JSLHR

Research Article

Late Language Emergence in 24-Month-Old Twins: Heritable and Increased Risk for Late Language Emergence in Twins

Mabel L. Rice,[a] Stephen R. Zubrick,[b] Catherine L. Taylor,[b] Javier Gayán,[c] and Daniel E. Bontempo[a]

Purpose: This study investigated the etiology of late language emergence (LLE) in 24-month-old twins, considering possible twinning, zygosity, gender, and heritability effects for vocabulary and grammar phenotypes.
Method: A population-based sample of 473 twin pairs participated. Multilevel modeling estimated means and variances of vocabulary and grammar phenotypes, controlling for familiality. Heritability was estimated with DeFries–Fulker regression and variance components models to determine effects of heritability, shared environment, and nonshared environment.
Results: Twins had lower average language scores than norms for single-born children, with lower average performance for monozygotic than dizygotic twins and for boys than girls, although gender and zygosity did not interact. Gender did not predict LLE. Significant heritability was detected for vocabulary (0.26) and grammar phenotypes (0.52 and 0.43 for boys and girls, respectively) in the full sample and in the sample selected for LLE (0.42 and 0.44). LLE and the appearance of Word Combinations were also significantly heritable (0.22–0.23).
Conclusions: The findings revealed an increased likelihood of LLE in twin toddlers compared with single-born children that is modulated by zygosity and gender differences. Heritability estimates are consistent with previous research for vocabulary and add further suggestion of heritable differences in early grammar acquisition.

Key Words: language acquisition, twin language, late language emergence, twin zygosity effects, twin early grammar

The relative contributions of nature and nurture are long-standing issues in studies of toddlers' language acquisition. Although language emergence is generally apparent by 24 months of age, some children show *late language emergence* (LLE), defined as language below age and gender expectations in children without other disabilities (Zubrick, Taylor, Rice, & Slegers, 2007). The cause of this unexplained individual variation is unknown. Twin children provide unique opportunities to differentiate genetic and nongenetic influences on language acquisition. Theoretically, monozygotic (MZ) co-twins inherit identical genomes, and dizygotic (DZ) co-twins share only half of their segregating DNA. Behavioral genetics methods yield estimates of the relative contribution of genetic and nongenetic factors to a specific phenotype. Previous studies have reported significant heritability for variation in language acquisition abilities in toddlers, as well as significant heritability of low levels of early language performance (Stromswold, 2001).

One limitation of the contemporary heritability estimates for toddlers is that the studies have not addressed possible *twinning effects,* defined here as a lower level of language performance for children who are twins compared with that of single-born children. A possible twinning effect on early language acquisition is generally accepted, although we argue here that further empirical evidence is needed to determine whether the generalization holds across samples, and, if the twinning effect appears, the implications for interpretations of outcomes. In the context of behavioral genetics studies, twinning effects could elevate the proportion of LLE in children who are twins, irrespective of zygosity. Under the assumptions of behavioral genetics methods, a twinning effect irrespective of zygosity would contribute to similarities across MZ and DZ co-twins, adding to estimated shared environment effects and reducing possible heritability effects. Clear interpretation of effects attributable to environment versus heritability thus requires investigation of

[a]University of Kansas, Lawrence
[b]Telethon Institute for Child Health Research and the University of Western Australia, Perth
[c]Bioinfosol, Seville, Spain
Javier Gayán is now affiliated with F. Hoffmann-La Roche AG, Basel, Switzerland.
Correspondence to Mabel L. Rice: mabel@ku.edu

Editor: Janna Oetting
Associate Editor: Laura DeThorne
Received November 5, 2012
Revision received March 27, 2013
Accepted August 23, 2013
DOI: 10.1044/1092-4388(2013/12-0350)

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

possible twinning effects. Another limitation is that the language phenotypes used in previous behavioral genetics investigations of toddlers have primarily, although not exclusively, focused on vocabulary size. Yet in this early period of development, children may show emergent indicators of grammatical abilities that warrant further evaluation as phenotypes. Gender differences may play a role, as well as possible zygosity effects, both of which are relatively unexplored. All told, there is a need for a behavioral genetics study that addresses possible twinning, zygosity, and gender effects in the same sample for which heritability effects are estimated. In a broader context, the possible effect of twinning on language acquisition has heightened social relevance given the dramatic increases in twin conceptions over the last three decades (Chauhan, Scardo, Hayes, Abuhamad, & Berghella, 2010; Pinborg, 2005).

The aims of this study were to (a) document language abilities of 24-month-old children who are twins; (b) describe possible twinning effects for vocabulary and grammar phenotypes, relative to population norms, zygosity, and gender; (c) calculate heritability estimates for vocabulary and grammar phenotypes; and (d) consider how the findings inform interpretations of heritability estimates.

Behavioral Genetics of Toddler Twin Language Acquisition: Heritability Estimates

The methods of behavioral genetics calculate the proportion of variance per measure due to genetic, shared environment, and nonshared environment factors. The underlying logic is that the greater the difference between MZ and DZ intraclass correlations for a trait, the higher the heritability. A general review of twin studies of language skills across a wide age range (Stromswold, 2001) reported heritability estimates between 0.35 and 1.0, suggesting a range of 35% to 100% of the variance accounted for by genetic factors.

Reported heritability levels are in a more narrow range for the toddler period of language acquisition. Vocabulary at age 2 years in a sample of more than 3,000 twins in the Twins Early Development Study (TEDS) was evaluated by Dale et al. (1998) using the short form of a parent-report questionnaire, the MacArthur Communicative Development Inventory (Fenson et al., 1993). They reported an individual differences heritability estimate of 0.25 and a shared environment estimate of 0.69 for the full sample of twins. When the sample was subdivided into two groups, according to whether or not at least one member of the twin pair was in the lowest 5% of the distribution for vocabulary, heritability of group differences was 0.73. The conclusions were that early language delay was a distinct disorder and that the effects of shared environment were less important for the language-delayed sample (18%) than for the entire sample (69%).

The same 2-year-old TEDS sample was investigated by Dale, Dionne, Eley, and Plomin (2000) who reported heritability estimates of 25% for vocabulary scores and 39% for sentence complexity scores (a task in which parents select sentences most like the ones their child says). The results are similar to those reported by Reznick, Corley, and Robinson (1997) with a parent-report language phenotype on 2-year-old twins. In a follow-up study of their 2-year-old TEDS sample, Dale, Price, Bishop, and Plomin (2003) found a greater risk for language impairment at 3 and 4 years of age (44.1%) for twin children who were below the 10th percentile at 2 years of age compared with children above the 10th percentile (8.5%), suggesting that LLE for twin children at 24 months establishes greater risk for language impairments later on. A companion study of the same sample reported higher heritability estimates when persistent difficulties were defined according to whether parents expressed concern about language at 3 years or whether a professional had been consulted about language difficulties at 4 years (h^2 of 0.41 for 2-year vocabulary; Bishop, Price, Dale, & Plomin, 2003). DeThorne, Petrill, Hayiou-Thomas, and Plomin (2005) evaluated parental report data of language and cognitive abilities in 4-year-old twins, comparing a group defined as "nonspecific expressive language impairment" (children with low vocabulary and low cognitive levels benchmarked to 1 *SD* or more below the mean) and a group with "specific language impairment" (children with low vocabulary, 1 *SD* or more below the mean, and no cognitive impairment). For the full sample, combined across the two groups, for a measure of expressive vocabulary they report a heritability index of 0.26, shared environmental influence of 0.64, and nonshared environmental factors of 0.10. For the nonspecific expressive language impairment group, the outcomes depended on the severity of impairments, with higher heritability as the severity of impairment increased, in the range of 0.24–0.48. For the specific expressive language impairment group, heritability estimates also increased with increased impairment, in the range of 0.25–0.48. Comparison of their outcomes with previous reports from the TEDS sample, with differences in sampling and measures, led the authors to conclude that heritability estimates around 0.25 should be expected for language impairment defined as 1 *SD* below the mean, increasing to around 0.45 for more severe impairment. These estimates are consistent with the results of Tomblin and Buckwalter (1998) with a smaller school-age sample of twins and triplets.

Twinning Effects

Children conceived as twins may be more likely to meet criterion for LLE. The studies reviewed above did not consider possible twinning effects on language acquisition, although the relatively low level of environmental effects in the group of children at the bottom of the sample would suggest that the environmental effects were probably not inflated due to twinning effects. Yet it would be difficult to assess whether the twin sample performed lower than what is expected in the general population of single-born children because standardized test data normed on the general population were not reported.

Recent reports that have focused on twinning effects are sparse, with two key studies. Rutter, Thorpe, Greenwood, Northstone, and Golding (2003) compared the language development of 20-month-old twins and single-born children in the Avon Longitudinal Study of Parents and Children (ALSPAC). They found lower levels of language in twins than in single-born children, with a vocabulary delay of about 1.7 months at 20 months (parental report on MacArthur Communicative Development Inventory; Fenson et al., 1993), increasing to 3.1 months delay in language at 3 years (based on McCarthy Scales of Children's Abilities [McCarthy, 1972], administered by examiners). Boys were more likely to be delayed than girls. Several limitations of the study work against interpretations of how twinning effects could affect LLE. An LLE group was not identified; the scores at 20 months are not interpretable in terms of a child's relative status in the full group; and the reported age equivalent scores are not comparable to standard scores (Bracken, 1988; Maloney & Larrivee, 2007). An investigation of 30-month-old children in the LaTrobe Twin Study (Hay, Prior, Collette, & Williams, 1987) reported an increased likelihood of speech and language impairment in preschool twins, with a greater effect for boys. Twin boys were 8 months behind twin girls and matched single-borns on expressive language and 6 months behind on receptive language.

If twinning effects are replicated across samples, the possible sources are likely to be complex, involving prenatal and perinatal differences between twins and single-born children, as well as unique parenting challenges posed by twins (Koeppen-Schomerus, Eley, Wolke, Gringras, & Plomin, 2000; Stromswold, 2006; Thorpe, 2006). A prominent conclusion is that the twinning effects are "almost certainly environmental" (Thorpe, 2006, p. 391), including adverse conditions in the prenatal and perinatal environment and disrupted social interactions. Some investigators have hypothesized that the disrupted interactions involve effects on caregivers' language interaction with their twin children that create differences relative to interactions with single-born children. This could play out as caregivers needing to divide attention between two infants amid the increased burden of childrearing, thereby reducing the verbal input directed to each member of the twin pair (Clark & Dickman, 1984; Lytton, Conway, & Sauve, 1977; Savic, 1980; Tomasello, Mannie, & Kruger, 1986). Thorpe, Greenwood, and Rutter (2003) compared, for 20-month-olds, naturalistic triadic interactions in the homes of parents with twins and parents with two closely spaced single-born children. They reported that there were differences between the two types of triads in interactions that explained differences in language development. They noted that the interactive differences may be attributable to the older sibling's effect in the single-born triads, which influenced the amount of time that the children received their mother's full attention, or the extent to which children's bids for attention were ignored. There were no significant triad group differences in time interacting with siblings or in exclusive interaction with the mother. Thus, environmental input differences are not straightforward when comparing caregiver interactions with twins versus single-born children with siblings. A limitation of this study is that children with LLE were not identified, and possible differences were not examined for children with LLE versus those without LLE, leaving unresolved how caregiver interactions may or may not influence twinning effects in children with LLE.

Although not highlighted in the literature, it is also conceivable that there are potentially advantageous effects on language acquisition for twin children, such as a consistent peer conversational partner, even at the earliest stages of development. If such effects occur, they would lessen the likelihood of a twinning effect on language as defined here, that is, a late appearance of language for twins during the toddler period of development.

Zygosity Effects

Possible zygosity effects have been overlooked in the studies focusing on possible twinning effects, although differences between MZ and DZ twins, if they exist, would help evaluate possible causal influences. For example, the hypothesis of reduced caregiver attention or adjusted caregiver interactions as a result of having two babies of the same age would suggest no zygosity effects, under the assumption that caregivers do not adjust their interactions according to the type of twin pair (alternatively, if there are zygosity effects, this would suggest the need to consider more than the effects due to two babies of the same age, perhaps involving other prenatal or perinatal differences between MZ and DZ twin pairs). On the other hand, the behavioral genetics studies have reported a DZ advantage for early language milestones at 24 months (Dale et al., 1998, 2000), although this factor is not discussed further. More information is needed about zygosity effects as possible mediators of twinning effects, because such effects could influence interpretations of heritability estimates and, ultimately, interpretations of the sources of the twinning effects.

Gender Effects

There is strong evidence of gender effects on language acquisition at 24 months. Early vocabulary assessments provide separate normative data for boys and girls because toddler boys, on average, have fewer words in their vocabularies than girls of the same age (Fenson et al., 1993). In a population-based single-born sample, boys were three times more likely to show LLE after controlling for associations with other predictor variables (based on a measure not adjusted for gender differences; Zubrick et al., 2007). The twinning effect is more apparent in boys than girls (Hay et al., 1987; Thorpe, 2006). Although a gender effect for lower language levels in twins suggests that gender effects might interact with twinning effects (Hay et al., 1987; Thorpe, 2006), there remains a need for consideration of gender effects in samples of twins in behavioral genetics investigations, with adjustment for possible gender effects, if needed, in modeled estimates of heritability.

Method

Approval to conduct this study was obtained from the Curtin University of Technology Human Research Ethics

Committee, the Department of Health Western Australia Human Research Ethics Committee, and the University of Kansas Human Research Committee.

Participants

The study design was a prospective cohort study of twins drawn from a total population sample frame comprising statutory notifications of all births in Western Australia (WA) in 2000–2003 (Gee & Green, 2004). The population of WA is demographically similar to some states in the Midwestern United States. For example, census data show that the population of the state of Kansas is 2.7 million and the population of WA is 2.1 million. In each state, most of the population is in urban areas. The states are predominantly white (86% for Kansas; 96% for WA), and the majority of the population are native speakers of English, well educated (86% of the population has completed high school in each state), and family oriented (in Kansas, 55% of all families are couples with children, and 9% are single-parent families; in WA, these rates are 49% and 15%, respectively). On a wide variety of behavioral and biological assessments of children and adults, distributional outcomes conform to normative expectations for instruments normed in the United States or the United Kingdom.

There were 1,135 sets of live twins born in this time period. We were able to trace 941 (83%) of these twin families by mail, and 698 (74%) consented to participate in the study. This represents ascertainment of 61% of all twins born in WA in 2000–2003. A comparison with data available for all twins born in 2000–2003 showed that the participants were broadly representative of the total twin population from which they were drawn. Participants and nonparticipants were compared on three postal area measures of socioeconomic status calculated by the Australian Bureau of Statistics (2001) from national census data. The measures were the Index of Disadvantage, the Index of Economic Resources, and the Index of Education and Occupation. There were no significant differences between participants and nonparticipants on the Index of Disadvantage or the Index of Economic Resources. On the Index of Education and Occupation, the only significant difference between the groups was at the second quintile. The majority of mothers in the participant group were white, 20–34 years old, married, and living in the metropolitan area. The participant group, compared with the nonparticipant group, contained more white mothers, married mothers, and fewer young mothers. This is consistent with other population level studies using the same recruitment methods, drawn from the same region with similar socioeconomic status distributions (Zubrick et al., 2007). There was no significant difference in the percentage of participants versus nonparticipants living in the metropolitan or rural regions of WA. Ten additional families with twins born 2000–2002, who were not recruited through the statutory notification of births, approached the study and were included in the cohort.

Exclusionary criteria were applied to form the final sample. Twins with exposure to languages other than English (52 twin pairs) or twin pairs in which at least one twin had hearing impairment, neurological disorders, or developmental disorders (14 twin pairs) were later excluded from the twin sample. Exclusionary conditions included Down syndrome, Angelman syndrome, cerebral palsy, cleft lip and/or palate, agenesis of the corpus callosum, and global developmental delay. The intent was to limit the sample to children who did not have concurrent conditions likely to affect language acquisition. The exclusionary criteria resulted in 633 twin pairs who were eligible to participate in the study.

At 24 months, data were collected by postal questionnaire. Questionnaires were sent 1 month prior to the twins' second birthday. The response rate to the questionnaires was 75%. Therefore, data were available for 474 eligible twin pairs of approximately 2 years of age (in days, mean age is 755.8, range, 687–899). There were 454 boys (47.9%) and 494 girls (52.1%). Twin zygosity was determined by molecular analysis of buccal swab samples. For twin pairs with unknown zygosity, a discriminant analysis of questionnaire items reported by parents was used to assign zygosity. The final twin counts were 313 DZ pairs and 160 MZ pairs, for a total of 473 pairs and 946 individuals. Sex within twin pairs was roughly evenly distributed: 36% female, 32% male, and 31% opposite sex. One twin pair did not provide a DNA sample, and the questionnaire analysis remained inconclusive; the pair was therefore excluded from heritability analyses.

Birth data were extracted from the Midwives' Notification System (MNS). These data are collected by statute on all live births, still births, and neonatal deaths in WA. The data comprise 150 variables on maternal demographics, pregnancy profile, labor and birth, and birth outcomes. Means and *SD*s for selected variables are reported in Table 1, broken out by zygosity. Maternal age at birth of twins differed by zygosity (t = 2.931, p = .003). The mothers of DZ twins were, on average, 12 months older than the mothers of MZ twins. Maternal height of DZ twins was, on average, about 2 cm taller than mothers of MZ twins (t = 3.863, p = .000). Other potential predictors of zygosity differences were not significant: maternal history of smoking cigarettes during pregnancy, number of previous pregnancies, education, or income. At birth, the DZ twins were longer (t = 1.936, p = .05) and heavier (t = 2.029, p = .04). There were no zygosity differences in Apgar scores at 1 or 5 min. Estimated gestation in weeks did not differentiate the twin types. As expected, the twins were smaller at birth than single-born children. The twins' mean (*SD*) birth weight in grams was 2,472.74 (543.42) compared with 3,397 (515.7) reported in a large population-based sample of single-born children (Zubrick et al., 2007).

Language Measures

Caregivers completed a questionnaire when the children were 24 months of age that included the MacArthur Communicative Development Inventories: Words and Sentences (CDI-WS; Fenson et al., 1993, 2007). This questionnaire's psychometric properties are well suited to the aims of the study. The demographics of the twin sample

Table 1. Maternal and child variables at birth.

Variable	DZ *M*	DZ *SD*	MZ *M*	MZ *SD*
Age	31.77	4.85**	30.78	4.98
Height (cm)	167.04	6.68***	165.10	6.79
Smoking during pregnancy[a]	1.88	0.32	1.91	0.29
Number of previous pregnancies	1.30	1.43	1.15	1.26
Education[b]	8.61	2.45	8.35	2.20
Income[c]	11.84	3.23	12.01	3.06
Child birth indices				
Body length (cm)	46.15	3.31*	45.74	2.85
Apgar score 1 min	7.83	1.49	7.73	1.51
Apgar score 5 min	9.00	0.80	8.92	0.76
Weight (g)	2,487.67	562.62*	2413.81	500.89
Estimated gestation (weeks)	35.53	2.63	35.69	2.35

Note. DZ = dizygotic; MZ = monozygotic.

[a]1 = yes, 2 = no. [b]Coded in 12 increments, where 1 = did not go to school, 7 = completed high school, 10 = diploma beyond high school, and 12 = postgraduate diploma/higher degree. [c]Coded according to 2001 Australian census, with income levels in 15 increments from $0 to $124,800 or more per year, where 11 = $36,400–$41,599 per year.

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

favorably compare with the U.S. population data and the CDI-WS sample reported in Fenson et al. (2007), as summarized in Table 2. Our twin sample is predominately white, somewhat more than the U.S. and CDI-WS data. To the extent that there is an advantage for white children on vocabulary development, this factor would work against a lower percentile on the CDI in our sample of twins. Table 2 shows that the twin sample falls between the U.S. Census estimates and the CDI reference sample means for mother's education. To the extent that higher mother's education is associated with enhanced vocabulary acquisition, this would act to lower the percentiles in the twin sample, but not as much as it would in a sample stratified according to the U.S. Census. The other consideration is exposure to multiple languages in the home. The CDI reference group includes such children, and our sample carefully excluded them. To the extent that early multilingual exposure delays early English word learning, the comparison with the CDI norms would inflate the percentiles of the twins, hence working against detection of a twinning effect. The number of participants in the CDI-WS 2007 norms is reported in monthly intervals. At 24 months, the reference sample is 72 girls and 63 boys. It is an advantage for this study to have norms broken out by gender. The CDI norms can be compared to the Bayley Scales of Infant Development (Bayley, 2006), which report approximately 40 children per month.

Table 2. Demographic profile of U.S. population compared with the sample in CDI Words and Sentences and the present study.

Characteristic	United States (%)	CDI Words and Sentences (%)	Twin sample (%)
Race			
White	75.1	73.7	89.0
Black	12.3	9.7	0.0
Asian	3.6	2.7	0.2
All other/mixed	8.9	6.9	10.7
Maternal education			
Some high school or less	15.6	8.2	12.9
High school diploma	26.2	24.1	35.4
Some college education	34.9	24.5	21.6
College diploma	23.2	43.2	30.1

Note. U.S. data from 2000 Census Overview of Race and Hispanic Origin. CDI Words and Sentences data based on 2007 norms from Fenson et al. (2007; Table 4.3). For the purposes of international comparisons, we are using the race category and not ethnicity for transparency in comparison with Australian information. Note that the Australian census does not have the category of black defined as African American. Instead, in this summary, Aboriginal and Torres Strait Islanders are included in the all other/mixed category. The ethnic category of Hispanic is not included in Australia census calculations, and the participants do not define themselves that way.

Another comparison in this age range is the Preschool Language Scales (Zimmerman, Steiner, & Pond, 1992), a test that does not break out norms in monthly levels in this age range; for the 2- to 2.5-year age interval, 96 children (not broken out by gender) compose the reference sample; in this same age interval, the CDI has 619 children, broken out by gender. Thus the number of participants in the normative groups for the CDI exceeds other widely used instruments in this age range. Reliability for the CDI is high (.95, .96, and .96 internal consistency for vocabulary scales; test–retest correlation for vocabulary production is .95). The standard error (*SE*) of the mean for words produced at 24 months is relatively low (14.7), and the difference between the group means of interest in our study is almost double that value (reported below), thereby well outside the expected error range. Validity values are also high for the CDI in comparison to other language assessments for this age.

With the permission of the authors, 24 Standard American English vocabulary items were replaced with Standard Australian English vocabulary items (e.g., *nappy* for *diaper, footpath* for *sidewalk*). There were separate questionnaires for first- and second-born twins, and caregivers were instructed to complete the questionnaire for their first-born twin before starting the questionnaire for their second-born twin. Five dependent variables were derived from the CDI-WS for evaluation. Three variables have appeared in previous studies: (a) the total number of words produced (Words Produced); (b) the gender-adjusted percentile for the words produced; and (c) response to a single item asking if a child had begun to combine words yet, scaled as *not yet, sometimes,* or *often,* and recoded as 0 = no, 1= yes (Word Combinations). A fourth variable (described below) tracked use of finiteness grammatical markers (Mini-Finite). The fifth variable was a yes/no designation as LLE. The criterion used to identify LLE was a CDI-WS Words Produced score at or below the gender-specific 10th percentile, based on the CDI norming study (Fenson et al., 1993). This is also the criterion used by Reilly et al. (2009) to identify LLE in a population-based sample of Australian children at 24 months of age.

The fourth variable was developed from the CDI-WS for this study as an estimate of a child's beginning use of finiteness grammatical markers. This variable was motivated by an earlier report of significant heritability influences on the grammar of 6-year-old twins (Bishop, Adams, & Norbury, 2005), as assessed by tasks requiring finiteness marking, that is, past tense and third person singular –s affixation (the tasks were experimental versions of the Test for Early Grammatical Impairment; Rice & Wexler, 2001). For this investigation, eight items were identified in the Complexity portion of the CDI-WS, asking the respondent to choose which of two options showing past tense sounded most like the way the child talked, for example, "I fall down" versus "I fell down." Four of the 8 items used past tense examples; four used examples of omitted or correct use of forms of *be* or *do,* for example, "that my truck" versus "That's my truck." These markers of finiteness have been identified as clinical markers of language impairments in young children, such that use of past tense, *be* and *do* forms are likely to be late appearing in children with language impairments (Hadley & Rice, 1996; Hadley & Short, 2005; Rice & Wexler, 1996; Rice, Wexler, & Cleave, 1995). In this study, these items constitute a Mini-Finiteness index calculated as a percentage of correct finiteness choices by caregiver among the eight items. Because only the more advanced children are beginning to use finiteness markers at 24 months of age, this variable measures the earliest stages of a grammatical requirement that has received extensive investigation at somewhat older ages.

Results

Descriptive and Preparatory Analyses

Data management and the descriptive and preparatory data analysis were performed using SPSS 17.0 and Stata (SPSS Statistics, 17.0, 2008; Stata Corp, 2011). Sample means are reported in Table 3, and model-estimated population means (at age = 24 months) are reported in Table 4 for the full sample. Given previous reports of zygosity and gender differences, we included these variables in the models. Combining Words and LLE variables are binary, and their distribution is binomial, requiring logistic regressions. The other three variables had pronounced right skew. Words Produced showed moderate skew (0.726) and an overall triangular distribution with continuous response over the range of 2 to 669; although technically a count, percentile Words Produced showed high skew (1.09), had more of an exponential distribution, and was semicontinuous (20 discrete values) over the range 5 to 99. Mini Finiteness showed the highest skew (3.17) as well as excess zero values (725 out of 946 twins scoring zero); this variable also had only nine distinct values corresponding to affirmative answers for up to 8 questions.

To examine the statistical consequences of dependencies due to twins nested within families, a multilevel model (Stata's xtmixed with random intercept for family) of Words Produced was fit to obtain an intraclass correlation coefficient (ICC) and associated design effect. This variable was selected for multilevel modeling because with only moderate skew and continuous response over a large range of values, the assumption of normally distributed between and within variances was more closely met than would have been the case with any of the other variables. Between twin pairs (Level 2), variance was estimated at 20,560.29, and within twin pairs, variance was an order of magnitude smaller, estimated at 2,169.70. This gives an ICC of .905, and with $n = 2$, the design effect is 1.91. The relatively small within-twin-pair variance and corresponding very large ICC indicates that co-twins are very similar to each other and that almost all of the variance is between families. A design effect of 1.91 indicates that the power of this sample is equivalent to a simple random sample of approximately half the size (i.e., 473 twins instead of 946 twins).

In response to a reviewer query, we explored possible effects of gestational age differences on Words Produced. We

Table 3. Twin children language phenotypes by zygosity and gender.

Sample means and *SDs*[a]	Words produced	Percentile words produced	Mini-Finite	Combining Words proportion	LLE proportion
Total (*N* = 946)					
M	212.39	26.63	7.36	0.70	0.38
SD	150.84	22.79	17.76		
DZ girls (*n* = 328)					
M	239.09	26.23	9.07	0.80	0.32
SD	144.65	20.86	19.42		
DZ boys (*n* = 298)					
M	209.96	30.55	6.75	0.66	0.33
SD	156.43	25.28	16.76		
MZ girls (*n* = 164)					
M	200.61	22.38	6.94	0.68	0.48
SD	152.13	21.54	16.63		
MZ boys (*n* = 156)					
M	173.24	24.46	5.37	0.63	0.48
SD	141.94	21.99	16.93		

LLE = late language emergence.

[a]Sample *SD*s not adjusted for clustering of twins within families.

used a multilevel model with the dependencies most accurately controlled and examined. Gestational age was not a significant predictor in either linear or quadratic form, and there was no indication that the main zygosity findings of the study are an artifact of gestation time. Therefore, this variable was not used in further modeling of Words Produced (or other dependent outcomes) to obtain model-estimated means. We note that there are few children in this sample with gestational ages less than 30 weeks, probably as a consequence of exclusion of children with neurological conditions known to delay language acquisition, a selection criterion that could make a difference for gestational age effects. More extensive exploration of possible pre- and perinatal influences will be completed in a separate follow-up study.

Turning to simpler single-level regression models, model-estimated means and determination of significant effects of zygosity and gender were obtained from regression models assuming distributions appropriate to each variable (i.e., negative binomial with mean dispersion for the skewed variables, logistic for the yes/no variables) and also using robust *SE*s based on clustering of twins within families (Rabe-Hesketh & Skrondal, 2012). These single-level robust regression models converged more readily and/or offered more appropriate distributions, such as the negative binomial. In all models, the Zygosity × Gender interaction was

Table 4. Twin children language phenotypes by zygosity and gender.

Model estimated	Words produced	Percentile words produced	Mini-Finite	Combining Words proportion	LLE proportion
Regression coefficients[a]					
MZ (*SE*)[b]	0.83* (0.06)	0.83 (0.07)	0.82 (0.19)	0.70 (0.14)	1.93** (0.37)
Male (*SE*)[b]	0.85** (0.05)	1.15* (0.08)	0.71* (0.12)	0.56*** (0.09)	1.00 (0.16)
Age (*SE*)[b]	1.12*** (0.03)	0.96 (0.03)	1.4*** (0.08)	1.33** (0.14)	1.08 (0.08)
Zygosity margins[c]					
DZ (*SE*)[b]	201.80 (9.02)	29.21 (1.51)	0.05 (0.01)	0.69 (0.03)	0.31 (0.03)
MZ (*SE*)[b]	167.90 (11.14)	24.16 (1.89)	0.04 (0.01)	0.60 (0.04)	0.47 (0.04)
Difference	–33.85**	–5.05*	–0.01	–0.08+	0.16***
Estimated *SE*	–2.69	–2.30	–0.85	–1.76	3.45
Cohen's *d*	0.18	0.15	0.06	0.12	0.23
Gender margins[c]					
Female (*SE*)[b]	200.30 (9.52)	24.75 (1.47)	0.06 (0.01)	0.71 (0.03)	0.38 (0.03)
Male (*SE*)[b]	169.20 (9.55)	28.51 (1.75)	0.04 (0.01)	0.58 (0.04)	0.39 (0.03)
Difference	–31.05**	3.76*	–0.02+	–0.13***	0.00
Estimated *SE*	–3.06	2.11	–1.91	–3.48	0.01
Cohen's *d*	0.20	0.14	0.12	0.23	0.00

[a]Exponentiated coefficients. Models 1–3 used negative binominal regression, and coefficients are incidence rate ratios. Models 4–5 used logistic regression, and coefficients are odds ratios. [b]Robust *SE* accounting for clustering of twins into 473 families. [c]Marginal means computed as balanced at age = 24 months.

+$p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

not significant, and models were rerun with only main effects of zygosity and gender. All models controlled for age in months centered at 24 months. No significant differences between first- and second-born twins were detected, and this variable was not considered further.

Sample Means and Variances

Table 3 reports the means and *SD*s per phenotype, collapsed across all twin pairs and broken out by zygosity and gender. A general twinning effect is apparent for three of the phenotypes. The percentile Words Produced variable provides a benchmark for normative expectations, which would be a mean near the 50th percentile. Instead, the total mean percentile for Words Produced is 26.63. It is also clear that the children were just beginning to combine words, with very limited use of finiteness markers. We can compare the outcomes for this twin sample to outcomes for a population-based sample of Australian children in the Early Language in Victoria Study (ELVS; Reilly et al., 2009). In the ELVS study, the CDI-WS (with 24 Standard Australian English substitutions) was administered to 1,691 twenty-four-month-old children. At 24 months, 33 children (19.7%) were identified as less than or equal to the gender-specific 10th percentile for expressive vocabulary. In comparison, the present twin sample yielded 358 children (38%) in the bottom 10%. Another benchmark for comparison is the proportion of twin children who were not combining words, 71%, compared to 17% of the ELVS sample.

Modeled Outcomes

The regression models reported in Table 4 revealed an advantage for DZ twins compared with MZ twins, and girls compared with boys. However, there was no interaction of zygosity and gender. The means in Table 4 differ from the values in Table 3 in that these are model-based estimates of population values. The models provide regression coefficients (tabulated in the upper portion of Table 4) for the simple effect of zygosity, gender, and age in comparison with a female dizygotic 24-month-old reference. In addition, the models afford marginal estimates for zygosities (controlling for gender) and for genders (controlling for zygosity), and in both cases controlling for age (i.e., centered at 24 months). These marginal means and their *SE*s are tabulated in the "margins" section of Table 4. Finally, modeling permits post hoc tests of zygosity and gender differences based on the robust *SE* that are adjusted for dependency due to twin pairs. These contrasts, associated critical ratios, effect sizes, and significance levels are tabulated in Table 4 in the sections for zygosity and gender margins.

The estimated means for four of the five phenotypes were lower for the MZ twins, with statistically significant disadvantages for the Words Produced (odds ratio [OR] = 0.832, $p < .01$), the Percentile of Words Produced (OR = 0.827, $p < .05$), for having LLE (OR = 1.934, $p < .001$), and a consistent trend toward disadvantage for Word Combinations (OR = 0.703, $p = .073$). The Mini-Finiteness variable did not yield significant differences for zygosity, possibly because of a floor effect such that few of the children were consistently using finiteness markers. The DZ outcome is consistent with previous reports (Dale et al., 1998, 2000) of a DZ advantage on similar questionnaire assessments at 24 months. The lower average performance of children who are in MZ pairs leads to a higher proportion of children in the MZ group who meet criterion as LLE: 46.5% of the MZ group versus 31.0% of the DZ group, difference = 15.5%, Cohen's $d = 0.23$. Effect sizes for other phenotypes with significant zygosity effects, based on contrasts of model-estimated means and robust SEs, are $d = 0.18$ for number of Words Produced, $d = 0.15$ for percentile Words Produced, and $d = 0.12$ for Combining Words.

Effects of gender were generally as strong or slightly stronger than zygosity effects, with boys at significant disadvantage compared with girls. ORs, model-estimated means, robust SEs, and Cohen's *d* measure of effect size for gender are also shown in Table 4. The relatively higher performance of boys on percentile Words Produced (as compared to poorer performance of boys on Words Produced) suggests the norming process may have overcompensated for maleness. Gender did not predict LLE; boys and girls did not differ, in contrast to the gender effect in single-born samples. This could be due to the general twinning effect that shifts the full distribution, boys and girls, to a lower level of performance that evens out the gender differences, or to effects of the normative sample's calculation of percentiles for boys.

Heritability Analyses

Subsequent to the examination of group differences with robust regression models, two continuous phenotypes (Words Produced and Mini-Finiteness) and two categorical variables (LLE and Word Combinations) were evaluated for heritability. Preliminary correlational analyses revealed that both quantitative variables were correlated (Pearson's $r = .59$). The dichotomous traits (LLE and Word Combinations) are also significantly correlated (Spearman's $r = -0.61$). These four variables are also significantly correlated among themselves (range, 0.36–0.80). Age and gender show only small ($r < 0.23$ for age and $r < 0.12$ for gender), and sometimes nonsignificant, correlations with these phenotypes. For heritability analyses, the two quantitative variables were age regressed and rank normalized (Blom method) to eliminate deviations from normality that can affect the estimation of parameters and confidence intervals (CIs). In addition, for DeFries-Fulker regression analysis, the variables were transformed as suggested by DeFries and Fulker (1985), using the current twin sample mean as population mean, and the MZ and DZ proband means, to produce direct estimates of the parameters of interest. These parameters will yield estimates for genetic and environmental effects after taking into account any general effect of twinning on language measures. Moreover, age and gender were either regressed out of the analysis, or controlled for, depending on the phenotype and the type of analysis. For example, in specific sex-limitation models, gender is included as a covariate to evaluate its effect in model fitting and parameter estimates. It is important to note that the effect of gender was similar across zygosity (i.e., no interaction in Table 4).

Models were calculated to estimate the heritability (h^2), the common-environment (c^2), and the unique-environment (e^2), which are, respectively, the proportion of the variability in a trait that is due to genes, the environmental effects common to the twins in each pair (making them alike), and the environmental factors unique to each twin (making them different). The unique environment estimate, by design, also includes error of measurement. Such biometrical models are widely used in psychology and psychiatry (Neale & Cardon, 1992).

For the quantitative variables, we performed DeFries-Fulker regression analysis (DeFries & Fulker, 1985, 1988; Purcell & Sham, 2003) to estimate heritability and environmental effects of group deficits in a sample selected so that probands were at least 1 *SD* below the sample mean. We also carried out a variance-component analysis in the full sample of twins to estimate genetic and environmental factors that affect variability throughout the total range. Data standardization was performed with specific awk scripts, and DF analysis was carried out with Mx software (Neale, Boker, Xie, & Maes, 2003; Purcell & Sham, 2003). Twin data were analyzed using specific sex-limitation models that evaluate potential gender differences in genetic effects.

Twin pair correlations give a first glimpse of the possible heritable nature of a trait. In the sample selected for LLE, for Words Produced, the MZ twin pair correlation is .96, and the DZ correlation is .81. For Mini-Finiteness, the MZ correlation is .88, and the DZ correlation is .71. These twin correlations for both phenotypes are quite high, and more so for MZ than DZ twin pairs, and are therefore suggestive of a heritable trait. Indeed, the parameter estimates obtained with model fitting in this selected sample indicate that group deficits in both traits are significantly heritable. For both variables, gender effects were not significant, and the best fitting model is one that includes all three components of variance, heritability (h^2_g), shared environment (c^2_g), and nonshared environment (e^2_g). Table 5 reports these parameter estimates for the selected sample, with 95% CIs. The heritability of group deficits in both phenotypes is about 42%–44%, whereas shared environment has a stronger influence for Words Produced (58%) than for Mini-Finiteness marking (48%).

Variance component analysis of the full dataset of 313 DZ and 160 MZ pairs was performed using Mx. Gender and age were included as covariates in a sex-limitation model to evaluate whether there are gender differences in the genetic effects. For Words Produced, there were no significant gender differences. Therefore, we report only the overall variance components estimates in Table 6. For Mini-Finiteness, the parameter estimates were significantly different for males and females ($p < .002$), and therefore each gender is presented separately. As in the selected sample analysis, variability in both language measures is significantly influenced by both genetic and shared environment effects. Table 6 includes parameter estimates, with 95% CIs in parentheses.

For the dichotomous traits defined by LLE and Combining Words, we estimated the proband-wise twin concordance rates (Carey, 2003) and also fitted a threshold model (Neale, Eaves, & Kendler, 1994), which aims to explain the phenotype with a model of underlying polygenic effects and environmental factors that influence the liability of being affected. For the dichotomous variable LLE, the MZ concordance rate is 0.93, and the DZ concordance rate is 0.77. These concordance rates are again quite high, higher for MZ than for DZ pairs, and therefore suggestive of heritable effects. Indeed, these concordance rates can be used to provide estimates of the relative influence of genes (h^2; 33%), common environment (c^2; 60%) and unique environment (e^2; 7%). For Combining Words, the MZ concordance rate is 0.94, and the DZ concordance rate is 0.89, providing estimates of 11% (h^2), 83% (c^2), and 6% (e^2). The reason for the high c^2 is the strikingly large DZ correlation (.89) due to 204 DZ concordant pairs and only 33 DZ discordant pairs.

We can also fit a threshold model to the raw data to provide more formal estimates of the variance components. Again, the best model for LLE is one that includes all three components of variance, with estimates similar to those obtained from the concordance rates. For LLE, there were no significant gender differences, not surprisingly, because the correlation with gender was nearly zero, and therefore the same model can be fitted to girls and boys. As reported in Table 6, the resulting parameter estimates were: h^2, 0.23 (0.10, 0.40); c^2, 0.75 (0.59, 0.87); e^2, 0.02. For Combining Words, boys and girls differed in prevalence (i.e., in variance), but nonetheless the variance components parameter estimates could be equated across genders: h^2, 0.22 (0.04, 0.43); c^2, 0.74 (0.54, 0.89); e^2, 0.04.

Table 5. MZ and DZ proband and co-twin means, parameter estimates (with 95% CIs) on heritability and environmental effects, and number of MZ and DZ twin pairs in a sample selected for LLE.

Variable	MZ proband mean	MZ co-twin mean	DZ proband mean	DZ co-twin mean	h^2_g	c^2_g	e^2_g	MZ twin pairs	DZ twin pairs
Words Produced	–1.48	–1.46	–1.56	–1.24	0.42 [0.18, 0.67]	0.58 [0.39, 0.78]	0.01	35	40
Mini-Finiteness	–1.45	–1.34	–1.56	–1.08	0.44 [0.04, 0.84]	0.48 [0.19, 0.76]	0.08	26	48

Note. h^2_g = heritability; c^2_g = shared environment; e^2_g = nonshared environment.

Table 6. Parameter estimates (with 95% CIs) on heritability and environmental effects in the full unselected sample.

Variable	h^2_g	95% CI	c^2_g	95% CI	e^2_g
Words Produced	0.26	0.21, 0.33	0.71	0.65, 0.77	0.03
Mini-Finiteness (boys)	0.52	0.39, 0.66	0.36	0.22, 0.47	0.13
Mini-Finiteness (girls)	0.43	0.28, 0.65	0.50	0.29, 0.65	0.07
LLE	0.23	0.10, 0.40	0.75	0.59, 0.87	0.02
Combining Words	0.22	0.04, 0.43	0.74	0.54, 0.89	0.04

Note. h^2_g = heritability; c^2_g = shared environment; e^2_g = nonshared environment.

Discussion

The language outcomes for 24-month-old twins in this study clearly demonstrate twinning effects in the form of lower mean levels of performance relative to single-born population expectations, and elevated risk for LLE. This held for a vocabulary estimate, Words Produced, and a grammar estimate, Mini-Finiteness, based on caregiver report of the use of finiteness markers in utterances. Overall, the mean percentile of Words Produced was 26.63, considerably below the expected population mean percentile of 50. Defining LLE as performance at the 10th percentile or below for Words Produced, the twin sample yielded 38% of the children as LLE, compared with population estimates of 19.7%, with a higher proportion for MZ twins (48.1%) than DZ twins (32.6%). Further investigation revealed that the twinning effect was more apparent in MZ twins, consistent with other reports in the literature of a twinning cost to children in MZ twin pairs compared with children in DZ twin pairs (Dale et al., 1998, 2000; Lytton, 1980). Similar outcomes were found for a finiteness marker not previously examined in the CDI-WS instrument, the Mini-Finiteness variable. Because no population norms are reported on this subset of the items on the CDI-WS, there is no reference sample estimate of a twinning effect relative to single-born children.

All traits analyzed in this study are significantly heritable. DF analysis suggests that the group deficit of the selected sample on the two continuous language measures (Words Produced and Mini-Finiteness) is heritable, with genes accounting for about 43% (95% CI = 18%, 67% for Words Produced and 4%, 84% for Mini-Finiteness) of the group deficit. Shared environment also significantly affects the group deficit on these language skills. In addition, population variability in these traits is also significantly affected by genetic and common-environment factors. For Words Produced, the difference in heritability of group deficits versus population variability is not significantly different ($p = .22$). For the grammar variable (Mini-Finiteness), this test is not appropriate because there are significant gender differences in the full range of variability, but not in the group deficit. Interestingly, in the full range of ability for grammar (Mini-Finiteness), the heritability in boys is slightly higher than in girls (h^2 of 0.52 vs. 0.43), although the 95% CIs overlap to a large extent (39%, 66% in males and 28%, 65% in females). Finally, the LLE and the Combining Words dichotomous variables also show a significant heritable nature, although its magnitude is smaller than for the quantitative variables. Again, the 95% CIs overlap, so these differences are just a trend, not statistically significant without confirmation in a larger sample.

The heritability estimates for this sample, in the range of .26 to .52 for the continuous traits, in the selected and full samples, are higher than the .25 reported by Dale et al. (2000) and in other investigations of the TEDS sample for words produced at age 24 months. The outcomes here suggest that the emergence of finiteness marking in the grammar of these young children may have higher heritability effects (0.44 [95% CI = 0.04, 0.84] in the selected sample; 0.52 [0.39, 0.66] and 0.43 [0.28, 0.65], for boys and girls, respectively, in the full sample) than are evident for vocabulary acquisition (0.42 [0.18, 0.67] and 0.26 [0.21, 0.33], respectively) suggesting relatively less shared environmental influence on the emergence of finiteness marking in toddlers, although such comparisons are not straightforward. The best estimate of heritability of LLE (defined with a cut-off level of 1 *SD* below the age gender-adjusted mean), 0.23 (95% CI = 0.10, 0.40), is much lower than the estimate of Dale et al. (1998) of 0.73 for a vocabulary measure, based on a cut-off at the 5th percentile, but similar to the estimate of DeThorne et al. (2005) of 0.26 for 4-year-old twins with a cut-off at the 10th percentile.

It is important to note here some considerations regarding the estimates for genetic and environmental effects reported in this study. Whether to analyze raw or transformed data can have implications for results and interpretation (Bishop, 2005). The large deviation from normality of some of the variables, especially the Mini-Finiteness task, most likely due to measurement problems arising from the young age of the children tested, produced uninterpretable parameter estimates. For this reason, we published results of the analysis of normalized variables, which are often used in this field. We also want to emphasize that some of the 95% CIs are large, a limitation possibly due to not having a large enough sample. For this reason, even though the parameter estimates for some analyses and variables differ, their CIs overlap. Therefore, we have mentioned that some point estimates are different, although often they are not statistically significant.

Further interpretation of the relative contributions of effects attributable to inheritance and shared environmental effects can be guided by the differences in language outcomes attributable to zygosity effects. Overall, children in MZ twin pairs were more likely to score lower than children in DZ twin pairs, leading to a higher proportion of MZ twin children in the LLE group.

The emerging picture reveals new factors to consider in the interpretation of twinning effects for language acquisition. This study shows, on average, a twinning effect, which is likely to contribute to a shared environmental influence. Yet the estimates of this shared environmental influence are lower in this sample than in previous reports of twins of the same age on similar measures. Further, the twinning effect is stronger, on average, for children in MZ pairs than children in DZ pairs, although a twinning effect is expected to contribute to similarity, not differences, across zygosity. This outcome is inconsistent with hypotheses that attribute a twinning effect in the form of delays in early language acquisition to a shared postnatal environment in which maternal input to children who are twins is reduced as a result of the additional demands of caring for two toddlers. The lack of support for a limited input interpretation of a twinning effect on language may be reassuring to busy parents who worry about providing sufficient individual attention to each child.

On the other hand, the zygosity effects may be attributable to differences at conception or other prenatal or perinatal environmental effects, or by shared prenatal or perinatal influences that interact with the early biological differences inherent to MZ versus DZ twin pairs. As noted in the participant section, at birth, DZ twin babies are, on average, somewhat longer and heavier than MZ babies, although in this study there were no differences in the other reported perinatal variables. The findings here suggest that zygosity does not interact with gender effects, although both factors warrant further investigation in follow-up studies. It is quite possible that factors such as prematurity and birth complications could disproportionately affect infants in MZ pairs more than infants in DZ pairs and could in turn contribute to how zygosity affects early language acquisition. Recall that in this sample of twins, children with neurological conditions known to influence early language acquisition were excluded from the sample, a selection criterion that may contribute to different outcomes across different samples of twins.

This study follows the precedents of other studies in documentation of the likely heritability of language based on observations of early language acquisition of 24-month-old twins, with increased heritability for children with lower levels of early language acquisition. This study provides a new estimate of the percentage of children who score low relative to their age reference group on a widely used parent questionnaire instrument, and documentation of increased heritability using a reference group–based criterion for LLE. Innovative elements of the study include exploration of an early finiteness marker for LLE, as well as results that suggest the need to study further possible effects of zygosity and gender at the early stages of language acquisition among twins. The findings raise the question of whether these possible early effects persist in the same children over time, a question we plan to pursue. This study adds to our appreciation of how the study of twin children contributes to our understanding of the sources of language acquisition for all children.

Acknowledgments

This work was made possible by grants from the National Institutes of Health (RO1DC05226, P30DC005803, P30HD002528). The second author is supported by a program grant from the National Health and Medical Research Council of Australia (No. 572742). The authors especially thank the children and families who participated in the study and the following members of the research team: Antonietta Grant, Erika Hagemann, Alani Morgan, Virginia Muniandy, Elke Scheepers, and Alicia Watkins. The authors also thank Denise Perpich for data management and data summaries, and the staff at the Western Australian Data Linkage Branch and the Maternal and Child Health Unit.

References

Australian Bureau of Statistics. (2001). *Census of Population and Housing: Selected Social and Housing Characteristics,* Australia, 2001, Cat.No.2015.0.

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development—Third Edition*. San Antonio, TX: Harcourt Assessment.

Bishop, D. V. M. (2005). DeFries-Fulker analysis of twin data with skewed distributions: Cautions and recommendations from a study of children's use of verb inflections. *Behavioral Genetics, 35,* 479–490.

Bishop, D. V. M., Adams, C. V., & Norbury, C. F. (2005). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain, and Behavior, 5,* 158–169.

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

Bracken, B. A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of Psychology, 26,* 155–166.

Carey, G. (2003). *Human genetics for the social sciences*. Thousand Oaks, CA: Sage.

Chauhan, S. P., Scardo, J. A., Hayes, E., Abuhamad, A. Z., & Berghella, V. (2010). Twins: Prevalence, problems, and preterm births. *American Journal of Obstetrics and Gynecology, 203,* 305–315.

Clark, P. M., & Dickman, Z. (1984). Features of interaction in infant twins. *Acta Geneticae Medicae et Gemellologiae, 33,* 165–171.

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

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

Dale, P. S., Simonoff, E., Bishop, D. V. M., Eley, T. C., Oliver, B., Price, T. S., . . . Plomin, R. (1998). Genetic influence on language delay in two-year-old children. *Nature Neuroscience, 1,* 324–328.

DeFries, J. C., & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavioral Genetics, 15,* 467–473.

DeFries, J. C., & Fulker, D. W. (1988). Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Genetica Medicae Gemellologiae, 37,* 205–216.

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

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E. A., Hartung, J. P., . . . Reilly, J. S. (1993). *MacArthur Communicative Development Inventories*. San Antonio, TX: The Psychological Corporation.

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

Gee, V., & Green, T. J. (2004). *Perinatal statistics in Western Australia, 2003 twenty-first annual report of the Western Australian Midwives' Notification System*. Perth, Western Australia: Department of Health.

Hadley, P. A., & Rice, M. L. (1996). Emergent uses of BE and DO: Evidence from children with specific language impairment. *Language Acquisition, 5,* 209–243.

Hadley, P. A., & Short, H. (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 1344–1362.

Hay, D. A., Prior, M., Collette, S., & Williams, M. (1987). Speech and language development in preschool twins. *Acta Medicae et Gemellologiae Twin Research, 36,* 213–223.

Koeppen-Schomerus, G., Eley, T. C., Wolke, D., Gringras, P., & Plomin, R. (2000). The interaction of prematurity with genetic and environmental influences on cognitive development in twins. *Journal of Pediatrics, 137*(4), 527–533.

Lytton, H. (1980). *Parent-child interaction: The socialization process observed in twin and singleton families*. New York, NY: Plenum.

Lytton, H., Conway, D., & Sauve, R. (1977). The impact of twinship on parent-child interaction. *Journal of Personality and Social Psychology, 35,* 97–107.

Maloney, E. S., & Larrivee, L. S. (2007). Limitations of age-equivalent scores in reporting the results of norm-referenced tests. *Contemporary Issues in Communication Science and Disorders, 34,* 86–93.

McCarthy, D. A. (1972). *The McCarthy Scales of Children's Abilities*. New York, NY: The Psychological Corporation/Harcourt Brace Jovanovich.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling*. Richmond: Virginia Institute for Psychiatric and Behavioral Genetics.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, the Netherlands: Kluwer Academic.

Neale, M. C., Eaves, L. I., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavioral Genetics, 33*(3), 271–278.

Pinborg, A. (2005). IVF/ICSI twin pregnancies: Risks and prevention. *Human Reproduction Update, 11*(6), 575–593.

Purcell, S., & Sham, P. C. (2003). A model-fitting implementation of the DeFries-Fulker model for selected twin data. *Behavioral Genetics, 33,* 271–278.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata: Categorical responses, counts, and survival* (Vol. 2, 3rd ed.). College Station, TX: Stata Press.

Reilly, S., Bavin, E. L., Bretherton, L., Conway, L., Eadie, P., Cini, E., . . . Wake, M. (2009). The Early Language in Victoria Study (ELVS): A prospective, longitudinal study of communication skills and expressive vocabulary development at 8, 12 and 24 months. *International Journal of Speech-Language Pathology, 11,* 344–357.

Reznick, J. S., Corley, R., & Robinson, J. (1997). A longitudinal twin study of intelligence in the second year. *Monographs of the Society for Research in Child Development, 62*(1), 1–154.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39,* 1239–1257.

Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: 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, Language, and Hearing Research, 38,* 850–863.

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. *Journal of Child Psychology and Psychiatry, 44,* 326–341.

Savic, S. (1980). *How twins learn to talk*. New York, NY: Academic Press.

Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Language, 77,* 647–723.

Stromswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition, 101,* 333–384.

Thorpe, K. (2006). Twin children's language development. *Early Human Development, 82,* 387–395.

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

Tomasello, M., Mannie, S., & Kruger, A. (1986). Linguistic environment of 1- to 2-year-old twins. *Developmental Psychology, 22,* 169–176.

Tomblin, J. B., & Buckwalter, P. (1998). Heritability of poor language achievement among twins. *Journal of Speech, Language, and Hearing Research, 41,* 188–199.

Zimmerman, L. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool Language Scale—3*. San Antonio, TX: The Psychological Corporation.

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