of a "double hit" versus "single hit" criterion. In twin studies, the group sizes influence sensitivity to heritable effects and variance estimates, thereby constraining interpretations.

Purpose of This Study

This study focuses on twin children with language impairments, in the form of SLI or NLI, following classic experimental methods of defining affectedness (Tomblin et al., 1997). The study is the first to explore the relationship between language impairment phenotypes and nonverbal IQ over time in the preschool age range. The study follows a previous report of language acquisition of 2-year-old twins, which documented an early twinning effect (Rice et al., 2014), and a subsequent report of language acquisition across multiple phenotypes of twins at 4 and 6 years of age (Rice et al., 2018) focused on the twinning effect and estimates of heritability in the full sample but did not estimate heritability of language impairment.

There is no available evidence for the distribution of a population-level twin sample according to SLI or NLI inclusionary/exclusionary criteria because previous population-

level twin samples have not been assessed with the full norm-referenced standardized assessments that provide standard deviation from age-level group means. This study aims to fill these gaps in the research base, using consistent criteria across the SLI and NLI groups, in a longitudinal sample of twins at 4 and 6 years of age.

Research Questions

This study used the following research questions:

1. What percentage of twins meet criteria for SLI, observed language impairment in the absence of deficits in nonverbal IQ (i.e., standard scores ≥ 85), and NLI (nonverbal standard scores < 85)?

2. How frequently do twins meet language (and speech) impairment criteria at increasingly strict, that is, lower levels of performance, per phenotype, age, group, and consistency of group assignment over two age levels?

3. How do rates of proband-wise twin concordance and heritability of language impairments differ when using increasingly strict criteria for its designation, per phenotype, at each age and also across ages for persistent language impairments? This question looks at the sample as a whole.

4. What are the estimated heritability rates in twins with SLI versus NLI per phenotype, per age level? This question looks at the sample grouped according to levels of impairments of language.

Method

Ethics

This study was approved by the University of Kansas Institutional Review Board (#12582) and two institutions in Perth, Western Australia: Curtin University of Technology Human Research Ethics Committee (HR3/2001) and the Department of Health Western Australia Human Research Ethics Committee (2010/6). The study collected identified information from participants in Western Australia and followed approved procedures for protecting confidentiality. Small reimbursements for effort were provided to participants, such as small toys for children and movie vouchers for adolescents and adults.

Participants

The full sample comprised 1,354 children from 677 pairs, which included 109 MZ girls, 105 MZ boys, 117 DZ girls, 108 DZ boys, and 238 DZ opposite sex pairs. Because nonverbal intelligence scores were used as a grouping variable in the analyses, individuals without a nonverbal IQ score were removed. At the age of 4 years, 598 pairs had IQ scores from both twins, 11 pairs had IQ scores from only one twin, and 68 pairs had no IQ scores. At the age of 6 years, 629 pairs had IQ scores from both twins, one pair had IQ scores from only one twin, and 47 pairs

had no IQ scores. The final analysis sample consisted of 1,207 and 1,259 individuals at ages 4 and 6 years, respectively. This would be the maximum number of children available for the analyses at each time point per phenotype.

Details about the study design, sampling, and exclusionary criteria are given in an earlier report (Rice et al., 2018). Of note here are details about how possible confounding variables were addressed with exclusionary criteria. The relevant section from the 2018 article (p. 3) states the criteria as follows: Twins with exposure to languages other than English were excluded, based on a parent report questionnaire. Birth records and parent questionnaires were consulted to exclude children with known hearing impairment, neurological disorders, or development disorders, including Down syndrome, Angelman syndrome, cerebral palsy, cleft lip and/or palate, agenesis of the corpus callosum, and global developmental delay. At 4 and 6 years of age, the children's hearing was assessed via pure-tone screenings (500, 1000, 2000, and 4000 Hz) under headphones in everyday ambient noise in field testing. A pass was defined as a participant responding to each frequency in either the right or left ear at 25 or 30 dB.

Measures and Procedure

As in the previous study, the variables used in this study were derived from standardized tests, selected for a range of dimensions of language including speech production and providing sound psychometric properties for reliability and validity. All have independently ascertained norm-based standard scores with the exception of mean length of utterance (MLU) in morphemes, which has independently generated age norms from Rice's lab. These included the Columbia Mental Maturity Scale (CMMS; Burgemeister et al., 1972) for nonverbal intelligence (IQ), a pointing task for assessing conceptual development; the Peabody Picture Vocabulary Test–III (PPVT-III; Dunn & Dunn, 1997), a frequently used receptive vocabulary test; and the Test of Language Development–Primary: Third Edition (TOLD-P:3; Newcomer & Hammill, 1997), a psychometrically robust language assessment across different dimensions of language. The latter provided three scores: Spoken Language (a combined score of the following two scores, considered to be an omnibus score collapsed over multiple dimensions of language), Semantics, and Syntax. We used age-adjusted standard scores for these outcomes. The Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000), a picture-naming task for evaluating target speech sound accuracy, which also provided percentile scores for speech development. The Rice–Wexler Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001) is a research-developed assessment of finiteness marking in sentences with age-referenced normative data. The TEGI Composite measures production of third-person singular –s, past tense, BE auxiliary and copula, and DO auxiliary in obligatory sentence contexts. The TEGI Screener measures third-person –s and past tense production in sentences. We calculated standard scores

from the means and standard deviations provided in the manual. Finally, an additional variable was collected from an analysis of spontaneous language samples collected in the test sessions at 4 and 6 years of age. The MLU (Miller & Chapman, 2002; Rice et al., 2006, 2010) was calculated from the coded transcripts, following procedures used in Rice's lab (Rice et al., 2010). Affectedness was determined using the age-level means and standard deviations reported for an independent singleton sample from Rice's lab (Rice et al., 2010).

MLU can be unreliable if not calculated properly. Reliability is heavily dependent on the total number of utterances in the sample and the total number of complete and intelligible utterances, which constitute the denominators for calculating the means. A total of 200 utterances is accepted as a good base for adequate reliability within the usual time constraints for collecting the utterances (Gavin & Giles, 1996). In the twin sample of this study, the mean number of total utterances at 4 years of age was 202.92 ($SD = 68.56$); at 6 years of age, it was 214.62 ($SD = 84.42$). Complete and intelligible utterances at 4 years of age was 151.58 ($SD = 60.09$); at 6 years of age, it was 171.05 ($SD = 67.78$). The results per zygosity groupings for total utterances were as follows: 4 years of age, MZ, 198.43 ($SD = 67.24$), and DZ, 204.95 ($SD = 60.09$); 6 years of age, MZ, 212.86 ($SD = 90.32$), and DZ, 215.44 ($SD = 81.52$). For complete and intelligible utterances, the results were as follows: 4 years of age, MZ, 147.87 ($SD = 61.38$), and DZ, 153.26 ($SD = 59.47$); 6 years of age, MZ, 167.91 ($SD = 71.14$), and DZ, 172.53 (66.12). Validity of the MLU estimates in this study was supported by expected group means, as shown in Table 1, with higher performance for unaffected children.

Results

RQ 1: SLI and NLI by Language Outcome and Age

To examine the frequency with which twins meet inclusionary and exclusionary SLI criteria (i.e., observed language impairment in the absence of deficits in nonverbal IQ), we formed groups for the categorical phenotype (presence or absence of each type of impairment) using a cutoff of approximately −1 SD, following the criterion used in earlier studies of SLI (Rice & Hoffman, 2015; Rice et al., 1999). Thus, impairment per nonverbal IQ and all phenotypes of speech and language was assigned for standard scores of < 85, whereas impairment was assigned for GFTA Speech as a percentile of < 15 (approximately equivalent to a standard score of 85). SLI was defined as impaired in speech or language phenotype but not nonverbal IQ; NLI was defined as impaired in speech or language phenotype with nonverbal IQ standard scores of below 85.

Because this is the first report of this kind of cross-classification in a sample of twins with standardized scores relative to age norms with singleton children and it is important to illustrate the twinning effect differences between nonverbal IQ and omnibus language, we first present

Table 1. Sample sizes (*N*), percentages (%), means, and standard deviations (*SD*s) for groups defined as affected (Aff) or unaffected (Unaff) in nonverbal IQ (NVIQ) by language (Lang) for each phenotype by age.

Outcome	Age	NVIQ	Lang	Group	*n*	%	NVIQ score M	NVIQ score SD	Lang score M	Lang score SD
PPVT-III Vocabulary	4	Unaff	Unaff		871	72	106	11	101	11
			Aff	SLI	182	15	95	8	78	6
		Aff	Unaff		89	7	80	4	94	7
			Aff	NLI	64	5	78	4	75	8
	6	Unaff	Unaff		1,122	89	106	11	104	9
			Aff	SLI	78	6	98	9	78	6
		Aff	Unaff		41	3	80	5	97	7
			Aff	NLI	15	1	79	4	74	12
TOLD-P:3 Spoken Language	4	Unaff	Unaff		702	59	107	11	96	9
			Aff	SLI	338	28	98	10	79	4
		Aff	Unaff		50	4	80	4	90	5
			Aff	NLI	98	8	79	4	79	4
	6	Unaff	Unaff		891	71	108	10	99	10
			Aff	SLI	310	25	100	9	77	6
		Aff	Unaff		20	2	79	6	92	6
			Aff	NLI	36	3	80	4	71	8
TOLD-P:3 Semantics	4	Unaff	Unaff		733	62	106	11	96	9
			Aff	SLI	307	26	99	10	79	3
		Aff	Unaff		69	6	80	4	91	6
			Aff	NLI	80	7	79	4	78	4
	6	Unaff	Unaff		938	75	107	11	99	9
			Aff	SLI	263	21	101	9	76	6
		Aff	Unaff		27	2	79	5	93	7
			Aff	NLI	29	2	80	4	71	8
TOLD-P:3 Syntax	4	Unaff	Unaff		726	61	107	11	97	10
			Aff	SLI	314	26	98	9	78	4
		Aff	Unaff		61	5	80	4	91	5
			Aff	NLI	87	7	79	4	79	4
	6	Unaff	Unaff		914	73	107	11	99	10
			Aff	SLI	287	23	100	10	75	6
		Aff	Unaff		20	2	79	4	91	7
			Aff	NLI	36	3	80	5	72	8
GFTA-2 Speech	4	Unaff	Unaff		996	83	104	11	58	22
			Aff	SLI	48	4	101	12	8	4
		Aff	Unaff		139	12	80	4	55	22
			Aff	NLI	12	1	78	5	9	4
	6	Unaff	Unaff		1,100	88	106	11	45	18
			Aff	SLI	97	8	106	11	8	4
		Aff	Unaff		48	4	80	5	39	15
			Aff	NLI	7	1	79	3	9	5
MLU	4	Unaff	Unaff		816	70	105	11	104	14
			Aff	SLI	202	17	100	11	75	10
		Aff	Unaff		95	8	80	4	101	15
			Aff	NLI	54	5	79	4	72	10
	6	Unaff	Unaff		1,007	81	106	11	106	15
			Aff	SLI	187	15	104	10	77	7
		Aff	Unaff		41	3	80	4	106	16
			Aff	NLI	13	1	78	7	76	9
TEGI Composite	4	Unaff	Unaff		625	54	106	11	100	8
			Aff	SLI	393	34	101	10	65	15
		Aff	Unaff		70	6	80	4	98	8
			Aff	NLI	77	7	79	4	61	14
	6	Unaff	Unaff		912	73	107	11	103	9
			Aff	SLI	286	23	104	10	58	21
		Aff	Unaff		31	2	79	5	100	9
			Aff	NLI	25	2	80	4	56	24

(table continues)

Table 1. *(Continued).*

| Outcome	Age	NVIQ	Lang	Group	n	%	NVIQ score M	NVIQ score SD	Lang score M	Lang score SD
TEGI Screener	4	Unaff	Unaff		685	58	106	11	101	7
			Aff	SLI	342	29	100	11	65	14
		Aff	Unaff		73	6	80	4	99	7
			Aff	NLI	76	6	79	4	61	14
	6	Unaff	Unaff		982	78	106	11	104	7
			Aff	SLI	217	17	103	11	60	22
		Aff	Unaff		39	3	80	5	103	8
			Aff	NLI	17	1	79	5	50	25

Note. PPVT-III = Peabody Picture Vocabulary Test–III; SLI = specific language impairment; NLI = nonspecific language impairment; TOLD-P:3 = Test of Language Development–Primary: Third Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; MLU = mean length of utterance as calculated by SALT software; TEGI = Rice–Wexler Test of Early Grammatical Impairment.

descriptive information—their distributions and scatter plots for their associations at each age. Figures 1 and 2 show the means, standard deviations, and overall shape of the distribution for the nonverbal IQ (CMMS) and omnibus language measure (TOLD Spoken Language) at ages 4 and 6 years. As shown in Figure 1, for nonverbal IQ, average performance is within the expected range of 100 at 4 years of age ($M = 101.02$) and 6 years of age ($M = 104.63$), with a significant increase in performance with age: CMMS

paired-samples t test corrected for twin dependency across 4 and 6 years of age, $t(1222) = 9.19$, $p < .0001$, $d = 0.53$. Figure 2 shows the same information for the omnibus language measure (TOLD Spoken Language). At the age of 4 years, average performance is slightly below the expected score of 100, with a mean of 89.49 at the age of 4 years and 92.35 at the age of 6 years, also with a significant increase in performance with age: TOLD Spoken Language

Figure 1. Distribution of standard scores on nonverbal IQ (CMMS) at 4 and 6 years of age. CMMS = Columbia Mental Maturity Scale.

Figure 2. Distribution of standard scores on omnibus language (TOLD Spoken Language) at 4 and 6 years of age. TOLD-P:3 = Test of Language Development–Primary: Third Edition.
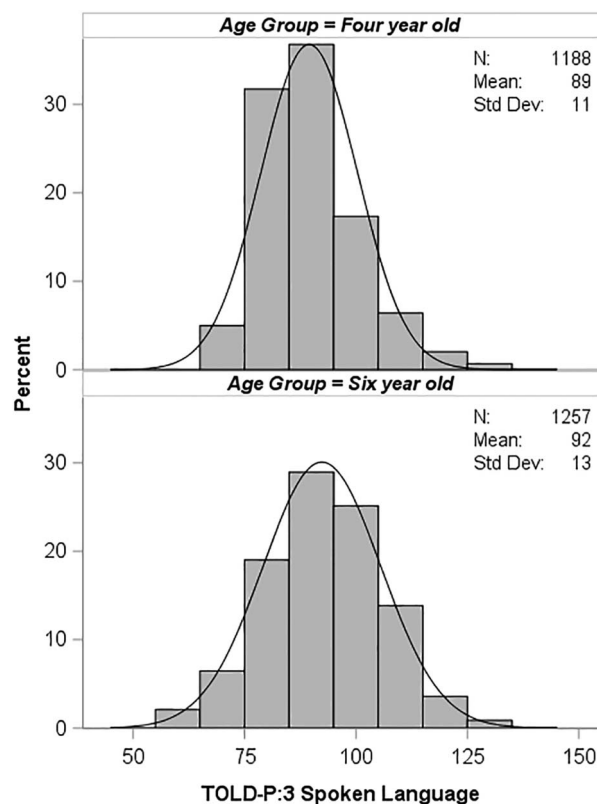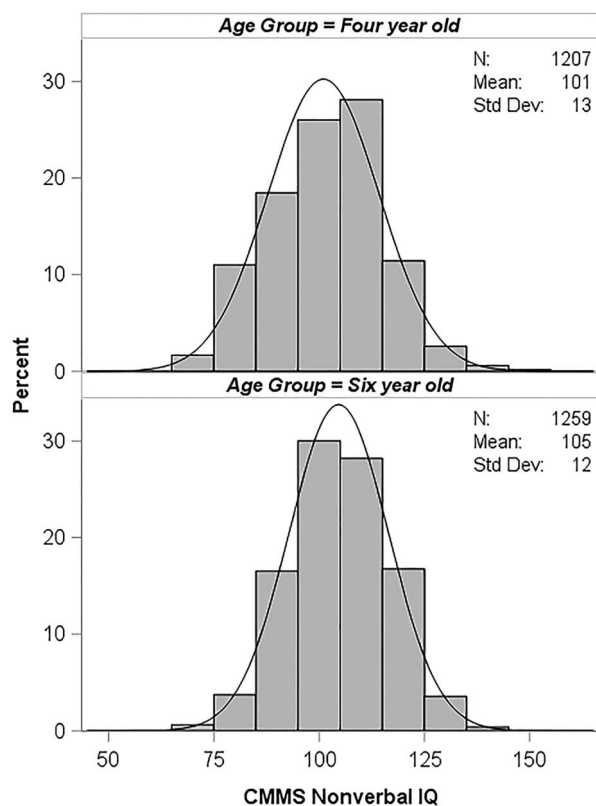
paired-samples *t* test corrected for twin dependency across 4 and 6 years of age, $t(1189) = 9.80$, $p < .0001$, $d = 0.57$. Note the twinning effect present in the distribution for the omnibus language measure in Figure 2 (i.e., a greater percentage of lower scores than expected) but not for nonverbal IQ in Figure 1, a finding we return to below.

Figures 3 and 4 show scatter plots between the CMMS and TOLD Spoken Language for ages 4 and 6 years, respectively, indicating how the twinning effects influence affectedness by SLI criteria at each age level. Figure 3 reports for the children at 4 years of age. For now, focus on the "−1 *SD* below the mean" line shown in the figures as the upper cut line differentiating groups, indicated by shading in the box. The SLI group, according to the −1 *SD* criterion, is 28% of the sample (bottom right) compared to the normative cell (unaffected on both variables, shown at the top right) of 59%; the percentage of children affected (low scores) on both measures is 8% (bottom left), with 4% affected on nonverbal IQ and within typical range on the language measure (top left). The Pearson correlation between the CMMS and TOLD Spoken Language at the age of 4 years shows a moderate positive relationship, $r = .482$, $p < .0001$. Shown in Figure 4 is the relation at 6 years of age. The SLI group is 25% (bottom right), the group unaffected on both measures is 71% (top right), the group affected on both is 3% (bottom left), and 2% of the sample is low on nonverbal IQ and unaffected on language (top left). The Pearson correlation between the CMMS and TOLD Spoken Language at the age of 6 years also indicates a moderate positive relationship, $r = .437$, $p < .0001$. From the distributional data, we can

Figure 3. Relationship between standard scores on nonverbal IQ (CMMS) and omnibus language (TOLD Spoken Language) at 4 years of age. CMMS = Columbia Mental Maturity Scale; TOLD-P:3 = Test of Language Development–Primary: Third Edition; SLI = specific language impairment.
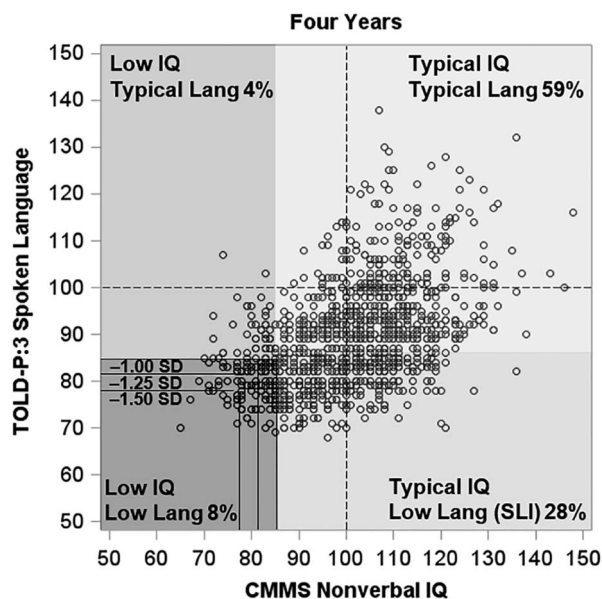
Figure 4. Relationship between standard scores on nonverbal IQ (CMMS) and omnibus language (TOLD Spoken Language) at 6 years of age. CMMS = Columbia Mental Maturity Scale; TOLD-P:3 = Test of Language Development–Primary: Third Edition; SLI = specific language impairment.
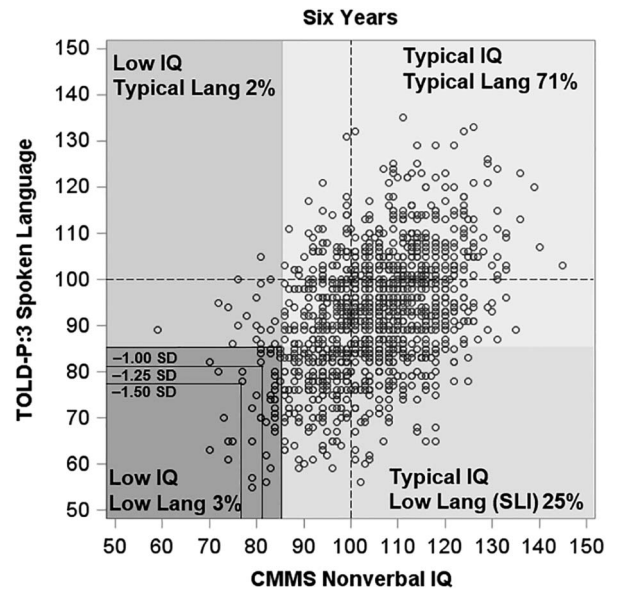
see that the pattern of twinning effects reported in the full sample (Rice et al., 2018) influences the categorical groupings by increasing the percentage of the unaffected group between ages 4 and 6 years from 59% to 71%, although the SLI group decreases only from 28% to 25%. Note that summing across the top row of children in Figure 3, with language standard scores above 85 (normal range and above), there are 63% at 4 years of age and, in Figure 4, 73% at 6 years of age, indicating accelerated language acquisition during this age range for some twins, although there is still a higher-than-expected percentage of twins with low language scores.

These patterns play out across the set of language measures, as shown in Table 1, with an exception for speech, which has notably lower levels of affectedness, a total of 5% at 4 years of age and 9% at 6 years of age, attributable to age-norming ceiling effects at the older age. Across all measures, as expected, the most frequently observed combination was no impairment in either nonverbal IQ or language, ranging from a minimum of 54% (TEGI Composite) to a maximum of 83% (GFTA Speech) at the age of 4 years and a minimum of 71% (TOLD Spoken Language) to a maximum of 89% (PPVT-III Vocabulary) at the age of 6 years.

Table 1 also shows that, although the number of children per group shifts over time and diagnostic grouping, the mean language/speech standard scores and standard deviations are consistent across groups, across time, and across phenotypes. Furthermore, in this sample of twins, there is no evidence of a "double hit"; the pattern is of nearly equivalent (within the standard error of measurement) means of children in the NLI group compared to those in the SLI group.

Table 1 shows replication of the existence of a group of children with typical or above language scores and nonverbal IQ scores of below 85. The pattern is higher percentages at 4 years of age than at 6 years of age throughout the phenotypes, generally more than 5% at 4 years of age (range: 4%–12%) than at 5 years of age (range: 2%–4%).

RQ 2: How Frequently Do Twins Meet Language (and Speech) Impairment Criteria at Increasingly Strict, i.e., Lower Levels of Performance, per Phenotype, Age, Group, and Consistency of Group Assignment Over Two Age Levels?

To examine how frequently twins would meet speech and language impairment criteria under increasingly strict criteria, we formed groups based on impairment using cutoffs of −1.00, −1.25, and −1.50 SD for each outcome, shown in Figures 3 and 4. For standard scores, these cutoffs were < 85, 81, or 77, respectively, whereas impairment was assigned for GFTA Speech as a percentile of < 15, 11, or 7, respectively. Of interest is the consistency of performance across the two different ages of samples. Table 2 provides the percentage of children affected on each measure at both times of measurement under increasingly strict definitions of affectedness. Note under this approach the Ns in the inconsistent cells (i.e., affected at one age only) are not shown in Table 2. With increasing strictness in definitions of affectedness, the percentage of the full sample in the affected category is reduced; that is, the sample size is

smaller when moving farther into the tail of the distribution. Across the different phenotypes, the percentage of affected children at −1.00 SD is greatest for language measures, particularly TOLD Spoken Language (19.6%) and TEGI Composite (17.9%). As shown in the table, children with no impairments at the age of 4 years were most likely to have no impairments at the age of 6 years. This effect is especially pronounced for nonverbal IQ, 96.9% unaffected at 4 and 6 years of age at −1.50 SD affectedness. From data not shown in the table, children were more likely to "outgrow" their affectedness between 4 and 6 years of age rather than the other way around. For example, for TOLD Spoken Language at −1.00 SD, 16.7% of children switched from affected at 4 years of age to unaffected at 6 years of age, versus 7.5% of children who were unaffected at 4 years of age but affected at 6 years of age. Similarly, at −1.25 SD on TOLD Spoken Language, 11.3% of children switched from affected at 4 years of age to unaffected at 6 years of age, versus 7.8% of children who were unaffected at 4 years of age but affected at 6 years of age.

RQ 3: Heritability Differences by Affectedness Severity Levels per Phenotype and Age

Moving to the level of twin pairs, the method of proband-wise twin concordance was used, in which at least one member of each pair is affected, generating a measure of the proportion of twins who have an affected twin. Concordance in SLI affectedness at each age was calculated separately for MZ and DZ pairs using a standard formula (Smith,

Table 2. Consistency in affectedness at the ages of 4 and 6 years by phenotype and affectedness criteria.

Outcome	−1.00 SD		−1.25 SD		−1.50 SD	
	Affected at 4 and 6 years of age	Unaffected at 4 and 6 years of age	Affected at 4 and 6 years of age	Unaffected at 4 and 6 years of age	Affected at 4 and 6 years of age	Unaffected at 4 and 6 years of age
PPVT-III Vocabulary	5.05%	78.56%	2.79%	86.58%	1.17%	92.43%
N = 1,110	N = 56	N = 872	N = 31	N = 961	N = 13	N = 1,026
TOLD-P:3 Spoken Language	19.60%	56.24%	10.21%	70.65%	3.19%	83.32%
N = 1,097	N = 215	N = 617	N = 112	N = 775	N = 35	N = 914
TOLD-P:3 Semantics	12.22%	57.52%	6.02%	74.38%	3.01%	81.95%
N = 1,097	N = 134	N = 631	N = 66	N = 816	N = 33	N = 899
TOLD-P:3 Syntax	16.86%	58.61%	7.38%	73.02%	4.65%	79.31%
N = 1,097	N = 185	N = 643	N = 81	N = 801	N = 51	N = 870
GFTA-2 Speech	1.91%	92.45%	1.18%	95.91%	0.82%	96.72%
N = 1,099	N = 21	N = 1,016	N = 13	N = 1,054	N = 9	N = 1,063
MLU	5.52%	68.78%	3.41%	77.16%	2.21%	85.27%
N = 1,086	N = 60	N = 747	N = 37	N = 838	N = 24	N = 926
TEGI Composite	17.90%	54.45%	15.58%	60.67%	12.52%	66.88%
N = 1,078	N = 193	N = 587	N = 168	N = 654	N = 135	N = 721
TEGI Screener	13.42%	61.03%	10.66%	66.36%	8.82%	70.77%
N = 1,088	N = 146	N = 664	N = 116	N = 722	N = 96	N = 770
CMMS Nonverbal IQ	1.26%	84.47%	0.09%	92.46%	0.00%	96.86%
N = 1,114	N = 14	N = 941	N = 1	N = 1030	N = 0	N = 1,079

Note. Ns on the far-left column only include those with data at both 4 and 6 years of age. Ns inside middle cells do not include participants who switched affectedness status between 4 and 6 years of age. PPVT-III = Peabody Picture Vocabulary Test–III; TOLD-P:3 = Test of Language Development–Primary: Third Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; MLU = mean length of utterance as calculated by SALT software; TEGI = Rice–Wexler Test of Early Grammatical Impairment; CMMS = Columbia Mental Maturity Scale.

1974) for samples in which both twins were independently ascertained—as was the case in this study. The proband-wise concordance rate is calculated as $2*C / (2*C + D)$, in which C is the number of twin pairs in which both twins are affected (i.e., concordant) and D is the number of twin pairs in which only one twin is affected (i.e., discordant). The resulting proband-wise concordance rate indicates the probability of affectedness among cotwins of affected twins. Concordance rates are reported here as percentages for ease of interpretability. Table 3 reports the percent affected and proband-wise concordance rates for the MZ and DZ pairs for each phenotype at three levels of severity (−1.00, −1.25, and −1.50 SD). Note that, as expected, for most measures, a higher percentage of MZ twins were affected than DZ twins (e.g., for PPVT vocabulary, −1.00 SD affectedness at the age of 4 years, 25.5% of MZ twins were affected vs. 18% of DZ twins). Similarly, proband-wise concordances were higher in MZ than DZ pairs, indicating stronger likelihood of both twins within a pair being affected for the MZ pairs than the DZ pairs. For example, at the age of 4 years on the PPVT at −1.00 SD affectedness, where one MZ twin is affected at −1.00 SD, the probability that the cotwin will also be affected is 60.4%. For DZ twins on the same measure and affectedness criterion, the probability of cotwin affectedness is 35.6%, almost half the probability as for MZ. With greater phenotypic severity, the concordance estimates decrease; for example, PPVT for MZ twins at 4 years of age is 60.4% at −1.00 SD affectedness versus 26.7% at −1.50 SD affectedness. Across phenotypes, TOLD Syntax and TEGI Screener have the highest levels of concordance; for example, at a criterion of −1.00 SD for affectedness, MZ concordance at 4 years of age was 74.6% and 78.7% for TOLD Syntax and TEGI Screener, respectively.

We then examined to what extent heritability of the affectedness designation varied by phenotype, affectedness criterion, and age using structural equation models estimated via diagonally weighted least squares in Mplus v.8 (Muthén & Muthén, 1998–2017). These models use a probit link function to predict binary affectedness from a random intercept factor, which captures the tetrachoric correlation of affectedness across twins from the same pair, making use of all available data. Given estimation difficulties arising from empty or near-empty cells in models for both ages at once, we instead report estimates derived from

Table 3. Percent affected and percent proband-wise concordance (Con) by phenotype, zygosity, age, and affectedness (Aff) criteria.

Outcome	Aff criterion	Age 4				Age 6			
		MZ twins		DZ twins		MZ twins		DZ twins	
		Aff	Con	Aff	Con	Aff	Con	Aff	Con
PPVT-III Vocabulary	−1.00 SD	25.5	60.4	18.0	35.6	10.7	51.2	5.9	16.3
	−1.25 SD	14.3	41.5	11.4	34.8	6.5	46.2	3.8	25.8
	−1.50 SD	8.3	26.7	6.1	28.6	3.7	53.3	1.9	26.7
TOLD-P:3 Spoken Language	−1.00 SD	41.9	71.1	34.3	57.5	29.0	68.4	26.8	47.4
	−1.25 SD	26.9	56.3	20.3	44.0	21.8	72.7	17.7	38.7
	−1.50 SD	9.9	45.7	8.2	25.0	14.6	64.4	10.7	31.1
TOLD-P:3 Semantics	−1.00 SD	37.1	59.7	30.5	47.7	23.3	53.2	23.2	38.6
	−1.25 SD	18.7	35.8	16.1	26.0	18.3	56.8	14.8	28.8
	−1.50 SD	10.1	28.6	10.1	22.8	14.1	56.1	11.3	25.0
TOLD-P:3 Syntax	−1.00 SD	39.7	74.6	31.0	50.4	29.2	69.5	24.0	46.1
	−1.25 SD	21.9	51.3	16.4	36.9	22.3	66.7	15.8	32.8
	−1.50 SD	15.7	42.9	10.2	30.0	18.8	71.1	12.7	31.8
GFTA-2 Speech	−1.00 SD	7.4	56.0	4.3	37.5	6.5	53.8	4.2	5.6
	−1.25 SD	4.5	66.7	2.9	43.5	4.3	58.8	1.8	13.3
	−1.50 SD	3.7	61.5	2.2	47.1	3.3	61.5	1.4	16.7
MLU	−1.00 SD	24.8	55.2	20.6	29.8	16.9	47.8	15.6	25.8
	−1.25 SD	17.9	50.8	15.0	29.3	12.8	27.5	10.3	16.1
	−1.50 SD	12.4	60.5	10.3	20.5	7.6	40.0	5.9	4.1
TEGI Composite	−1.00 SD	46.3	78.0	37.7	61.4	31.3	69.8	21.7	41.1
	−1.25 SD	40.7	74.3	32.2	58.3	26.9	75.9	19.4	38.8
	−1.50 SD	34.6	73.9	27.1	55.1	22.4	75.6	15.7	37.3
TEGI Screener	−1.00 SD	41.8	78.7	32.7	57.6	23.4	55.3	16.4	40.0
	−1.25 SD	35.1	76.8	28.3	55.8	19.9	60.0	13.7	37.6
	−1.50 SD	30.4	77.8	24.5	54.5	16.9	61.8	11.6	36.4
CMMS Nonverbal IQ	−1.00 SD	12.8	25.0	12.6	25.5	5.7	52.2	3.9	6.1
	−1.25 SD	6.8	8.0	5.8	25.0	2.0		1.9	
	−1.50 SD	2.9	20.0	1.8	13.3	1.2		0.9	

Note. Proband-wise concordance for −1.25 and −1.50 SD affectedness on CMMS Nonverbal IQ at the age of 6 years was incalculable due to the small number of affected individuals (i.e., there were no instances of both members of a twin pair affected). MZ = monozygotic; DZ = dizygotic; PPVT-III = Peabody Picture Vocabulary Test–III; TOLD-P:3 = Test of Language Development–Primary: Third Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; MLU = mean length of utterance as calculated by SALT software; TEGI = Rice–Wexler Test of Early Grammatical Impairment.

Table 4. Intraclass correlation estimates (Est), standard errors (SEs), and lower and upper 95% confidence limits (LCL, UCL) by phenotype, zygosity, age, and affectedness (Aff) criteria.

Outcome	Aff criterion	Age	MZ twins				DZ twins			
			Est	SE	LCL	UCL	Est	SE	LCL	UCL
PPVT-III Vocabulary	−1.00 SD	4	.72	.08	.51	.84	.39	.09	.16	.54
		6	.75	.10	.50	.90	.32	.16	.00	.58
	−1.25 SD	4	.59	.12	.28	.79	.51	.10	.26	.67
		6	.76	.11	.46	.95	.57	.15	.20	.81
	−1.50 SD	4	.47	.18	.01	.76	.54	.12	.22	.74
		6	.85	.10	.58	1.00	.66	.18	.21	.95
TOLD-P:3 Spoken Language	−1.00 SD	4	.73	.07	.55	.84	.55	.07	.38	.65
		6	.79	.06	.63	.89	.46	.07	.27	.58
	−1.25 SD	4	.63	.09	.40	.78	.52	.08	.32	.65
		6	.87	.05	.75	.95	.46	.08	.24	.60
	−1.50 SD	4	.72	.11	.43	.91	.42	.13	.10	.63
		6	.84	.06	.68	.94	.47	.10	.21	.64
TOLD-P:3 Semantics	−1.00 SD	4	.55	.09	.31	.70	.41	.08	.22	.54
		6	.61	.09	.38	.77	.35	.08	.14	.49
	−1.25 SD	4	.41	.13	.07	.63	.26	.11	.00	.44
		6	.72	.08	.51	.86	.33	.10	.08	.50
	−1.50 SD	4	.47	.17	.04	.74	.34	.12	.02	.54
		6	.77	.08	.56	.90	.34	.11	.05	.52
TOLD-P:3 Syntax	−1.00 SD	4	.80	.06	.65	.89	.46	.07	.27	.58
		6	.80	.06	.65	.90	.48	.07	.30	.61
	−1.25 SD	4	.61	.10	.36	.78	.45	.09	.22	.60
		6	.83	.06	.68	.92	.39	.09	.15	.54
	−1.50 SD	4	.57	.12	.26	.77	.46	.11	.18	.64
		6	.88	.05	.76	.96	.44	.10	.18	.60
GFTA-2 Speech	−1.00 SD	4	.84	.09	.61	.98	.72	.11	.44	.90
		6	.84	.09	.62	.99	.08	.23	.00	.46
	−1.25 SD	4	.93	.06	.78	1.00	.79	.10	.53	.96
		6	.92	.07	.74	1.00	.47	.25	.00	.88
	−1.50 SD	4	.91	.08	.70	1.00	.84	.10	.58	1.00
		6	.91	.08	.70	1.00	.56	.24	.00	.96
MLU	−1.00 SD	4	.64	.09	.40	.79	.22	.10	.00	.38
		6	.62	.10	.35	.79	.25	.10	.00	.41
	−1.25 SD	4	.66	.10	.40	.83	.34	.10	.08	.51
		6	.36	.15	.00	.61	.17	.13	.00	.38
	−1.50 SD	4	.83	.08	.63	.95	.28	.13	.00	.49
		6	.67	.13	.33	.88	−.07	.21	.00	.28
TEGI Composite	−1.00 SD	4	.82	.05	.69	.91	.58	.06	.43	.68
		6	.78	.06	.63	.88	.43	.08	.23	.56
	−1.25 SD	4	.81	.06	.66	.90	.60	.06	.44	.70
		6	.88	.04	.77	.95	.43	.08	.21	.56
	−1.50 SD	4	.83	.06	.69	.92	.61	.06	.45	.72
		6	.90	.04	.79	.96	.47	.09	.25	.61
TEGI Screener	−1.00 SD	4	.85	.05	.73	.93	.58	.06	.42	.69
		6	.65	.09	.42	.79	.50	.08	.29	.64
	−1.25 SD	4	.86	.05	.74	.94	.61	.06	.45	.71
		6	.75	.08	.56	.88	.51	.09	.28	.65
	−1.50 SD	4	.89	.04	.78	.96	.64	.06	.47	.74
		6	.80	.07	.62	.91	.54	.09	.30	.69
CMMS Nonverbal IQ	−1.00 SD	4	.32	.16	.00	.58	.32	.11	.03	.50
		6	.81	.10	.56	.98	.12	.24	.00	.51
	−1.25 SD	4	.08	.27	.00	.52	.48	.13	.14	.70
		6	.43	.09	.20	.57	.24	.10	.00	.40
	−1.50 SD	4	.53	.27	.00	.98	.44	.25	.00	.85
		6	.61	.06	.44	.72	.51	.06	.36	.61

Note. MZ = monozygotic; DZ = dizygotic; PPVT-III = Peabody Picture Vocabulary Test–III; TOLD-P:3= Test of Language Development–Primary: Third Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; MLU = mean length of utterance as calculated by SALT software; TEGI = Rice–Wexler Test of Early Grammatical Impairment; CMMS = Columbia Mental Maturity Scale.

Table 5. Heritability, common environment, and residual error estimates (Est), standard errors (SEs), and lower and upper 95% confidence limits (LCL, UCL) by phenotype, zygosity, age, and affectedness (Aff) criteria.

Outcome	Aff criterion	Age	h^2 heritability				c^2 common environment				e^2 residual error			
			Est	SE	LCL	UCL	Est	SE	LCL	UCL	Est	SE	LCL	UCL
PPVT-III Vocabulary	−1.00 SD	4	.64	.24	.03	1.00	.07	.20	.00	.39	.29	.08	.08	.41
		6	.86	.37	.00	1.00	.00	.33	.00	.43	.36	.10	.01	.41
	−1.25 SD	4	.16	.31	.00	.66	.43	.23	.00	.81	.41	.12	.10	.61
		6	.37	.37	.00	.98	.39	.31	.00	.90	.24	.11	.00	.43
	−1.50 SD	4	−.15	.43	.00	.56	.61	.30	.00	1.00	.53	.18	.08	.82
		6	.37	.41	.00	1.00	.48	.36	.00	1.00	.15	.10	.00	.32
TOLD-P:3 Spoken Language	−1.00 SD	4	.36	.19	.00	.67	.37	.15	.00	.61	.28	.07	.10	.39
		6	.65	.19	.16	.96	.13	.16	.00	.39	.22	.06	.06	.32
	−1.25 SD	4	.23	.24	.00	.62	.40	.18	.00	.70	.37	.09	.13	.52
		6	.83	.19	.34	1.00	.05	.17	.00	.33	.13	.05	.01	.20
	−1.50 SD	4	.59	.34	.00	1.00	.13	.28	.00	.58	.28	.11	.00	.47
		6	.73	.24	.12	1.00	.11	.21	.00	.46	.16	.06	.00	.27
TOLD-P:3 Semantics	−1.00 SD	4	.28	.24	.00	.67	.27	.18	.00	.56	.45	.09	.21	.60
		6	.52	.25	.00	.93	.09	.19	.00	.40	.39	.09	.15	.54
	−1.25 SD	4	.30	.34	.00	.85	.11	.25	.00	.52	.59	.13	.25	.81
		6	.79	.26	.12	1.00	.00	.22	.00	.29	.34	.08	.07	.41
	−1.50 SD	4	.26	.41	.00	.94	.21	.30	.00	.70	.53	.17	.10	.80
		6	.87	.28	.16	1.00	.00	.24	.00	.29	.33	.08	.02	.37
TOLD-P:3 Syntax	−1.00 SD	4	.69	.19	.21	.99	.11	.16	.00	.37	.20	.06	.05	.30
		6	.63	.19	.15	.94	.17	.16	.00	.43	.20	.06	.05	.30
	−1.25 SD	4	.33	.27	.00	.77	.28	.21	.00	.62	.39	.10	.13	.55
		6	.88	.22	.31	1.00	.00	.20	.00	.27	.23	.06	.03	.27
	−1.50 SD	4	.21	.32	.00	.74	.36	.25	.00	.77	.43	.12	.12	.63
		6	.90	.22	.34	1.00	.00	.20	.00	.32	.13	.05	.00	.19
GFTA-2 Speech	−1.00 SD	4	.24	.28	.00	.70	.60	.24	.00	.99	.16	.09	.00	.31
		6	1.52	.50	.25	1.00	.00	.47	.00	.10	.84	.09	.00	.30
	−1.25 SD	4	.28	.24	.00	.67	.66	.21	.10	1.00	.07	.06	.00	.17
		6	.89	.52	.00	1.00	.03	.51	.00	.86	.08	.07	.00	.20
	−1.50 SD	4	.14	.25	.00	.56	.76	.21	.22	1.00	.10	.08	.00	.23
		6	.69	.51	.00	1.00	.22	.49	.00	1.00	.10	.08	.00	.23
MLU	−1.00 SD	4	.83	.27	.14	1.00	.00	.21	.00	.15	.56	.09	.12	.52
		6	.74	.29	.00	1.00	.00	.23	.00	.25	.51	.10	.12	.55
	−1.25 SD	4	.64	.29	.00	1.00	.02	.23	.00	.40	.34	.10	.08	.51
		6	.38	.40	.00	1.00	.00	.30	.00	.47	.67	.15	.25	.90
	−1.50 SD	4	1.09	.30	.32	1.00	.00	.27	.00	.18	.44	.08	.00	.30
		6												
TEGI Composite	−1.00 SD	4	.48	.16	.06	.75	.34	.13	.00	.56	.18	.05	.04	.26
		6	.71	.20	.19	1.00	.08	.17	.00	.36	.22	.06	.06	.32
	−1.25 SD	4	.42	.17	.00	.70	.39	.14	.04	.62	.19	.06	.05	.29
		6	.91	.19	.43	1.00	.00	.17	.00	.25	.15	.04	.01	.19
	−1.50 SD	4	.43	.17	.00	.71	.40	.14	.04	.63	.17	.06	.03	.26
		6	.86	.19	.37	1.00	.04	.18	.00	.33	.10	.04	.00	.17
TEGI Screener	−1.00 SD	4	.54	.16	.13	.80	.31	.14	.00	.54	.15	.05	.03	.23
		6	.29	.24	.00	.69	.36	.19	.00	.66	.36	.09	.13	.50
	−1.25 SD	4	.51	.16	.10	.77	.36	.14	.01	.58	.14	.05	.02	.21
		6	.49	.23	.00	.86	.27	.19	.00	.58	.25	.08	.06	.37
	−1.50 SD	4	.50	.15	.11	.75	.39	.13	.04	.61	.11	.04	.01	.18
		6	.52	.23	.00	.90	.28	.20	.00	.60	.20	.07	.02	.32
CMMS Nonverbal IQ	−1.00 SD	4	.01	.39	.00	.65	.31	.27	.00	.76	.68	.16	.27	.95
		6	1.39	.51	.07	1.00	.00	.48	.00	.22	.76	.10	.00	.35
	−1.25 SD	4	−.81	.60	.00	.18	.89	.38	.00	1.00	.92	.27	.23	1.00
		6	.37	.26	.00	.80	.06	.21	.00	.41	.57	.09	.35	.72
	−1.50 SD	4	.18	.74	.00	1.00	.35	.57	.00	1.00	.47	.27	.00	.92
		6	.20	.17	.00	.48	.41	.13	.07	.63	.39	.06	.23	.50

Note. Negative heritability estimates occurred when the intraclass correlation for dizygotic twins was greater than the intraclass correlation for monozygotic twins, and heritability estimates greater than 1 occurred in models where the intraclass correlation for monozygotic twins was .5 greater than the intraclass correlation for dizygotic twins. LCLs and UCLs were truncated at 0 and 1, respectively; c^2 estimates ≤ 0 were converted to 0. Blank rows indicate that the model did not estimate. PPVT-III = Peabody Picture Vocabulary Test–III; TOLD-P:3 = Test of Language Development–Primary: Third Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; MLU = mean length of utterance as calculated by SALT software; TEGI = Rice–Wexler Test of Early Grammatical Impairment; CMMS = Columbia Mental Maturity Scale.

separate models for each phenotype and age. Given the arbitrary assignment of Twins 1 and 2, all parameters were constrained to be the same across Twins 1 and 2 within zygosity, but all parameters differed by zygosity. Results from these models are shown in Tables 4 and 5, including estimates, standard errors, and 95% confidence intervals for each source of variance.

First, Table 4 provides the intraclass correlations (ICCs) by age and zygosity for each outcome, calculated as ICC = random intercept variance divided by random intercept variance + 1 (in which 1 is the residual variance, as fixed for identification in the probit link function using the THETA parameterization in Mplus). These ICCs reflect the proportion of total variance due to mean differences between twin pairs (with a range of 0–1); said differently, the ICC is the correlation of twins from the same pair. As expected, these ICCs were generally higher for MZ twin pairs, with two exceptions at 4 years of age: PPVT at −1.50 SD and CMMS at −1.25 SD. Note that ICCs were highest for speech (GFTA) and increased with level of affectedness severity but not with age. ICCs for TEGI Composite increased with severity and, inconsistently, with age; ICCs for TOLD Spoken Language and Syntax increased with severity and age.

Table 5 provides the heritability results by age for each outcome. The proportion of variance due to heritability (i.e., shared genes) was calculated using the difference in ICC between MZ and DZ twins as $h^2 = 2*(ICC_{MZ} − ICC_{DZ})$. The proportion of variance due to common environment was calculated as $c^2 = ICC_{MZ} − h^2$ (constrained to be ≥ 0), and the proportion of unexplained variance was calculated as $e^2 = 1 − (h^2 + c^2)$. The proportions of variance attributable to heritability (h^2), as expected, increased with age and severity and varied across phenotypes. We note two instances of negative heritability values, as expected, for the PPVT and CMMS at 4 years of age, given the higher ICC for DZ than MZ twins noted in Table 4. Also, several heritability values in Table 5 exceed 1.0, an expected outcome when the difference of MZ minus DZ ICCs is more than 0.5: GFTA at 6 years of age and −1 SD, MLU at 4 years of age and −1.50 SD, and CMMS at 6 years of age and −1.00 SD. Otherwise, the highest h^2 is .91 for the TEGI Composite at 6 years of age and −1.25 SD, followed by TOLD Syntax at 6 years of age (.88 and .90 at −1.25 and −1.50 SD, respectively), TOLD Semantics at 6 years of age (.87 at −1.50 SD), TOLD Spoken Language at 6 years of age (.83 at −1.25 SD), and PPVT at 6 years of age (.86 at −1.00 SD). Overall, the heritability estimates across the phenotypes suggest inherited influences, albeit with some noise, as expected given the binary affectedness designation.

RQ 4: What Are the Estimated Heritability Rates in Twins With SLI Versus NLI per Phenotype, per Age Level?

We next compared heritability of affectedness at −1.00 SD for each phenotype at 4 and 6 years of age

between twins with SLI and NLI, where SLI is defined as affected at −1.00 SD per phenotype and unaffected on CMMS nonverbal IQ at −1.00 SD and NLI is defined as affected at −1.00 SD on both the phenotype and CMMS nonverbal IQ. This level provides the highest numbers of affected children per group. Models were estimated as described above, with values derived from separate models for each phenotype, grouping (SLI/NLI), and age. Results from these models are shown in Table 6 and in Table 7 are compared to the heritability values from models where SLI and NLI were combined in the previous section. For some phenotypes, only a small number of twins met the criteria for NLI (e.g., N = 7 NLI for GFTA at the age of 6 years; see Table 1). Because of the small NLI sample size for some measures, heritability values are missing from Table 6 for the following: PPVT at 4 and 6 years of age, GFTA Speech at 4 and 6 years of age, MLU at 6 years of age, and TEGI Composite and Screener at 6 years of age. Similarly, there were four instances of negative heritability values (i.e., not supportive of heritability) within the NLI group at the age of 4 years, as expected: TOLD Spoken Language and Semantics, MLU, and TEGI Composite. There was also one instance of a heritability value exceeding 1.0 (supportive of heritability) in the SLI group, for GFTA Speech at 6 years of age, an expected outcome when the difference of MZ minus DZ ICCs is more than .5.

For the remaining models, as shown in Table 6, in the SLI group, the highest h^2 is .91 for MLU at the age of 4 years and .74 at the age of 6 years, followed by .71 for TEGI Composite and .64 for PPVT, both at 6 years of age. In the NLI group, the highest h^2 values are at 6 years of age for TOLD Semantics (.86), Syntax (.85), and Spoken Language (.74).

Table 7 presents a comparison of heritability values across the full sample, SLI and NLI groups discussed in the previous section. Despite the incomplete heritability results for the NLI group, heritability tends to increase with age, except for the TEGI Screener, TOLD Syntax, and MLU in the full sample group and TEGI Screener and MLU in the SLI group. Ignoring h^2 values greater than 1 (either negative or positive), the highest h^2 values are for PPVT in the full sample group at 6 years of age of .86, versus .64 in the SLI group, and for TOLD Spoken Language (.74), Semantics (.86), and Syntax (.85) in the NLI group at 6 years of age. MLU also yielded high heritability estimates at 4 years of age (.83 for the full sample group, .91 for the SLI group) and at 6 years of age (.74 for the full sample group and for the SLI group). The TEGI Composite at 6 years of age had heritability estimates of .71 for the full sample group and the SLI group. These heritability estimates are greater than the .5 level often reported in earlier studies. The findings do not replicate the earlier report of heritability for NLI but not SLI (Hayiou-Thomas et al., 2005).

Discussion

A brief recap of the motivation for this study is in order. It is a follow-up to a previous study reporting a twinning

Table 6. Heritability, common environment, and residual error estimates (Est), standard errors (SEs), and lower and upper 95% confidence limits (LCL, UCL) by phenotype, zygosity, and age for the specific language impairment (SLI) and nonspecific language impairment (NLI) groups at −1.00 SD affectedness.

Outcome	Group	Age	h^2 heritability				c^2 common environment				e^2 residual error			
			Est	SE	LCI	UCI	Est	SE	LCI	UCI	Est	SE	LCI	UCI
PPVT-III Vocabulary	SLI	4	.35	.32	.00	.87	.14	.25	.00	.55	.51	.12	.21	.71
		6	.64	.44	.00	1.00	.00	.37	.00	.57	.44	.14	.04	.63
	NLI	4												
		6												
TOLD-P:3 Spoken Language	SLI	4	.34	.26	.00	.76	.18	.19	.00	.50	.48	.10	.23	.64
		6	.45	.22	.00	.81	.24	.17	.00	.52	.32	.08	.11	.45
	NLI	4	−.50	.62	.00	.52	.55	.39	.00	1.00	.95	.28	.24	1.00
		6	.74	.57	.00	1.00	.03	.52	.00	.88	.23	.13	.00	.44
TOLD-P:3 Semantics	SLI	4	.44	.28	.00	.91	.00	.21	.00	.34	.58	.11	.29	.75
		6	.60	.27	.00	1.00	.00	.21	.00	.31	.46	.10	.16	.60
	NLI	4	−.42	.63	.00	.62	.56	.41	.00	1.00	.86	.28	.14	1.00
		6	.86	.53	.00	1.00	.03	.51	.00	.86	.12	.09	.00	.26
TOLD-P:3 Syntax	SLI	4	.45	.24	.00	.84	.17	.19	.00	.48	.38	.09	.15	.52
		6	.55	.22	.00	.90	.17	.17	.00	.46	.28	.08	.09	.40
	NLI	4	.27	.47	.00	1.00	.24	.34	.00	.79	.49	.19	.00	.80
		6	.85	.56	.00	1.00	.00	.52	.00	.81	.24	.12	.00	.39
GFTA-2 Speech	SLI	4	.00	.36	.00	.60	.72	.27	.02	1.00	.28	.14	.00	.50
		6	1.22	.32	.39	1.00	.00	.30	.00	.14	.48	.07	.00	.24
	NLI	4												
		6												
MLU	SLI	4	.91	.34	.03	1.00	.00	.26	.00	−.02	.98	.13	.21	.75
		6	.74	.32	.00	1.00	.00	.24	.00	.22	.62	.12	.14	.63
	NLI	4	−.75	.65	.00	.32	.89	.44	.00	1.00	.87	.28	.16	1.00
		6												
TEGI Composite	SLI	4	.33	.21	.00	.66	.37	.16	.00	.62	.31	.08	.11	.43
		6	.71	.21	.17	1.00	.07	.18	.00	.36	.23	.06	.06	.33
	NLI	4	−.06	.50	.00	.75	.45	.35	.00	1.00	.62	.20	.10	.95
		6												
TEGI Screener	SLI	4	.26	.22	.00	.63	.38	.17	.00	.66	.36	.09	.15	.50
		6	.26	.27	.00	.70	.32	.20	.00	.66	.42	.10	.17	.58
	NLI	4	.18	.45	.00	.92	.34	.34	.00	.89	.48	.17	.04	.77
		6												

Note. Negative heritability estimates occurred when the intraclass correlation for dizygotic twins was greater than the intraclass correlation for monozygotic twins, and heritability estimates greater than 1 occurred in models where the intraclass correlation for monozygotic twins was .5 greater than the intraclass correlation for dizygotic twins. LCLs and UCLs were truncated at 0 and 1, respectively; c^2 estimates of ≤ 0 were converted to 0. Heritability estimates are not reported for combinations in which one of the four cells in the Twin × Affectedness combination was empty within zygosity (for which the associated tetrachoric correlation is not reliable due to small Ns within a group). PPVT-III = Peabody Picture Vocabulary Test–III; TOLD-P:3 = Test of Language Development–Primary: Third Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; MLU = mean length of utterance as calculated by SALT software; TEGI = Rice–Wexler Test of Early Grammatical Impairment; CMMS = Columbia Mental Maturity Scale.

effect on language acquisition in a relatively large longitudinal sample of twins at the ages of 4 and 6 years (Rice et al., 2018). The twinning effect, by definition, is one of the ways in which twins may differ from singleton children. In this case, twin children, as a group, are likely to be delayed in early language acquisition relative to the norms for singleton children, an outcome more likely in MZ than DZ twin pairs. Furthermore, although the twinning effect on language acquisition is persistent over time, up to 6 years of age and possibly beyond, it is not evident for nonverbal cognitive development. These new facts were not woven into questions, methods, and interpretations of previous studies of the heritability of language impairments in children, which is what this study does. It provides new evidence of the implications of a twinning effect on the identification of SLI and NLI in a sample of twins at young ages when

language growth is dynamic and grouping status could change. It reports heritability estimates for each of the clinical groups. The empirical strengths of the study include (a) an unprecedented population-based sample of twins with direct behavioral assessments yielding standardized scores interpretable relative to population-based age level distributions, thereby supporting identification of children not meeting age-level expectations for various measures of speech and language, as well as nonverbal IQ; (b) a sample size that is robust for examination of distributions of children on measures and calculation of heritability estimates for children classified as affected, either SLI or NLI; (c) longitudinal assessments on the same measures at 4 and 6 years of age that reduce possible measurement error over times of measurement, an empirical strength especially important for detection of a twinning effect on language and

Table 7. Comparison of heritability estimates (Est) for the full sample, specific language impairment (SLI), and nonspecific language impairment (NLI) groups at −1.00 *SD* affectedness.

Outcome	Age	Full sample		SLI		NLI	
		Est	*SE*	Est	*SE*	Est	*SE*
PPVT-III Vocabulary	4	.64	.24	.35	.32		
	6	.86	.37	.64	.44		
TOLD-P:3 Spoken Language	4	.36	.19	.34	.26	−.50	.62
	6	.65	.19	.45	.22	.74	.57
TOLD-P:3 Semantics	4	.28	.24	.44	.28	−.42	.63
	6	.52	.25	.60	.27	.86	.53
TOLD-P:3 Syntax	4	.69	.19	.45	.24	.27	.47
	6	.63	.19	.55	.22	.85	.56
GFTA-2 Speech	4	.24	.28	.00	.36		
	6	.52	.50	1.22	.32		
MLU	4	.83	.27	.91	.34	−.75	.65
	6	.74	.29	.74	.32		
TEGI Composite	4	.48	.16	.33	.21	−.06	.50
	6	.71	.20	.71	.21		
TEGI Screener	4	.54	.16	.26	.22	.18	.45
	6	.29	.24	.26	.27		

Note. Negative heritability estimates occurred when the intraclass correlation for dizygotic twins was greater than the intraclass correlation for monozygotic twins, and heritability estimates greater than 1 occurred in models where the intraclass correlation for monozygotic twins was .5 greater than the intraclass correlation for dizygotic twins. Heritability estimates are not reported for combinations in which one of the four cells in the Twin × Affectedness combination was empty within zygosity (for which the associated tetrachoric correlation is not reliable due to small *N*s within a group). PPVT-III = Peabody Picture Vocabulary Test–III; TOLD-P:3 = Test of Language Development–Primary: Third Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; MLU = mean length of utterance as calculated by SALT software; TEGI = Rice–Wexler Test of Early Grammatical Impairment; CMMS = Columbia Mental Maturity Scale.

change over the critical age range of 4–6 years as children enter school; and (d) estimates of nonverbal IQ that are cross-classified with language levels to create SLI and NLI groups for further examination of twinning effects.

The outcomes are discussed according to the research questions, highlighting new information about twinning effects, as well as consistency or inconsistency with previous reports or generalizations in the literature.

RQ 1: SLI and NLI by Language Outcome and Age

Twinning effects on language acquisition are strong in the SLI group as shown by elevated rates of language impairments compared to population estimates at 4 and 6 years of age, as reported in Figures 3 and 4. Using the most generous criterion level of −1 *SD* for SLI, the expected 7%–8% of the children in the SLI group (Norbury et al., 2016; Tomblin et al., 1997) is instead 28% at 4 years of age and 25% at 6 years of age. The twinning effect on language also increases the percentage of children with NLI although at 4 years of age only: 8%, dropping to 3% at 6 years of age, compared to 2%–3% as the expected population estimate reported in previous studies (Norbury et al., 2016; Tomblin et al., 1997). Thus, the twinning effect for language yields a higher percentage of children and is more persistent for children in the SLI group relative to children in the NLI group. Perhaps, this is because the children with NLI might pass through the SLI group on their way to better language. However, recall that, collapsing across all children, there is a twinning effect for language

overall, but not nonverbal IQ (Rice et al., 2018), indicating the full sample is relatively stable in rank within the group on nonverbal IQ across the two times of measurement. In a 2 × 2 sorting of children according to affected/unaffected on language and/or nonverbal IQ measures, children's language levels relative to singleton children change from 4 to 6 years of age. Another way this is evident is in the percentage of children in the group with omnibus language scores in normal range or above, that is, a standard score of 85 or above: At 4 and 6 years of age, 63% and 73% of the children overall are in the normal or above group, respectively.

Another robust feature of the twinning effect is that it replicates across multiple dimensions of language, although this was represented here in Figures 3 and 4 only with the omnibus standard score, a conventional phenotype in twin studies. The outcomes suggest a need for further replication with multiple phenotypes of speech and language, as this is the first report of many of the phenotypes studied here with the same children at the same age levels.

The data do not support the generalization that children with SLI tend to score higher than children with NLI. As shown in Table 1, in the dynamic shift from low to higher performance on language phenotypes from 4 to 6 years of age, the mean scores of children remaining in the designated groupings are quite stable. The generalization from previous studies may not hold in this age range for twins due to the twinning effects that are especially notable in the SLI group; that is, the twinning effect depresses the level of the SLI group to that of the NLI group. The generalization of higher performance in the SLI group

may hold for older twins if the twinning effects resolve after 6 years of age, an empirical question for future studies.

The results in Table 1 replicate the finding in other population-based studies of the existence of children with low nonverbal IQ and language scores in typical or above range, a pattern that replicates across phenotypes with a reduction in percentage from 4 to 6 years of age. This evidence supports the conclusion that low nonverbal IQ is neither necessary nor sufficient for language impairments. This group warrants further study to clarify the children's relative strength in language.

RQ 2: How Frequently Do Twins Meet Language (and Speech) Impairment Criteria at Increasingly Strict, i.e., Lower Levels of Performance, per Phenotype, Age, Group, and Consistency of Group Assignment Over Two Age Levels?

Twinning effects influence the consistency of affectedness status across 4 and 6 years of age, as children are more likely to move into unaffected status on language than nonverbal IQ. This is evident in the classification of unaffected at both times (see Table 2). Highest consistency is evident for nonverbal IQ (84.47% consistently above the −1 SD criterion level across ages, 92.46% at −1.25 SD, and 96.86% at −1.50 SD). Consistency is lower for language assessments (e.g., TOLD Spoken Language, 56.24%; TEGI Composite, 54.45%; consistently above the −1 SD across ages). A comparison of Figures 3 and 4 shows, between 4 and 6 years of age, the twins, as a group, improve in language relative to their age peers, shifting from standard scores below 85 to 90 and above. Yet, the twins still show an elevated rate of SLI at 6 years of age.

The overall importance is that there is great momentum in twin acquisition of language during this age period, with the SLI group gaining on their age peers with an accelerated rate of growth. If they sustain their projected rate of growth, they will exceed their age peers in language acquisition within a few years, which clearly does not happen, so there must be an onset and offset of unusual acceleration. Such unexpected acceleration is unlikely in singleton children with SLI or NLI, as their growth curves over various language phenotypes do not differ from unaffected control children (Rice & Hoffman, 2015; Rice et al., 2009, 2006, 1998, 2000). There is no known explanation for the mechanisms that could drive such a change in acceleration relative to unaffected children, nor is there a clear understanding of how the typical growth curve becomes activated in toddlers or bends with downward inflection points in pre-adolescence. Longitudinal follow-up of the twin sample is needed to document the trajectory of language acquisition through childhood to determine when their apparent language acceleration levels off to align with normative expectations.

RQ 3: Heritability Differences by Affectedness Severity Levels per Phenotype and Age

The rate of proband-wise twin concordance rates for language measures, as expected for genetic effects, was higher for MZ than DZ twins, and with greater severity, the concordance estimates seemed to decrease. In the previous study of the full sample without stratification according to language impairments (Rice et al., 2018), the twinning effect was statistically significantly greater in MZ than DZ twins, which probably affected the concordance estimates based on affectedness status. The dynamic movement across the criterion may have contributed to some discrepancies within twin pairs as children resolved the twinning effect, which could have affected the MZ group more than the DZ group, thereby creating "noise" in the estimates of sameness within MZ pairs. If so, despite this possibility, fairly high levels of concordance were obtained. The patterns of within-pair intercorrelations and concordances for language measures did not replicate for the nonverbal IQ phenotype, which indicated less predictability of twin concordances and led to inconsistent heritability estimates within the age and ability level groupings.

Overall, various speech and language measures replicated previously reported patterns of substantial heritability estimates, although with some differences across phenotypes. Higher heritability with age was consistently replicated. In the age of 4–6 years, this is the first study to reveal how the decreased twinning effect could contribute to increased heritability of language impairment—as the effect is moving toward resolution at 6 years of age, some children shift into normal range, thereby reducing error variance in heritability estimates.

Heritability estimates for low levels of nonverbal IQ did not show the expected patterns of heritability across ages and levels of severity, as evident in many of the language measures. Instead, the outcomes were inconsistent across age and severity levels with some high levels of error variance. Overall, the outcomes suggest high value for examining nonverbal IQ as concurrent measurements in studies of language impairments in children, allowing for more precise information about twinning effects on language (but not nonverbal IQ) and how heritability estimates may differ.

Empirical limitations include some noise in heritability estimates and some noise in phenotype measures. The GFTA's psychometric properties include a ceiling effect at 6 years of age (capturing the natural ceiling as young children master production of speech sounds). The two measures from TEGI also show expected age-level effects related to sensitivity. The TEGI Screener and MLU have less sensitivity to SLI at 6 years of age due to restricted variance in that age range, but the TEGI Composite score picks up sensitivity at 6 years of age due to the greater difficulty of some items in the Composite that are not included in the Screener.

RQ 4: Heritability Rates for SLI Versus NLI per Phenotype and Age

Comparisons are provided for three ways of grouping the participants: the full sample, the SLI-only group, and the NLI-only group. Thus, the group comparisons are confounded with variations in sample size, and the full sample is not independent of either SLI or NLI. Furthermore, some phenotypes are confounded with others: TOLD Spoken Language is a composite of TOLD Semantics and TOLD Syntax, although the Semantics and Syntax subtests are independent phenotypes; TEGI Composite includes the TEGI Screener. Keeping these issues in mind, patterns of heritability across age, phenotypes, and grouping criteria can be compared to previous generalizations in the literature.

The most obvious is that the findings do not replicate the earlier report of heritability for NLI but not SLI in the TEDS sample of twins (Hayiou-Thomas et al., 2005). The nonreplication may be related to technical differences in data analysis. As noted above, the TEDS sample phenotypes were calculated as percentiles within the group of twins, thereby obscuring possible twinning effects on language versus nonverbal cognition. Second, analyses were restricted to twin pairs in which at least one twin was affected, which aligns with the logic of proband-wise concordance calculations, as reported in Table 3. The modeling methods differ, however. In the TEDS, the model looked at the relatedness of cotwins of the proband without considering the full distribution. The modeling method of this study included all participants, modeling heritability by considering unaffected and affected twins. With this approach, there is clearly heritability for SLI, which is greater at 6 years of age and is evident in multiple phenotypes: PPVT h^2 = .64, TOLD Semantics h^2 = .60, TOLD Syntax h^2 = .55, Speech h^2 = 1.00 (truncated), MLU h^2 = .74, and TEGI Composite h^2 = .71. Comparison to NLI is complicated by the lower numbers of children in the NLI group, the negative values nonsupportive of heritability in the NLI group only, and missing values. The missing values also appear only for the NLI group, caused by counts of 0 for affectedness within twin pairs; that is, within a pair of twins, either DZ or MZ, neither twin was affected (see footnotes for Tables 5–7). The only phenotypes available at 6 years of age for the NLI group with positive estimates were from the TOLD phenotypes, which yielded the following h^2 values: Spoken Language, .74; Semantics, .86; and Syntax, .85. Although these heritability estimates are relatively high, interpretation across the phenotypes is complicated by missing models for PPVT and GFTA and four negative models only for the NLI group. Overall, there is less evidence of heritability for the NLI group because the criterion for low nonverbal cognition + low language does not pick up as many children to be counted as affected, or when it picks up affectedness, the similarity within twin pairs, that is, concordances, can be opposite of a heritability effect, that is, with higher within-pair phenotype correlations in DZ twin pairs compared to MZ twin pairs.

Comparison of h^2 estimates across the group phenotypes (full sample, SLI, and NLI) could be informative as to which method to use for sensitivity to heritability, where the advantage could be expected to go to the combined grouping due to the computational benefits of a larger number of children in the affected group. Table 7 outcomes provide at best mixed evidence for this approach, given that the pattern differs by age, affectedness grouping, and phenotype. Lumping everyone together comes at the risk of confounds with nonverbal IQ and/or other unidentified sources of unexplained variance. As shown in Table 5, heritability estimates for the CMMS phenotype are highly variable across ages and levels of severity. It is likely that this variability could introduce unexplained variance in the estimates of h^2 per phenotype within each of the groupings, in turn working against a coherent pattern of outcomes. However, as shown in Figure 1, outcomes from the CMMS are normatively distributed around the expected mean values, with expected means and standard deviations at 4 years of age and again at 6 years of age, suggesting suitable psychometric properties for estimating heritability and stability over time, as shown in Figures 3 and 4 as well. Furthermore, the percentage of NLI children at 6 years of age (3%) is similar to the 5% reported in the Iowa study (Tomblin et al., 1997). Thus, the psychometric properties of the CMMS seem suited to phenotyping, and in the intraclass correlations of Table 4, CMMS scores did not stand out as obviously different from some of the other phenotypes. Further investigation of nonverbal phenotypes and replication studies with CMMS in these age ranges are needed.

Implications for Causal Pathways

The evidence reported here is not consistent with a shared causal pathway for language impairments and low levels of nonverbal cognition. Instead, there are multiple indications of independent pathways for language and nonverbal cognition. First, the twinning effect is evident in phenotypes for language but not in nonverbal cognition. Second, as reported in Table 1 and in Figures 3 and 4, children can have typical or above levels of language scores relative to their age and score low on the nonverbal cognition phenotype. This outcome replicates across phenotypes, across ages, and is not rare. Third, there is no evidence of a "double hit" on language impairments in children who are in the NLI group. Instead, as reported in Table 1, the mean levels of performance for the SLI and NLI groups are very similar across phenotypes. Fourth, consistency in affectedness across both times of measurements is lowest for nonverbal cognition out of all phenotypes, as shown in Table 2. Fifth, as shown in Table 5, the patterns of heritability of nonverbal cognition are more variable across phenotypes and ages and not aligned with patterns of heritability within the same sample of children. Sixth, the patterns of heritability do not align well across the SLI and NLI groups across ages and phenotypes, with more negative values and modeling difficulties for the NLI group. This twin sample provides clear indications of how children's

nonverbal cognition and language acquisition can follow nonoverlapping pathways. The SLI and NLI groups are not equivalent, counter to the arguments for collapsing them into one group for research or clinical practice under the term "DLD."

Comparisons across phenotypes replicate previous findings of heritability across dimensions of speech and language. Support for the heritability of speech impairments is replicated, as is heritability of grammar as well as vocabulary and omnibus measure sampling across different dimensions at different ages. Although theoretical predictions of heritability for grammar were at the forefront of contemporary molecular genetic studies of language acquisition (Fisher et al., 1998), there are now replicated findings that heritability of language impairments in children is not unique to grammar. At the same time, the TEGI phenotype is the only one benchmarked to a technical requirement of the grammar that is not explicitly taught to children, and the phenotype performance can be interpreted as progress toward the obligatory requirements of the adult grammar. This phenotype consistently yields substantial heritability estimates in each previous study in which it has been evaluated (Dale et al., 2018; Rice et al., 2014, 2018), an outcome replicated in this study. The ways in which inherited influences affect language acquisition remain to be identified.

Clinical Relevance

It is often noted that the entry point for best practice is accurate diagnosis. This study adds new insights into diagnostic criteria for SLI and NLI. Note that collapsing across the two categories to form the new combination label of DLD would increase unexamined heterogeneity, as noted earlier (Stark & Tallal, 1981), which in turn would obscure the important differences revealed in this study about the details in how language changes between 4 and 6 years of age in twins and the replicated finding that language and nonverbal cognition are not on the same causal pathway. A twin with language impairments cannot be assumed to have low nonverbal cognition, nor can it be assumed that the language delays of twins are due to low nonverbal cognition. Twins do not align with the normative data for singleton children in early language acquisition, although they are likely to meet normative levels for nonverbal cognition.

The results of this study also suggest caution about assuming there are quantitative differences between children with SLI and NLI, such that children with NLI are likely to score at lower levels than the children with SLI across various measures, whereas there are no qualitative differences in symptoms of language impairments across the two groups (Bishop et al., 2017). In the case of twins, there seems to be a quantitative difference in early language acquisition demonstrated as a twinning effect that looks like late language acquisition, which in turn introduces qualitative differences in developmental trajectories.

In the age range of 4–6 years (as well as at 24 months; Rice et al., 2014), the twinning effect calls for caution in arriving at a clinical diagnosis of language impairments in twin children ages 4 and 6 years. When assessing a young twin who seems to have "immature" language, a default interpretation of limited maternal input or limited nonverbal cognitive abilities is not warranted, and instead, a twinning effect on language should be considered. Furthermore, a diagnosis of SLI or NLI at 4 years of age may overrepresent true cases of SLI or NLI as twins work through the twinning effect. On the other hand, the findings here suggest that some of the twin children may have a true persistently elevated risk of SLI, at least through 4 and 6 years of age, given the similarity of proportions of children in the SLI group across the ages and the relatively high prevalence. Parental concern is a driving factor in accessing clinical services for children (Skeat et al., 2010). For parents of twins, regular opportunities to raise concerns with service providers in the health, child care, and education sectors are likely to play an important role in the identification and treatment of SLI and NLI in twins. Practitioners familiar with the research on twins can consider a monitoring approach in combination with parental counseling in the 2- to 6-year age range. Standard scores based on singleton norms must be interpreted cautiously, and repeated measurements must be provided during the preschool years.

Another approach could be to enroll referred children in preventive services to facilitate closure of a possible twinning gap, that is, enhancing language acquisition to a rate to catch up with age expectations. This approach would also provide early identification of children with SLI or NLI for ongoing intervention services. The finding that variability in performance over time characterizes typical and atypical language development in twins is also observed in singleton children (Christensen et al., 2014; Rice et al., 2008). A question for future research is whether the persistence of atypical language performance over time could be a better predictor of SLI or NLI than language status at a given point in time.

Overall, the outcomes of the study provide new perspectives about how to interpret inherited effects on early speech and language development. The sizes of the heritability effects are substantial, are replicated across different speech and language phenotypes, and add strength to the generalization that heritability increases with age. The most generous criterion of affectedness for defining SLI or NLI is a language level -1 SD below the age norm, usually a standard score of 85 or lower. The findings here document that this level revealed consistent heritability across different phenotypes. In other words, inheritance is not limited to "severe" language disorders but also plays a role in language disorders in the range of the bottom 15th percentile of children. The new suggestion is that the inherited mechanisms for speech and language are robust despite, in the special case of twins, a twinning effect on timing of acquisition early in childhood. The twinning effect may present "noise" for the estimation of inherited contributions to speech and language impairment, relatively independent

of nonverbal IQ. If so, this could contribute to the higher heritability of the phenotypes at 6 years of age. Another possibility could be that, in the causal pathways, mechanisms driving the heritability of speech and language impairments are intertwined with mechanisms contributing to a twinning effect. Overall, the outcomes indicate the high informativeness of language studies of twin children.

Acknowledgments

References

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Blickstein, I., & Keith, L. G. (2005). *Multipe pregnancy: Epidemiology, gestation and perinatal outcome* (2nd ed.). CRC Press. https://doi.org/10.1201/b14615

Bloomfield, F. H. (2011). Epigenetic modifications may play a role in the developmental consequences of early life events. *Journal of Neurodevelopmental Disorders, 3*(4), 348–355. https://doi.org/10.1007/s11689-011-9096-1

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

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142–1157. https://doi.org/10.1044/1092-4388(2002/093)

Christensen, D., Zubrick, S. R., Lawrence, D., Mitrou, F., & Taylor, C. L. (2014). Risk factors for low receptive vocabulary abilities in the preschool and early school years in the longitudinal study of Australian children. *PLOS ONE, 9*(7), e101476. https://doi.org/10.1371/journal.pone.0101476

Dale, P. S., Dionne, G., Eley, T. C., & Plomin, R. (2000). Lexical and grammatical development: A behavioural genetic perspective. *Journal of Child Language, 27*(3), 619–642. https://doi.org/10.1017/S0305000900004281

Dale, P. S., Rice, M. L., Rimfeld, K., & Hayiou-Thomas, M. E. (2018). Grammar clinical marker yields substantial heritability

for language impairments in 16-year-old twins. *Journal of Speech, Language, and Hearing Research, 61*(1), 66–78. https://doi.org/10.1044/2017_JSLHR-L-16-0364

DeThorne, L. S., Hart, S. A., Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., & Davison, M. D. (2006). Children's history of speech-language difficulties: Genetic influences and associations with reading-related measures. *Journal of Speech, Language, and Hearing Research, 49*(6), 1280–1293. https://doi.org/10.1044/1092-4388(2006/092)

Dunn, L. M., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test–III (PPVT-III).* AGS. https://doi.org/10.1037/t15145-000

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics, 18,* 168–170. https://doi.org/10.1038/ng0298-168

Gavin, W. J., & Giles, L. (1996). Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research, 39*(6), 1258–1262. https://doi.org/10.1044/jshr.3906.1258

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2).* AGS. https://doi.org/10.1037/t15098-000

Hay, D. A., Prior, M., Collett, S., & Williams, M. (1987). Speech and language development in preschool twins. *Acta Geneticae Medicae et Gemellologiae: Twin Research, 36*(2), 213–223. https://doi.org/10.1017/S000156600000444X

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233–249. https://doi.org/10.1111/j.1467-7687.2011.01119.x

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2014). Language impairment from 4 to 12 years: Prediction and etiology. *Journal of Speech, Language, and Hearing Research, 57*(3), 850–864. https://doi.org/10.1044/2013_JSLHR-L-12-0240

Hayiou-Thomas, M. E., Kovas, Y., Harlaar, N., Plomin, R., Bishop, D. V. M., & Dale, P. S. (2006). Common aetiology for diverse language skills in 4 1/2-year-old twins. *Journal of Child Language, 33*(2), 339–368. https://doi.org/10.1017/S0305000906007331

Hayiou-Thomas, M. E., Oliver, B., & Plomin, R. (2005). Genetic influences on specific versus nonspecific language impairment in 4-year-old twins. *Journal of Learning Disabilities, 38*(3), 222–232. https://doi.org/10.1177/00222194050380030401

Martin, J. A., Hamilton, B. E., & Osterman, M. J. K. (2012). *Three decades of twin births in the United States, 1980–2009.* NCHS Data Brief No. 80. National Center for Health Statistics.

Miller, J., & Chapman, R. (2002). SALT for Windows. Standard Version 7.0 [Computer software]. Madison, WI: Language Analysis Laboratory, Waisman Research Center.

Muthén, L. K., & Muthén, B. O. (1998–2017). Mplus user's guide (8th ed.). Muthén & Muthén.

Newcomer, P., & Hammill, D. (1997). *Test of Language Development–Primary: Third Edition (TOLD-P:3).* Pro-Ed.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Oliver, B. R., & Plomin, R. (2007). Twins' Early Development Study (TEDS): A multivriate, longitudinal genetic investigation of language, cognition and behavior problems from childhood

through adolescence. *Twin Research and Human Genetics, 10*(1), 96–105. https://doi.org/10.1375/twin.10.1.96

Olson, R. K., Keenan, J. M., Byrne, B., Samuelsson, S., Coventry, W. L., Corley, R., Wadsworth, S. J., Willcutt, E. G., DeFries, J. C., Pennington, B. F., & Hulslander, J. (2011). Genetic and environmental influences on vocabulary and reading development. *Scientific Studies of Reading, 15*(1), 26–46. https://doi.org/10.1080/10888438.2011.536128

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie, P., & Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from Early Language in Victoria Study. *Pediatrics, 126*(6), 1530–1537. https://doi.org/10.1542/peds.2010-0254

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,* Article no. 27. https://doi.org/10.1186/1866-1955-4-27

Rice, M. L. (2020). Clinical lessons from studies of children with specific language impairment. *Perspectives of the ASHA Special Interest Groups, 5*(1), 12–29. https://doi.org/10.1044/2019_PERSP-19-00011

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment (SLI): A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58*(2), 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150

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*(6), 1417–1433. https://doi.org/10.1044/1092-4388(2009/08-0171)

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research, 49*(4), 793–808. https://doi.org/10.1044/1092-4388(2006/056)

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*(2), 333–349. https://doi.org/10.1044/1092-4388(2009/08-0183)

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research, 51*(2), 394–407. https://doi.org/10.1044/1092-4388(2008/029)

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*(4), 816–834. https://doi.org/10.1044/1092-4388(2004/061)

Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment.* 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*(6), 1412–1431. https://doi.org/10.1044/jslhr.4106.1412

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*(5), 1126–1145. https://doi.org/10.1044/jslhr.4305.1126

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*(4), 943–961. https://doi.org/10.1044/jslhr.4204.943

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. https://doi.org/10.1044/1092-4388(2013/12-0350)

Rice, M. L., Zubrick, S. R., Taylor, C. L., Hoffman, L., & Gayán, J. (2018). Longitudinal study of language and speech of twins at 4 and 6 years: Twinning effects decrease, zygosity effects disappear, and heritability increases. *Journal of Speech, Language, and Hearing Research, 61*(1), 79–93. https://doi.org/10.1044/2017_JSLHR-L-16-0366

Rudolph, J. M., Dollaghan, C. A., & Crotteau, S. (2019). The finite verb morphology composite: Values from a community sample. *Journal of Speech, Language, and Hearing Research, 62*(6), 1813–1822. https://doi.org/10.1044/2019_JSLHR-L-18-0437

Rutter, M., Thorpe, K., Greenwood, R., Northstone, K., & Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay: I: Design; twin-singleton differences in language, and obstetric risks. *The Journal of Child Psychology and Psychiatry, 44*(3), 326–341. https://doi.org/10.1111/1469-7610.00125

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research, 42*(6), 1461–1481. https://doi.org/10.1044/jslhr.4206.1461

Skeat, J., Eadie, P., Ukoumunne, O., & Reilly, S. (2010). Predictors of parents seeking help or advice about children's communication development in the early years. *Child: Care, Health and Development, 36*(6), 878–887. https://doi.org/10.1111/j.1365-2214.2010.01093.x

Smith, C. (1974). Concordance in twins: Methods and interpretation. *American Journal of Human Genetics, 26*(4), 454–466.

Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 46*(2), 114–122. https://doi.org/10.1044/jshd.4602.114

Taylor, C. L., Rice, M. L., Christensen, D., Blair, E., & Zubrick, S. R. (2018). Prenatal and perinatal risks for late language emergence in a population-level sample of twins at age 2. *BMC Pediatrics, 18,* Article no. 41. https://doi.org/10.1186/s12887-018-1035-9

Thorpe, K. (2006). Twin children's language development. *Early Human Development, 82*(6), 387–395. https://doi.org/10.1016/j.earlhumdev.2006.03.012

Thorpe, K., Rutter, M., & Greenwood, R. (2003). Twins as a natural experiment to study the causes of mild language delay: II: Family interaction risk factors. *The Journal of Child Psychology and Psychiatry, 44*(3), 342–355. https://doi.org/10.1111/1469-7610.00126

Tomblin, J. B., & Nippold, M. A. (2014). *Understanding individual differences in language development across the school years.* Psychology Press. https://doi.org/10.4324/9781315796987

Tomblin, J. B., Records, N. L., Buckwalter, 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*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245