Causal Pathways for Specific Language Impairment: Lessons From Studies of Twins

Mabel L. Rice

Purpose: This review article summarizes a program of longitudinal investigation of twins’ language acquisition with a focus on causal pathways for specific language impairment (SLI) and nonspecific language impairment in children at 4 and 6 years with known history at 2 years.

Method: The context of the overview is established by legacy scientific papers in genetics, language, and SLI. Five recent studies of twins are summarized, from 2 to 16 years of age, with a longitudinal perspective of heritability over multiple speech, language, and cognitive phenotypes.

Results: Replicated moderate-to-high heritability is reported across ages, phenotypes, full population estimates, and estimates for clinical groups. Key outcomes are documentation of a twinning effect of risk for late language acquisition in twins that persists through 6 years of age, greater for monozygotic than dizygotic twins (although zygosity effects disappear at 6 years); heritability is greater for grammar and morphosyntax than other linguistic dimensions, from age 2 years through age 16 years, replicated within twin samples at subsequent age levels and across twin samples at age 16 years.

Conclusion: There is consistent support for legacy models of genetic influences on language acquisition, updated with a more precise growth signaling disruption model supported by twin data, as well as singleton data of children with SLI and nonspecific language impairment.

Presentation Video: https://doi.org/10.23641/asha.13063727

Specific language impairment (SLI) is defined as “a communication disorder that interferes with the development of language skills in children who have no hearing loss or intellectual disabilities…one of the most common developmental disorders, affecting approximately 7 to 8 percent of children in kindergarten. The impact of SLI usually persists into adulthood” (https://www.nidcd.nih.gov/health/specific-language-impairment). Although causes of SLI are unknown, there has been significant progress (reported in an extensive research literature) documenting the linguistic and communicative symptoms of SLI, long-term outcomes, and related cognitive and linguistic processing abilities, among other topics (Leonard, 2014; Norbury et al., 2016; Tomblin & Nippold, 2014). As a more complete picture emerges of the clinical presentation of SLI, the core question remains: What causes difficulty in language acquisition in children who appear to have the basic prerequisite abilities that allow other children to acquire a native language with little apparent effort and little explicit tutorial instruction?

Almost all the extensive research literature on SLI is based on samples of singleton children, in keeping with research conventions that want to avoid possible complications due to conditions of multiple births, which could cause unexplained error variance in experimental studies. However, twin studies employ powerful methods for differentiating causal pathways for language abilities, which have enriched our understanding of the sources of language abilities in singleton and twin children. In particular, the study of twins generates estimates of inherited (genetic) effects versus shared environmental effects (such as the family setting), given the uniquely informative aspects of twin biology. The purpose of this review article is to summarize the outcomes of a recent program of research designed to bridge the gap, from what is known about singleton children with SLI to what is known about late language emergence (LLE) and possible SLI in twins. Following the methods of population-based studies of twin children, and previous reports of heritability of language acquisition, the program of investigation summarized here focused on SLI, with appropriate assessments and documentation of ways in which twin children, and twin
children with SLI, may differ from singleton children. The outcomes identify new insights about the causal pathways that support language acquisition in singleton children and in twins, and the complexities of nature versus nurture.

This review article proceeds in the following order of topics. First, there is a brief overview of the theoretical underpinnings of the core question of the source of humans’ language ability, which appears in the first 2 years of life, across many different societies and languages, without need of explicit instruction. In this context, children with SLI are of interest because they seem to have the wherewithal needed for language to develop but, instead, acquisition of language is delayed and, once underway, is likely to remain lower than expected relative to age peers, a gap that can continue into adulthood. This overview of theoretical papers covers the period relevant to early contemporary research on SLI; more immediate precedents for the research on twins (i.e., empirical studies of the heritability/environmenal effects) can be found within the literature reviews of the cited papers. Second, a concise summary of the legacy of contemporary research about SLI is provided. Third, an overview of the logic and methods of twin studies is provided as a framework for interpretation of outcomes, followed by the fourth section, a summary of the goals, methods, and outcomes of four recent empirical studies of twins that establish important new information about possible heritability of SLI. New findings are highlighted, including complexities in comparing twin outcomes to those of singleton children, evidence of differences between nonverbal cognitive abilities compared to language abilities, and high heritability for various measures of speech and language and for the condition of SLI. The fifth and final section is a discussion of the implications of the outcomes for further research and clinical issues.

Origins of Human Language: Inherited Versus Environmental Influences

The modern study of the biological foundations of child language was launched in a controversial book by the cognitive psychologist, Eric Lenneberg (Lenneberg, 1967). Prior to the modern period, from the earliest historical writings, there was fascination with the question of how children acquire language. A classic experimental approach was to deprive infants of language input, assumed to be a method for discovering the fundamental character of human nature including the origin of language, inspired by the apparent belief that children left to themselves will develop language. Of course, we now know this is a deeply flawed (and certainly unethical) method of inquiry. It is, however, indicative of a long-standing, if yet not fully resolved, philosophical and empirical question. See a concise historical summary by Otto Marx (Marx, 1967).

In the modern era of science, Lenneberg (1967) proposed the thesis that biological considerations are necessary. He stated a need

…to conduct investigations with the nature, structure, and history of natural languages, and then to see whether the empirically determined principles underlying language are, indeed, represented in the experimental paradigms. From this consideration it follows that a biological investigation of language must not only study the organism that speaks but must also investigate the behavior itself—language—much the way the zoologist who studies the badger must study its physique together with its habits in order to give a complete picture of that animal…biological aspect of man and some of the biological aspects of language. (Lenneberg, 1967, p. 3)

These directions of research have been the focus of extensive research on child language, enhanced by the invention of portable digital audio-recording devices and systematic data collection of spontaneous language samples collected from young children, beginning with the groundbreaking work of Roger Brown (Brown, 1973).

Lenneberg anticipated key outcomes of our contemporary language acquisition research in his list of observable facts consistent with the notion of innate drivers for language acquisition, such as:

The regularity of language-onset as a milestone that fits into an order and fairly constant sequence of other maturational milestones is…an observable fact and so is the apparent similarity in language acquisition strategies, the universal similarity of primitive states, and the difference in outer form between primitive stages and adult language. (Lenneberg, 1967, p. 379)

Subsequent decades of empirical research across languages have confirmed these general characterizations (with the expected dissent or clarification about details). In addition, Lenneberg recognized the unique experimental value of comparing fraternal and identical twins, based on the limited empirical evidence available at that time. He notes:

All studies agree that fraternal twins are much more prone to differences in language development than identical twins…The developmental history of identical twins tends to be much more synchronous than that of fraternal twins…including onset of speech…It is the general consensus of the investigators that these divergences cannot be simply explained on the grounds of imitation or differential treatment by parents…If the developmental history as a whole is compared, that is the occurrence of milestones and/or the occurrence and nature of speech and language abnormalities, the difference between identical and fraternal twins is even greater. (Lenneberg, 1967, p. 253)

At the time of Lenneberg’s writings, knowledge of the condition of SLI was very limited and not well known. The array of empirical methods for studying twins and genetic influences was equally limited. The modern era of theoretical formulations of the structure of universal language arguably began with the work of Chomsky (1957, 1967). This review article updates these threads of
science established in 1967 by focusing on children with SLI and addresses how twin studies not only reveal relative contributions of heredity and environmental influences in all children but also help clarify possible causal mechanisms for the symptoms of speech and language impairment (across dimensions of language) in children with SLI, as well as clarification of the role of nonverbal cognitive abilities.

Legacy Research in Children With SLI

Modern legacy research in children with SLI appeared first in the work of generative linguists inspired by Chomsky’s theory of universal grammar. The theory provided a framework for investigating whether these children’s grammar deviated from theoretical expectations. Paula Menyuk led the way with an awareness that it would be interesting and informative to look carefully at the grammars of children with language impairments who did not have other developmental disorders, even before the diagnostic label of SLI was established (Menyuk, 1964). Her paper was followed by an important paper with a provocative title, “What is deviant language?” (Leonard, 1972), which inspired other investigators to follow the trail of possible questions to ask.

Early on, investigators recognized that precise definitions of (SLI) affected children, out of all children with language disorders, would be necessary to avoid problems of replication related to noncomparability of participants across studies. This issue was tackled explicitly in a paper published in 1981 (Stark & Tallal, 1981), whose authors recognized the essential value of precise criteria for participant definition. They defined children with “Specific Language Deficits” as a “...developmental disorder characterized by severe problems of comprehension and/or expression of spoken language in the absence of hearing loss, mental retardation, or emotional disorder” (p. 114), previously referred to as “childhood dysphasia” with unknown etiology (Benton, 1964). The goal of their study aligns well with the many studies of children with SLI that followed their paper. Furthermore, they anticipated the current National Institutes of Health focus on replicability across studies:

- The development and testing of treatment procedures appropriate for children with different types of language disorders can only be carried out if sub-groups can be identified reliably by clinicians and/or investigators working in different types of settings...The selection process is likely to influence the results obtained from any study, the generalization of conclusions to other groups of language-impaired children, and the suggested implications for treatment. (Stark & Tallal, 1981, p. 114)

They followed with a further statement late in the paper: “...the need for care in selecting children with specific language impairment for both research and clinical intervention studies. Otherwise the conclusions drawn from such studies may be generalized incorrectly to all language-impaired children” (Stark & Tallal, 1981, p. 119).

Their point, almost 40 years old at the time of this writing, is as relevant today as ever, given current attempts to override the criteria for SLI with more inclusive definitions of research groupings (Rice, 2020). By the 1990s, the term “deficit” was replaced by “impairment” as a more appropriate term, and more precise measurement criteria were developed for identification of children with SLI (Leonard, 2014; Volkers, 2018). The issue of precision of classification criteria is also highly pertinent to the topic of this review article, because precise definition of participant characteristics is a well-known asset for the methods of twin studies.

Relevant to the summary of twin research to follow are two widely accepted generalizations from recent programs of research on children with SLI, noted in the current National Institute on Deafness and Other Communication Disorders definition of SLI (https://www.nidcd.nih.gov/health/specific-language-impairment). One conclusion is that, although SLI is a developmental condition that appears early in children, it is likely to persist into adulthood. Similar to their age peers, the language of children with SLI changes over time. In fact, children with SLI follow similar growth trajectories to their age peers across different dimensions of language early in childhood, followed by a leveling of rate of growth in pre-adolescence as do their age peers, leaving the children with SLI with a lower-than-expected level of language as they transition into adulthood (Rice, 2012, 2019, 2020; Rice & Hoffman, 2015). A second conclusion is that certain properties of grammar (best documented for finiteness marking in morphosyntax in languages such as English and many other languages, but certainly not all) are reliable clinical markers of SLI, meeting standards of sensitivity and specificity for identification of affected children, especially well documented for the age range of 3–9 years (Redmond et al., 2019; Rice, 2019, 2020; Rice & Wexler, 2001).

Logic and Methods of Twin Studies

The biology of twin conception provides a logic and method for differentiating the impact of inheritance versus environmental causal pathways on behavioral traits such as speech, language, and cognition. Twin studies adopt the terminology of genetics investigations, in which behaviors of interest are referred to as “phenotypes” defined at the biological level as observable properties of an organism (Strachan & Read, 2011), or, at the behavioral level, as the set of observable characteristics of an individual resulting from the interaction of its genotype with the environment and random variation. The usage here follows conventions of behavioral twin studies, where “phenotype” refers to measurable behaviors that are thought to tap into underlying dimensions of complex cognitive traits of “language,” “speech,” and “nonverbal cognition,” based on previous research evaluating theoretical constructs supported by reliable and valid measurements. In the overview to follow, the term “phenotype” is working at two levels of interpretation. Conceptually, it focuses on behaviors hypothesized to have possible genetic influence. This would not include
socially established habits such as table manners or noninherited outcomes such as brain injury. It also does not assume that a behavioral phenotype captures the full scope of a conceptual construct such as “human language,” “universal grammar,” or “intelligence.” Instead, an open question is whether genetic analytic methods can help inform the extent to which discrete speech, language, or cognitive dimensions, for example, share common causal pathways as measured in twin studies. At the measurement level, in the following discussion, “phenotype” also can be interpreted as a variable in a particular study, a common usage in scientific reports.

The logic of twin studies is based on the biological status of twins. Twin pairs are subdivided into two types: Identical (monozygotic [MZ]) twin pairs share a fertilized egg at conception, whereas fraternal (dizygotic [DZ]) twins do not share a fertilized egg at conception. Thus, both kinds of twin pairs share the time in uterus during pregnancy and are usually born at the same time, but, unlike the MZ pairs, the two children in a DZ pair are no more genetically similar than their other full siblings. MZ and DZ twins, reared by their parents, usually share the same home environment.

With the advent of assistive reproductive technology worldwide and older mothers, 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). Twins are at 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).

Given their biological similarities and differences, the study of twins is the most widely used methodology for evaluating the role of genetics and environment in language and cognition. Mathematical models are used to estimate heritability and environmental effects, based on the assumptions that, roughly speaking, MZ twins share 100% of their genome and DZ twins share, on average, 50% of their segregating genome. A useful (although simplistic) model assumes that MZ twins share genes and an environment (therefore are expected to be similar), whereas DZ twins share an environment but not genes (therefore may differ). It follows that, if MZ twins are more alike than DZ twins, the difference can be attributed to genetic influences. It is helpful to keep in mind that the core of the computational methods is the correlation between the two children within twin pairs (MZ vs. DZ) on any given phenotype. As shown in Figure 1, for inherited traits, it is expected that the correlation within twin pairs will be higher for MZ pairs (shared genomes and environments) than for DZ pairs (different genomes and shared environments).

In anticipation of content to follow, it is also important to keep in mind that correlational data cannot tell us how well children perform on the task. Correlations measure the extent to which scores from each member of a twin pair are similar. This association would hold across a full range of scores. The association does not reveal the level of performance of the group relative to age norms. These generalizations are illustrated in idealized (not actual) outcomes in a pair of figures in Figure 2, contrasting correlations with levels of performance. The figure on the left depicts hypothetical correlations between two twin children within a twin pair on vocabulary standard scores. A correlation (r) of 1.00 is a perfect association, such that, in a group of twins, as one member of the pair’s vocabulary score increases, so does that of the other member of the pair. Less perfect associations are indicated as dashed lines for r = .80 or r = .30, showing that gains in vocabulary standard score may not align well such that one twin within the pair is different from the other. The figure on the right in Figure 2 illustrates how a hypothetically similar level of r = .80, as in the left figure, may mask a dissimilarity of levels of accuracy, such that the association between vocabulary scores for twins within the same twin pair is highly likely to be similar across two groups of twin pairs, yet the accuracy of performance on the two measures could consistently differ between the two groups. Thus, MZ twins could have high correlations within the twin pair and yet also score low on assessments relative to age norms.

**Twinning Effects on Language Performance**

Previous studies documented a “twinning” effect on language acquisition such that twin children appeared to be later than singleton children in reaching early language benchmarks. One study compared the language development of 20-month-old twins and single-born children, reporting a vocabulary delay in twins of about 1.7 months at 20 months (Rutter et al., 2003). This study had limited metrics; the reported age-equivalent scores are not comparable to standard scores, and the scores at 20 months are not interpretable in terms of a child’s relative status in the full group. Another study (Hay et al., 1987) reported an increased likelihood of speech and language impairment in preschool twins, again reported as age-equivalent scores. The gap in evidence from these two studies is the need for standardized scores that allow for estimates of how twin children perform relative to their singleton age peers and the stability of their performance over time. This gap was not filled by the more recent large Twins Early Development Study (TEDS), a longitudinal study of twins born in England and Wales in 1994, 1995, and 1996 (Oliver & Plomin, 2007). Given a large
sample size and limited time for assessing all dimensions of interest, they used truncated/adjusted versions of language tests instead of the more time-consuming fully standardized versions. To interpret the data, they calculated language acquisition metrics referenced within the sample of twins, such that, for example, a child with a score at the 15th percentile was at that rank within the twin group (Kovas et al., 2005). This left unexamined the comparison of the level of performance at the 15th percentile within the twin group to same-age singleton children, a comparison that would be possible if using standardized assessments, which are primarily normed on singleton children.

Limitations of Legacy Twin Studies

Studies of twin children are relatively rare in the language acquisition literature because many challenges are involved; for example, (a) large sample sizes are required for the quantitative methods. (b) Population-based samples are preferred, to avoid overrepresentation of MZ pairs (in nature, for humans, about one third of twin pairs are MZ pairs). (c) It can be difficult to identify twins, especially in the United States, under the distributed medical care system. (d) Language phenotypes are often limited and lacking in metrics interpretable for comparisons to singleton age peers, in part because the requisite standardized language assessments require time-consuming administration by trained personnel. The consequence is that the resulting information is limited in its interpretation and lacks necessary details of various linguistic dimensions. (e) Also lacking, due to the practice of previous studies using different assessments with different psychometric properties at different ages, is documentation of growth in language dimensions that captures individual change over time, in ways that can be directly compared to singleton children. (f) As a result, little is known about possible twinning effects on language, although such effects need to be clarified in order to avoid confusing twin children with true language impairments versus late language acquisition attributable to twinning as a unique phenomenon.

General Goals of the Program of Twin Investigation Reported Here

The general intent of the program of twin study summarized here was to design and carry out a relatively large-scale longitudinal study of a population-based sample of twins, beginning at 2 years of age, with multiple dimensions of speech and language assessments that yield standardized scores referenced to same-age singleton children. The scope and level of assessment detail was tied to the need for information to clarify causes of SLI in children and to unpack the phenomenon of possible twinning effects as they might bear on the identification of SLI in young twin children. Longitudinal growth trajectories are the ultimate long-term goal. With funding from the National Institutes of Health (R01DC005226), data collection was centered in Perth, Australia, a city with favorable demographic similarity to the Kansas City metropolitan area (which straddles the states of Missouri and Kansas in the United States). Highlights of the outcomes reported so far are summarized in the following sections, in chronological order from 2 to 6 years of age, with the addition of a recent TEDS that included relevant linguistic phenotype outcomes.

Participants

Participants for the first three studies summarized below are from the longitudinal twin sample. The sample was recruited from public health records representative of the general population. Preliminary analyses verified that the sample was representative of the general population in family- and child-level assessments. As expected, about one third of the twin sample were MZ twins.
Relevant exclusionary criteria are described in an earlier report (Rice et al., 2018) and summarized here. 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 developmental 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.”

**Design**

The study recruited a sample of twins over a 3-year window to compile a full sample for longitudinal assessment. The first time of measurement was via parent report questionnaire data collection when children were 2 years old, followed by direct assessments of the children at 4 and 6 years of age.

**Measures**

A wide array of information was obtained from parental questionnaires and direct assessment of the children. The ones of interest here are (a) family demographic variables used as covariates in modeling outcomes (e.g., maternal education and family socioeconomic status) and (b) age-appropriate standardized assessments of speech, language, and cognitive outcomes, with established levels of validity and reliability for age-referenced scores. Direct assessments were carried out in appropriate rooms at a research building or, for families unable to travel, in mobile vans customized for the purpose of off-site data collection. Details of measurement can be found in previous reports (Rice et al., 2020, 2014, 2018).

**Twin Outcomes at 2 Years of Age: Identification of a Twinning Effect**

Earlier studies documented that twins sometimes showed a delayed onset of language acquisition, referred to as a “twinning effect” (Hay et al., 1987; Rutter et al., 2003; Thorpe et al., 2003). In effect, twin children seemed to score lower than expected for their age in language assessments. However, the methods of assessment in early twin studies of toddler-age children were limited because assessments did not meet the expected psychometric standards for sensitivity to language delays relative to same-age peers. Additionally, because variance in performance at 24 months can be considerable, there is a high degree of error, which makes it difficult to reliably estimate the standard deviation around the mean level of performance. Age-equivalent scores are unreliable due to a lack of adjustment for variance within an age level (Rutter et al., 2003). The TEDS method of using a study-internal norming system that did not provide a reference to population norms made it impossible to evaluate twin performance compared to same-age singleton peers.

Here, I briefly summarize a twin study of 473 twin pairs (N = 964 children), the first in a longitudinal investigation of twins, beginning with children aged 24 months (Rice et al., 2014). Data collection utilized the full version of the well-documented parent questionnaire for early language assessment, the MacArthur Communicative Development Inventories: Words and Sentences (Fenson et al., 1993), which provides age-referenced norms broken out for boys versus girls. The dependent variables of interest here are the number of words produced, percentile of words produced for age, the proportion of combining words, and a new measure of early finiteness marking in grammar, developed from the item pool of sample utterances reported by mothers to be like those produced by their children. These items were selected because they indicate knowledge of required marking of finiteness in English morphology, for a variable called “Mini-Finite.” Details are available in the full report (Rice et al., 2014). Predictor variables of maternal and child status at birth as well as demographic variables (race, maternal education) were used in modeling outcomes and estimating heritability.

Two major outcomes are of interest here, described in detail in the full report (Rice et al., 2014). First was the precise documentation of twinning effects such that twins have lower average language scores than age norms for single-born children, a gap more pronounced for MZ than DZ twins and more for boys than girls. This is a robust effect, such that the mean percentile for words produced is about the 23rd percentile for MZ twins, or roughly about two thirds of a standard deviation below the expected 50th percentile for single-born children. The conclusion is that LLE, or a twinning effect, is a notable risk for twins at 2 years. Heritability models were calculated based on correlational analyses; the covariates were nonsignificant or had only mild effects and are not discussed here. Statistically significant heritability was found for vocabulary (.26) and grammar phenotypes (.52 for boys and .43 for girls). LLE and the appearance of word combinations were also significantly heritable (.22–.23). These outcomes can be compared to Figure 1 and Figure 2. The significant heritability suggests that correlations within MZ twin pairs are higher than correlations within DZ twin pairs (even after adjusting for the covariates), a pattern shown in Figure 1. The size of these correlations, for example, the .26 for vocabulary, is a little below the red dotted line of .30 shown on the left side of Figure 2. The heritability estimates were consistent with previous research for vocabulary in this age group (Dale et al., 2000); this was the first report of heritability of grammar in this age group. Note that the statistically significant heritability outcomes do not reveal the considerably large twinning effect, which requires age-referenced measurements.

The outcomes raised new questions, some more empirical in nature and some more interpretive. On the empirical
side, we can wonder how long the twinning phenomenon persists, and how much the twinning effect may contribute to the significant heritability of performance on early language measures in twins as they age. On the interpretive side, we can wonder what causes the twinning effect. Earlier studies hypothesized that the cause could be attributed to the demands of two babies of the same age on maternal attention and quality of verbal interactions (Thorpe, 2006). However, the greater twinning effect for MZ than DZ twins, across multiple measures, works against this hypothesis, given that both types of twin pairs would pose problems for maternal time-sharing or vocal interactions. Another possibility is that perhaps there were environmental effects (such as maternal education or socioeconomic disadvantages) that account for the twinning effect. A follow-up study (Taylor et al., 2018) found no association with LLE in the 24-month-old twin sample for sociodemographic risk factors, such as low maternal education and socioeconomic area disadvantage, similar to outcomes for a large sample of singleton children at 24 months (Zubrick et al., 2007).

We also evaluated whether prenatal or perinatal effects differentiated twins versus singletons, which, if found, could contribute to the twinning effects. Yet prenatal and perinatal risk factors paralleled risk factors identified in singleton births (i.e., gestational diabetes, prolonged time to spontaneous respiration, and fetal growth restriction). The only risk factor unique to twin pregnancies was monozygosity (Taylor et al., 2018). Thus, the prenatal, perinatal, and sociodemographic predictive relationships for LLE are not notably different in twin versus singleton births, and there is consistent evidence suggesting differences between MZ and DZ twin pairs.

For the purpose of studies of SLI, the twinning effects raised the empirical question of how we could disentangle effects of twinning on delayed early language acquisition versus LLE as seen in singleton children as a possible precursor of SLI (Rice et al., 2008; Zubrick et al., 2007). On the interpretive side, the twinning effects could complicate our understanding of causal pathways for twin language acquisition versus the classic SLI of singleton children. The opportunity to follow the twin children as they aged provided valuable empirical and interpretive information.

Outcomes for the Full Group of Twins at 4 and 6 Years: Reduction of the Twinning Effect and Increased Heritability

The longitudinal sample of twins was assessed for language acquisition and covariates again at 4 and 6 years of age, making it possible to document the possible persistence of twinning effects into the school entry age range as well as possible increases in heritability of speech and language outcomes. A consistent finding from previous twin studies (reviewed above) is that heritability for language increases with age, although this effect has not been well documented for speech, language, and nonverbal cognition phenotypes within this age range. Furthermore, there was no documentation available about whether a twinning effect shown at 2 years could be resolving during this age period, such that some of the twins accelerate their language acquisition to catch up with age peers. If such a change is underway, it would create a dynamic shift in individual children’s age-referenced standard scores during this time, which could influence the stability of children in the low end of performance.

The general purpose of this study (Rice et al., 2018) was to investigate the heritability of language, speech, and nonverbal cognitive development of twins at 4 and 6 years. Possible confounding effects of twinning and zygosity (evident at 2 years) were investigated among other possible predictors of outcomes. The population-based sample included most of the twin pairs from the 2-year-old study with the addition of additional twin pairs who met the same inclusionary and exclusionary criteria, for a total of 627 twin pairs (197 MZ and 431 DZ, 610 boys and 645 girls) and one twin without a co-twin (whose missing score was estimated with modeling methods), for a total of 1,255 children. The study reported an unusually large number of speech, language, and nonverbal cognition standardized test phenotypes (a total of nine), assessing receptive vocabulary using the Peabody Picture Vocabulary Test–Third Edition (Dunn & Dunn, 1997), omnibus language using the Test of Language Development–Primary: Third Edition (TOLD-P:3; Newcomer & Hammill, 1997; subtests Spoken Language, Semantics, and Syntax), mean length of utterance and grammar/morphosyntax using the Test of Early Grammatical Impairment (Composite Score and Screener; Rice & Wexler, 2001), articulation/speech using the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000), and nonverbal IQ using the Columbia Mental Maturity Scale (Burgemeister et al., 1972). Each of the assessments contributed one phenotype for evaluation with the exceptions of the omnibus language test, TOLD-P:3 (three phenotypes: Spoken Language standard score and two subtests, namely, Semantics and Syntax) and Test of Early Grammatical Impairment (two phenotypes: Screener and Mini-finiteness).

The main outcomes were as follows. The expected patterns of correlations in support of heritability were evident across the language phenotypes (see Table 4, p. 89), in which within-pair correlations were higher for MZ than DZ twin pairs, and the correlations for both types of twins for within twin assessments were consistently positive, in the range of .45–.68, showing overall stability in a child’s rank within the group at age 4 years and again at age 6 years, and also supporting the reliability and validity of the phenotype measurements. The nonverbal IQ phenotype yielded heritability estimates of .21 at 4 years and .59 at 6 years. In the predictor models, maternal education was the only consistent significant predictor across language and nonverbal cognition phenotypes, although not for the speech phenotype. The nonverbal IQ phenotype yielded heritability estimates of .21 at 4 years and .59 at 6 years.

There was a persistent twinning effect across various language and speech phenotypes for ages 4 and 6 years, and twin type. Yet there was no twinning effect for the nonverbal
cognition phenotype. The twinning effect for language and speech phenotypes was consistent across phenotypes and not small. For example, standardized score averages for the omnibus language measure and the grammar measure were roughly 10 points below the expected mean of 100 (50th percentile) with an SD of 15, indicating that the groups at 4 and 6 years were, on average, about two thirds of a standard deviation below expectations compared to singleton children, approximately the 25th percentile. Yet the twinning effects decreased from 4 to 6 years. The MZ twins had significantly lower standard scores at 4 years than the DZ twins for six of the 10 phenotypes, but these differences were not evident at age 6 years for any phenotype.

Thus, regarding the twinning effect and possible twin type/zygosity effects, the conclusions were that twinning effects were still evident at 4 and 6 years, although they declined with age. The twin type/zygosity difference, such that MZ twins scored lower than DZ twins, was still evident at 4 years but resolved by 6 years. This is the first documentation of the age at which the zygosity effect is resolved. The causal mechanism driving this “catch-up” period for MZ twins is not known.

Similarly, substantial heritability estimates were replicated across language phenotypes and, as predicted based on earlier reports, increased with age. The heritability estimates were higher than reported previously, in the range of .44–.92 at 6 years. The highest estimate was .92 for the clinical grammar marker. These values would be in the moderate-to-high range.

**Outcomes for SLI and Nonspecific Language Impairment of Twins at 4 and 6 Years: Dynamic Categorical Change Over Time**

A subsequent study (Rice et al., 2020) used the same sample as the previous study for the purpose of examining how twinning effects influence the identification of children with language impairments at 4 and 6 years of age, comparing children with SLI and children with nonspecific language impairment (NLI). The total number of twin pairs was 677 (214 MZ and 463 DZ). The twins were sorted into two clinical groups who met the same criteria for language impairments (i.e., language scores more than 1 SD below their age mean, with standard scores of 84 or below). They differed in the exclusionary criteria for nonverbal IQ, following precedent established in an earlier epidemiological study (Tomblin et al., 1997). As before, each child in the SLI group (and the nonaffected children in the sample) met the inclusionary criteria of a nonverbal IQ of 85 or higher (i.e., greater than 1 SD below the mean for the test); the NLI group included children with nonverbal IQs at or below 84. This criterion has the further advantage of precedents in the previous twin studies defining groups by distance from the population mean, in that 85 is 1 SD below the population mean in the outcome measures. Note that this psychometric definition works against identification of twinning effects if there is more “wobble” around this cut-point, so it is not biased toward detecting heritability. At the same time, this cut-point is not so stringent as to leave very few children in affected groups, which can undermine the correlational analyses. Finally, it provides some room to evaluate deeper levels into lower performance on the phenotypes, as has been done in previous twin studies.

The purposes of the 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 SLI and children with NLI, and estimated levels of heritability for SLI and NLI. We predicted that the twinning effects would result in elevated rates of language impairments (defined in terms of below-age levels of performance) in twins. We also predicted, based on previous findings, that the heritability estimates would be higher for the clinical groups (i.e., SLI and NLI) than for the full sample of children and increase with age.

The key outcomes were as follows. The twinning effects, as predicted, elevated the proportion of children who met the SLI definition (28%) compared to 7%–8% in singleton population studies. The twinning effects were only slightly reduced from age 4 years (28%) to age 6 years (25%). For the NLI group, 8% were identified at 4 years, reduced to 3% at 6 years, compared to 2%–3% reported for population studies of singletons (Norbury et al., 2016; Tomblin et al., 1997). Recall there was no twinning effect for nonverbal cognition; 85% of the children had a standardized nonverbal IQ score of 85 or above at 4 and 6 years of age. The shift in clinical groups was attributed to gains in language, relative to age peers, between 4 and 6 years; across the phenotypes, about 50%–60% of the children were 85 or consistently above on language assessments at 4 and 6 years. There was no “double hit” for nonverbal IQ and language impairments (i.e., language scores were not the lowest for the low/low group).

As expected, heritability levels higher than those of the full sample in the first study were found for the clinical group outcomes, and, as expected, they increased with age. Separate heritability models were estimated for each phenotype, grouping (SLI/NLI), and age. Heritability was significant for the SLI group in multiple language phenotypes and, as in the full sample, was generally greater at 6 years (.55–.71), although not higher than in the full group. Because of the small NLI sample size for some measures, and change over time, interpretable heritability values for this group were limited to age 6 years TOLD-P:3 scores. Semantics (.86), Syntax (.85), and Spoken Language (.74). Heritability for nonverbal IQ was inconsistent across age and did not show congruence with speech and language phenotype heritability estimates. The grammar marker for finiteness outcomes (Rice & Wexler, 2001) is the only heritable phenotype theoretically and empirically linked to obligatory properties of the adult grammar, showing highest heritability for the Composite Score at age 6 years (.92).

Overall, the study reveals that nonverbal IQ is not on the same causal pathway as language impairments. Language acquisition is accelerated between 4 and 6 years as some of the twin children “outgrow” a twinning effect that delays early language acquisition. The twinning effect,
and subsequent acceleration, was not evident on a psychometrically robust assessment of nonverbal cognition, and there was no “double hit” of low language–low nonverbal cognition. Note that although the NLI group did not have lower language levels than the SLI group, the NLI group is nevertheless informative because the overall patterns of heritability are not the same across the two clinical groups, related to the new finding that the twinning effect is not the same across language and nonverbal IQ. Had the two groups been combined, this important difference may have been obscured. Heritability estimates are most consistent in the SLI group; calculations for some phenotypes were affected by the relatively low sample size in the NLI group.

**Grammar Clinical Marker Yields High Heritability for Language Impairments in 16-Year-Old Twins**

Convergent outcomes were reported in a recent study of twins at 16 years of age (Dale et al., 2018). The twin sample for this study was part of the TEDS, a population-based longitudinal study of language, cognition, and behavior problems from childhood through adolescence (Oliver & Plomin, 2007). The TEDS subsample for this study consisted of 810 16-year-old twins (405 pairs). The purpose of the study was to evaluate if a grammatical judgment measure previously shown to be a clinical marker of language impairment across a wide age range in young children (Rice et al., 2009) would be a valid clinical marker of language impairment in older adolescents. Three language phenotypes were assessed: a grammaticality judgment (GJ) task for finiteness marking, vocabulary, and figurative language. The GJ task was adapted from earlier studies of children with SLI (Rice et al., 2009) and is theoretically related to the finiteness assessments of the studies of the younger twins described above, with items appropriate for older children. The validity of the GJ task in this study was supported by these outcomes for the children scoring in the lowest 10%. Their mothers had significantly lower levels of education, the children scored significantly lower on TEDS tests of vocabulary and figurative language at the same time of assessment as the GJ task, and the children had significantly lower scores on school-administered assessments of English and mathematics.

As expected, the within–twin pair concordances (number of twin children whose co-twins are affected/total number of twin children) were higher for MZ twins than DZ twins for the lowest 10% of the sample (.28/.19), the lowest 7% (.40/.24), and the lowest 5% (.43/.15). Heritability for the GJ marker was substantial and increased with the severity of the language impairment: lowest 10%, .36; lowest 7%, .47; and lowest 5%, .74. These heritability outcomes range from moderate to high, depending on the criterion used for low performance. Note that the range for the heritability outcomes is similar to those reported above (Rice et al., 2020) using a criterion of a standard score of 85 as a cutoff, approximately the 15th percentile. Overall, the outcomes support the potential value of the GJ task as a clinical marker of SLI. It is also a good candidate as a phenotype for molecular genetics research (https://www.phenxtoolkit.org).

**Recap of the Empirical Outcomes of Recent Twin Studies**

In the context of previous studies, the outcomes supported a priori predictions for the overall patterns of heritability of language and SLI. The replicated pattern of medium-to-high levels of heritability (replicated longitudinally, across ages and phenotypes, within the same sample of twins and replicated across independently ascertained twin samples) supports the conclusion that language phenotypes are under genetic influence. In particular, the evidence is consistent with a genetic contribution to SLI across the age span of 2–16 years and presumably into adulthood. At the same time, the studies reveal a twinning effect on language acquisition, such that twin language acquisition has risk for delay relative to singletons in the preschool years and is not fully resolved at school entry (6 years). In short, compared to singleton acquisition data during the preschool period, a higher percentage of twins than expected are at the low end of language abilities, and this twinning effect is more evident for MZ than DZ twins at 4 years but not 6 years. This twinning effect is not evident for nonverbal cognitive abilities. There is no indication of a “double hit” (i.e., the pattern is of nearly equivalent language phenotype means for children in the NLI group compared to the SLI group).

The consequence of these new facts about twin language acquisition (twinning effect for language, yet no twinning effect for nonverbal cognition, no indication of double hits, and an early higher risk for MZ than DZ twins) is that precision in estimating heritability requires differentiation between SLI and NLI groups of children with language impairment especially in the age range of 2–6 years. Consistency of low levels of performance is lower for language phenotypes and more dynamic, as the twin children accelerate to close the gap with their singleton age peers, whereas low levels of performance are more consistent for low levels of nonverbal IQ. Furthermore, the interpretation requires attention to the distinction between levels of performance and associations/correlations within twin pairs. Higher within-pair correlations for MZ twin pairs support high heritability estimates, yet the MZ pairs are more likely to score lower than DZ pairs until 6 years. Overall, the empirical, and convergent, evidence has benefited from a legacy of previous scientific studies of twins and more recent programmatic longitudinal investigation across different groups of scientists.

**Revisiting the Bigger Picture of Genetic Effects on Language Acquisition and Language Impairments in Children**

Lenneberg’s prescient thesis anticipated two of the key outcomes of the program of investigation summarized...
here: (a) Innate drivers of language acquisition are revealed in replicated findings of moderate-to-high heritability of language acquisition in twins. The outcomes are among the strongest indicators of inheritance for behavioral phenotypes. (b) Although heritability estimates are relatively high across different dimensions of language, the outcomes are not uniform across different speech and language phenotypes, even when considering the variations across ages.

The one consistently high level of heritability is reported for the morphosyntactic feature of finiteness, first reported in twins by Bishop et al. (2006), where the finiteness phenotype was based on early studies of finiteness and a theory proposed by Rice et al. (1995). The twin study of Bishop et al. (2006) evaluated two phenotypes in 173 twin pairs, two experimental measures of verb tense/finiteness markers very similar to what subsequently appeared in a standardized test (Rice & Wexler, 2001) and a delay test of phonological short-term memory. The twins were grouped according to risk for language impairments versus no risk, based on parental report at 4 years of age. The conclusions were that both phenotypes were significantly heritable, although there was minimal overlap between the two phenotypes, “suggesting that different genes are implicated in causing these two kinds of language difficulty. From an evolutionary perspective, these data are consistent with the view that language is a complex function that depends on multiple underlying skills with distinct genetic origins” (Bishop et al., 2006, p. 158).

In an updated theoretical proposal, based on extensive longitudinal data outcomes on finiteness markers and vocabulary, Rice (2012) postulated that two genetic elements were implicated in the developmental course of language acquisition by children with SLI and unaffected children. The general pattern (extrapolating across growth sensitive measures [not standard scores] of vocabulary, mean length of utterance, and morphosyntax/finiteness marking in obligatory sentence contexts) is that children with SLI start language later but, once started, continue in trajectory parallel to that of unaffected children, meaning they do not “catch up” to their age peers and, at early adolescence, begin to decline in rate of acquisition, as do their age peers. Such timing and rate differences in development are not specific grammatical structures and perhaps are more likely to comprise a cross-linguistically valid broad phenotype than any one individual phenotype. The neuronal mechanisms for start-up and acceleration of language acquisition are unknown. Yet investigations of the immune system provide a model of how acceleration and deceleration could be regulated (Sharma et al., 2011). In effect, T cells require two signals to attack a target effectively, one that functions as an “ignition switch” and another as a “gas pedal.” These two signals are in a timed relationship with a third signal that eventually inhibits T cell activity. This discovery led to vital advances in drug treatment for cancer and a 2018 Nobel Prize for James Allison. Such closely synchronized epigenetic effects are now widely recognized as operative in molecular genetics. In the 2012 paper, I proposed a growth signaling disruption model, which posits that faulty timing mechanisms at the cellular level, somewhere in the complex underlying neuronal networks for language onset and growth regulation, are at the core of SLI.

I suggest that the new findings for twins are consistent with this model, with the additional suggestion that twinning brings unique risk to the language onset signals, more likely in MZ than DZ twins. This source of risk is not exactly the same as for singleton children with SLI because it can be followed by an acceleration burst (relative to singleton language acquisition rates) to compensate for the initial delay, at least in a proportion of the twins. This kind of an acceleration burst appears in some singleton children as well, where it is known colloquially as “outgrowing” early immature language acquisition. The best evidence of LLE is found in population-based studies, which identifies about 15% of singleton children who show LLE at 24 months (Zubrick et al., 2007). At 7 years of age, comparing a group of children documented with LLE at 2 years with a group without LLE at 2 years (Rice et al., 2008), 20% of children with the positive history performed below normative expectations, as SLI, compared to 11% of the children without LLE at 2 years. The persistent delays were most likely in general language ability, syntax, and morphosyntax, but not vocabulary or semantics. Thus, the compensatory acceleration burst in singletons is least likely for the language phenotypes with higher levels of heritability in the twin studies.

The acceleration mechanisms are unknown for the twins but quite remarkable. The risk for LLE is greater for twins than singletons with the important distinction that the twin type matters: MZ twins are more likely to have early delays than DZ twins, although, for unknown reasons, this difference has largely disappeared by 6 years. In other words, the MZ twins are accelerating language acquisition faster than the DZ twins, and both kinds of twin pairs are at approximately singleton norms by 6 years. If these acceleration mechanisms were available to all children, the condition of SLI would not exist.

The point here is that, as molecular-level genetics regulatory mechanisms are better understood in the domain of neurobiological underpinnings of language acquisition, there is reason to expect an initial onset mechanism followed by a longitudinal regulatory mechanism that sustains a mostly linear acquisition curve for vocabulary development and more curvilinear functions as grammar reaches adult levels, all of which begin to decelerate in preadolescence. Twins may provide valuable future clues for the earliest beginnings of the genetic underpinnings for inherited language acquisition skills that serve as the base platform for the environmental effects of learning, including the relatively quick adoption of the native language operative in a child’s home environment with the associated differences in the surface manifestation of underlying linguistic principles. Twin study outcomes of differences between language acquisition and nonverbal cognition are important elements in moving our science forward. The lack of a twinning effect for nonverbal cognition versus the notable twinning effect for language acquisition is an important
hint of nonoverlap in some of the hypothesized underlying neurodevelopmental triggers for early development and mechanisms to maintain ongoing developmental progress. Note that the distinction between SLI and NLI is lost in the newly coined term “developmental language disorders,” which would combine the two groups into one larger, more diverse group (Bishop et al., 2017). Extensive discussions of the consequence of this new, more diverse clinical group have appeared elsewhere (Rice, 2020; Volkers, 2018) and will not be repeated here. Scientific investigations in the modern era are enhanced by specificity of outcomes, as in the example of T cell function and in well-understood clinical conditions such as breast cancer, where subtypes of breast cancer have replaced the much less useful legacy term “lump in the breast.” Recent modeling studies make it clear that more precise phenotypes for depression, for example, have greater scientific informativeness than broad/minimal phenotypes in genetic studies (Cai et al., 2020). In the same way, a separate grouping for SLI reveals differences in developmental outcomes and heritability estimates. Furthermore, morphosyntax as a precise phenotype consistently reveals differences from vocabulary phenotypes.

Lenneberg did not anticipate the enormous advance in our understanding of molecular genetics, the advanced modeling methods for estimating heritabilities and controlling for covariates, and the increased knowledge of children’s acquisition of language and particular properties of grammar in the area of morphosyntax. Overall, we have a more extensive research base to build on in future studies of SLI than ever before. Menyuk’s important questions about the details of children’s grammars are as relevant as ever although notable progress has been achieved. Our leading sources of financial support for research emphasize the need for empirical research that meets the standards for rigor and reproducibility. The specificity of the new evidence base emerging for SLI is an important asset for adding to our research discoveries and is necessary for scientific progress moving forward.

Acknowledgments

This article stems from the 2019 Research Symposium at ASHA Convention, which was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award R13DC003383. Preparation of this review article and much of the research reported here was supported by National Institutes of Health Grants R01DC001803 and R01DC005226 (PI: M. L. Rice), as well as Grant T32DC000052 as doctoral student support (PI: M. L. Rice). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Appreciation for editorial feedback goes to Kathleen Kelsey Earnest.

References

Benton, A. L. (1964). Developmental aphasia and brain damage. Cortex, 1(1), 40–52. https://doi.org/10.1016/S0010-9452(64)80012-5

Bishop, D. V. M., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. Genes, Brain and Behavior, 5(2), 158–169. https://doi.org/10.1111/j.1601-183X.2005.00148.x

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the 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 outcomes (2nd ed.). Taylor & Francis. https://doi.org/10.1201/b14615

Brown, R. (1973). A first language: The early stages. Harvard University Press.

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

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., Clarke, T.-K., Forstner, A. J., Grabe, H. J., Hamilton, S. P., Levinson, D. F., Lewis, C. M., Lewis, G., Martin, N. G., Milaneschi, Y., Mors, O., Müller-Myhsok, B., Penninx, B. W. J. H., Perlis, R. H., … Flint, J. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nature Genetics, 52, 437–447. https://doi.org/10.1038/s41588-020-0594-5

Chomsky, N. (1957). Syntactic structures. Mouton.

Chomsky, N. (1967). The formal nature of language. In E. Lenneberg (Ed.), Biological foundation of language (pp. 397–442). Wiley.

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., Rinfield, K., & Hayiou-Thomas, M. E. (2018). Grammar clinical marker yields substantial heritability and genetic influences on grammar and phonological short-term memory skills in 4.5-year-old twins. Journal of Speech, Language, and Hearing Research, 61(1), 66–78. https://doi.org/10.1044/2017_JSLHR-L-16-0364

Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test–Third Edition. AGS.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E. A., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). MacArthur Communicative Development Inventories. The Psychological Corporation.

Goldman, R., & Fristoe, M. (2000). Goldman-Fristoe Test of Articulation–Second Edition. AGS.

Hay, D. A., Prior, M., Collette, 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

Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V. M., & Plomin, R. (2005). Genetic influences in different aspects of language development: The etiology of language skills in 4.5-year-old twins. Child Development, 76(3), 632–651. https://doi.org/10.1111/j.1467-8624.2005.00868.x

Lenneberg, E. (1967). Biological foundations of language. Wiley. https://doi.org/10.1080/21548331.1967.11707799

Leonard, L. B. (1972). What is deviant language. Journal of Speech and Hearing Disorders, 37(4), 427–446. https://doi.org/10.1044/jshd.3704.427

Leonard, L. B. (2014). Children with specific language impairment (2nd ed.). MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001
Martin, J. A., Hamilton, B. E., & Osterman, M. J. K. (2012). Three decades of twin births in the United States, 1980-2009. *NCIHS Data Brief, 80*, 1–8.

Marx, O. (1967). The history of the biological basis of language. In E. Lenneberg (Ed.), *Biological foundations of language* (pp. 443–469). Wiley.

Menyuk, P. (1964). Comparison of grammar of children with functionally deviant and normal speech. *Journal of Speech and Hearing Research, 7*(2), 109–121. https://doi.org/10.1044/jshr.0702.109

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

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vannvakas, 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 multivariate, 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

Redmond, S. M., Ash, A. C., Christopolous, T. T., & Pfaff, T. (2019). Diagnostic accuracy of sentence recall and past tense measures for identifying children’s language impairments. *Journal of Speech, Language, and Hearing Research, 62*(7), 2438–2454. https://doi.org/10.1044/2019_JSLHR-L-18-0388

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 27. https://doi.org/10.1186/1866-1955-4-27

Rice, M. L. (2019). Longitudinal study of language and speech of twins: New insights into possible causal pathways. In E. Grigorenko, Y. Shtyrov, & P. McCadde (Eds.), *All about language: Science, theory, and practice* (pp. 82–96). Brookes.

Rice, M. L. (2020). Clinical lessons from studies of children with specific language impairment. *SIG 1 Language Learning and Education, 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: 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(200908-0171)

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(20080209)

Rice, M. L., Taylor, C. L., Zubrick, S. R., Hoffman, L., & Earnest, K. K. (2020). Heritability of specific language impairment and nonspecific language impairment at ages 4 and 6 years across phenotypes of speech, language, and nonverbal cognition. *Journal of Speech, Language, and Hearing Research, 63*(3), 793–813. https://doi.org/10.1044/2019_JSLHR-19-00012

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

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research, 38*(4), 850–863. https://doi.org/10.1044/jslhr.3804.850

Rice, M. L., Zubrick, S. R., Taylor, C. L., Gayán, J., & Bontempo, D. E. (2014). Late language emergence in 24-month-old twins: Heritable and increased risk for late language emergence in 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., & Gayan, 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), 79–83. https://doi.org/10.1044/2017_JSLHR-L-16-0366

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

Sharma, P., Wagner, K., Wolchok, J. D., & Allison, J. P. (2011). Novel cancer immunotherapy agents with survival benefit: Recent successes and next steps. *Nature Reviews: Cancer, 11*, 805–812. https://doi.org/10.1038/nrc3153

Stark, R., & 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

Strachan, T., & Read, A. (2011). *Human Molecular Genetics*. Garland Science, Taylor & Francis Group.

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 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., Backwalter, 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

Volkers, N. (2018). Diverging views of language disorders: Researchers debate whether the label “developmental language disorder” should replace “specific language impairment.” *The ASHA Leader, 23*(12), 44–53. https://doi.org/10.1044/leader.FTR1.23122018.44

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50*(6), 1562–1592. https://doi.org/10.1044/1092-4388(2007/106)