Outcomes of Early Language Delay: II. Etiology of Transient and Persistent Language Difficulties

Dorothy V. M. Bishop
Thomas S. Price
University of Oxford,
Oxford, U.K.

Philip S. Dale
University of Missouri–
Columbia

Robert Plomin
Institute of Psychiatry,
London

Genes are known to play an important role in causing specific language impairment, but it is unclear how far a similar etiology is implicated in transient language delay in early childhood. Two-year-old children with vocabulary scores below the 10th centile were selected from a cohort of over 2,800 same-sex twin pairs whose language was assessed by parental report at 2, 3, and 4 years of age. These children with early language delay (ELD) were divided into cases of transient and persistent language difficulties on the basis of outcome at 3 and 4 years. A DeFries–Fulker analysis (J. C. DeFries & D. W. Fulker, 1985) was used to compute group heritability (h^2_g) of 2-year vocabulary delay separately for those with transient and persistent difficulties. When 3-year and 4-year language attainments were used to categorize outcomes, h^2_g was similar and modest (.25 or less) for both transient and persistent difficulties. However, when persistent difficulties were defined according to whether parents expressed concern about language at 3 years or according to whether a professional had been consulted about language difficulties at 4 years, heritability was significantly higher. For 289 children with no professional involvement at 4 years, heritability of 2-year vocabulary delay was close to zero, whereas for 134 children with professional involvement, a significant h^2_g of .41 (*SE* = .127) was found. Early language delay appears largely environmental in origin for 2-year-olds whose parents do not go on to seek professional help.

KEY WORDS: language delay, preschoolers, genetic influence, longitudinal outcome

Over the past decade, there have been major advances in our understanding of the etiology of specific language impairment (SLI), with growing recognition that genes play an important role. Three twin studies have found substantially higher concordance for SLI in monozygotic (MZ) compared to dizygotic (DZ) twins (Bishop, North, & Donlan, 1995; Lewis & Thompson, 1992; Tomblin & Buckwalter, 1998), giving heritability estimates of .45 or above, suggesting that shared genes rather than shared environment are principally responsible for the similarities between twins. However, heritability estimates can vary substantially depending on how SLI is defined (Bishop, 1994).

The importance of phenotype definition is re-emphasized by contrasting recent findings from molecular genetic studies. Lai, Fisher, Hurst, Vargha-Khadem, and Monaco (2001) studied a three-generation

family, half of whom had a severe speech and language disorder, and found that all the affected family members had a point mutation of a gene on Chromosome 7. However, the phenotype in this family was both severe and distinctive, with major impairments in the areas of syntax, phonological processing, and orofacial praxis (Watkins, Dronkers, & Vargha-Khadem, 2002). The same team (SLI Consortium, 2002) subsequently studied 98 other families with an affected child, defining the phenotype in terms of scores being more than 1.5 *SD*s below the mean for age on receptive and/or expressive language measures, with a performance IQ of 80 or above. In this sample there was no linkage to Chromosome 7, although two other sites, on Chromosomes 16 and 19, were identified as quantitative trait loci for SLI.

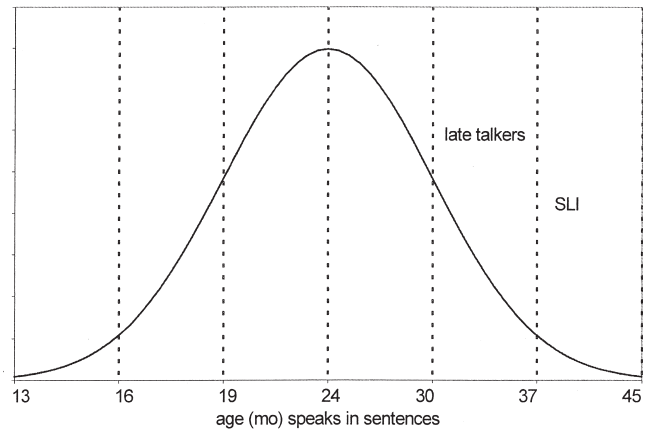
If we are to make progress in uncovering the genetic basis of SLI, we need a better specification of which subtypes of impairment are related to genetic factors. In particular we need to know whether heritable language impairments differ from nonheritable impairments in terms of severity, nature of the language difficulties, or the age at which they are identified. In this article we compare the heritability of early language delay for two groups of 2-year-old children: those who have continuing evidence of language difficulties at 3 or 4 years of age and those who appear to grow out of their difficulties.

Depending on their conceptualization of SLI, researchers have adopted contrasting views of the relationship between transient and persistent language difficulties. One view is that SLI is a qualitatively distinct disorder resulting from a genetic deficiency affecting a specialized system for learning syntax (e.g., Gopnik & Crago, 1991; Van der Lely, Rosen, & McClelland, 1998). The linguistic deficiencies are seen as long-lasting, so that even when apparently normal syntax is observed, it is generated by explicit application of learned rules, rather than by implicit syntactic knowledge (Gopnik, 1999). In this view, transient language delay is distinct from SLI and is likely to have different causes. Nevertheless, a case of SLI will be difficult to distinguish from a "late bloomer" at the outset of language acquisition, because both will have delayed language milestones and will continue to use a small vocabulary of single words at a time when other children are putting words together (Haynes & Naidoo, 1991).

An alternative perspective is to view transient and persistent language difficulties as points on a continuum of severity, with SLI as the tail of a normal distribution of language ability. This view was cogently stated by Leonard (1987), who proposed that it was mistaken to regard SLI as a distinct pathological condition. Language ability, like many other human traits, shows variation in the population and may be regarded as continuously distributed. Those at the lower end of this continuum will be disadvantaged and it may be appropriate to offer special services to help boost their language skills. However, according to Leonard, differences between children labeled as having SLI and those with skills at the low end of the normal range are quantitative, not qualitative (see also Dale & Cole, 1991). Leonard's position was adopted by Rescorla, Dahlsgaard, and Roberts (2000), who argued that children vary not just in their level of language ability, but also in the age at which language milestones are passed. They suggested that late talkers who subsequently recover occupy an intermediate position on a continuum that has SLI at one extreme and normal development at the other (see Figure 1). In this view, the language deficits of late talkers are caused by the same factors and are similar in kind to those seen in SLI, but are milder in form.

The theoretical contrast between these positions highlights the question of whether there is a categorical distinction between cases of maturational lag and cases of persistent language impairment. Those who argue for such a divide typically have done so on the basis of phenotypic characteristics, in which grammatical impairments are particularly pronounced, with the language profile of typical SLI differing from that of normally developing children at any age. Often implicit in such accounts is the idea that an abnormal profile of language skills can be used as a marker for long-term language disorder rather than transient language delay. However, this conceptualization is not universally accepted. For instance, Rice, Wexler, and Hershberger (1998) discussed an account of SLI in which a specific module of the developing grammatical system, which controls use of the infinitive, is late to mature. According to this account,

Figure 1. Continuum model of language impairment, which regards persistent SLI and transient language delay as differing only quantitatively. Estimates of age at speaking in sentences were made by fitting a logarithmic function to percentile values from Neligan and Prudham (1969).



the pattern of infinitive usage by children with SLI will resemble that of younger typically developing children and will change over time. At particular stages of development, children with SLI will demonstrate distinctive domain-specific language difficulties. However, this is seen as due to slow maturation rather than to a damaged language module. This, then, is a maturational account that nevertheless predicts a distinct profile of grammatical impairment in SLI.

It is, clearly, difficult to decide if there is a difference in the underlying mechanisms causing persistent versus transient language difficulty purely on the basis of the phenotype. An alternative source of evidence, however, is available from etiology. If transient language delay were associated with one set of risk factors, and persistent language impairment with another, then this would provide support for the notion that these conditions differ qualitatively, not just in terms of severity. As noted above, studies with severe and persistent language impairment in older children and adults have found ample evidence of strong genetic influence. It is therefore of interest to know whether transient language delay is equally heritable. Such a finding would not prove it is the same condition, as different genes could be implicated in transient versus persistent language impairment. However, if we were to find lower levels of heritability for transient delay than for persistent impairment, this would add weight to the notion that transient and persistent difficulties are distinct conditions.

To date, the data on this issue are mixed. Some circumstantial evidence is provided by studies of familiality of language impairment in late talkers. Any group of late talkers will contain a mixture of children who will prove to be late bloomers and others who will turn out to have persistent language impairments. However, in studies that have recruited late-talking 2-year-olds, the majority do improve spontaneously, to score within normal limits on language measures a few years later (Paul, 2000). Thus, a sample selected for small vocabulary at 2 years of age is likely to consist predominantly of cases of transient language impairment. Paul (1991) reported that family history of language, speech, and learning problems was three to four times more common in children with early language delay than in controls matched on social background; this figure is comparable to those reported in family studies of older children with SLI (Stromswold, 1998). Similarly, Dale et al. (1998) found substantial heritability for delayed vocabulary development in 2-year-old twins. These two studies offer indirect evidence that transient delays, like more persistent SLI, may be under strong genetic influence.

However, a rather different picture is given by two studies, from the United States and Finland, that adopted a longitudinal perspective, contrasting etiological factors in late talkers whose difficulties resolved with those whose problems persisted. In the United States, Whitehurst et al. (1991) found that a sample of 2-year-olds with very limited expressive vocabulary did not differ from a matched control group with respect to family history of language problems. The discrepancy with findings of Paul (1991) may reflect the fact that Whitehurst et al. selected their sample to exclude children with receptive language impairment. In a follow-up of the same sample, the researchers considered whether otitis media might act as a nongenetic risk factor for early language delay. They found that children with early expressive language delay did not differ from controls in the frequency, duration, or timing of episodes of otitis media, but history of middle ear disease between 12 and 18 months of age predicted expressive language improvement in the language delayed group (Lonigan, Fischel, Whitehurst, Arnold, & Valdez-Menchaca, 1992). They concluded that the population of children with language delay may contain a mixture of two etiological groups. The first contains those whose language delay arises from early middle ear disease, who will recover once the transient effects on hearing have resolved. The second group contains children who suffer early language delay for no known reason, in whom the problems are more persistent.

In the Finnish study, Lyytinen, Poikkeus, Laakso, Eklund, and Lyytinen (2001) compared language outcomes of children with a family history of dyslexia to those of a control group. In both samples, some children were identified as late talkers at 2 years of age, but the outcome at 3 years 6 months (3;6) differed depending on family history. The problems of late talkers with no family history had completely resolved by age 3;6, whereas late talkers with a family history of dyslexia had persisting language difficulties. Thus, although the two longitudinal studies differed substantially in methodology and findings, they both suggested that a sample of children with early language delay contains a mixture of cases with different etiologies, and that these etiologies relate to outcome.

In this article we address the question of whether there are any etiological differences between children whose early delays resolve and those who have longer term problems, using data from a large sample of twin pairs. These include the 1994 cohort reported by Dale et al. (1998), as well an additional cohort born in 1995, who were assessed using the same methods. Language development in these children was studied using parental report measures at ages 2, 3, and 4 years. In our companion article in this issue (Dale, Price, Bishop, & Plomin, 2003), we described outcomes at 3 and 4 years for children who had vocabulary scores below the 10th centile at age 2 years. Consistent with previous literature, we found that a substantial proportion of these children no longer appeared language impaired when assessed a year or two later. In

the companion article, we looked for phenotypic differences between language delayed 2-year-olds in relation to their outcome at 3 or 4 years and found these were slight and not adequate for predicting outcome with any degree of accuracy. In this article we again focus on children who had evidence of early language delay at 2 years, this time contrasting those with transient and persistent problems in terms of etiology.

We identified language delay at 2 years (ELD) using parental report of expressive vocabulary size. We focused on vocabulary because grammatical measures are insensitive at this age (most late talkers are not combining words), and parental report of comprehension is likely to be unreliable by this stage, when receptive vocabulary has become very large.

Three sets of heritability analyses were carried out. First, we estimated heritability of 2 year language delay, without regard to outcome at 3 and 4 years. This was a replication of analyses previously conducted on a subset of these children by Dale et al. (1998). We used a quantitative method of genetic analysis known as DeFries–Fulker (DF) analysis (DeFries & Fulker, 1985), which considers whether the degree of similarity between twins is affected by the genetic relationship between them. The rationale is that if genes are important in causing disorder, then identical (MZ) twins, who share all their genes, should be more similar in language status than nonidentical (DZ) twins, who share on average half their segregating genes. In the second set of analyses, we subdivided the ELD group in terms of language outcomes at 3 and 4 years of age, to see whether genetic influence on 2-year-old vocabulary deficit differed for those with persisting versus transient difficulties. Outcome at 3 and 4 years was classified in three different ways: (a) according to parental report of language attainments, (b) according to parental expression of concern, and (c) in terms of whether parents had sought professional help for speech and language difficulties. These analyses suggested that heritability of 2-year language delay did differ depending on outcome, and so we conducted a third set of analyses to test for the significance of differences in heritability between subgroups. To do this we used an augmented form of DF analysis, which is explained in more detail below. Six of these augmented analyses were done: one for each of the three outcome measures at both ages (3 and 4 years).

Method

Participants

Full details of methods are given in our companion article. In brief, the core sample consisted of 5,208 twin pairs born in 1994 or 1995, all of whom had parental report data available for 2 years of age and for 3 and/or 4 years. A parent-report measure was used to assign twin zygosity (Price et al., 2000), followed up with DNA testing in uncertain cases. The analyses reported here focus on 187 MZ twin pairs and 169 same-sex DZ twin pairs where one or both twins met our criteria for ELD (see below). As shown in Table 1, follow-up data were available at both 3 and 4 years for 62% of this sample, at 3 years only for 27% of the sample, and at 4 years only for 10%. The 1,392 opposite-sex DZ twin pairs included in our companion article are excluded here, because inclusion of opposite-sex pairs complicates behavior genetic analyses, especially when there are sex differences in rates of impairment.

The 1994 cohort overlaps with that reported by Dale et al. (1998), but is not identical. Of the 3,039 twin pairs

Table 1. Number of twin pairs included in the analysis, and details of those excluded.

Sample	Cohort		
	1994	1995	Total
Included in analysis: one or both met criteria for ELD at 2 yr	190	166	356
With language data at 2, 3, and 4 yr	131	89	220
Data at 2 and 3 yr only	38	61	99
Data at 2 and 4 yr only	21	16	37
MZ	110	77	187
DZ same sex	80	89	169
Excluded from analysis			
Non-English background, medical factors, unclear zygosity	466	549	1,015
Opposite sex	766	626	1,392
Same sex, neither twin met criteria for ELD at 2 years	1,443	1,002	2,445
Total sample (excluded and included)	2,865	2,343	5,208

Note. ELD = early language delay; MZ = monozygotic; DZ = dizygotic.

described in that report, 660 were excluded in the current report because of missing or late data at 3 and 4 years, and a further 38 were excluded because of medical conditions or syndromes such as autism that were discovered only after the age of 2 years. The current sample included an additional 58 pairs born in 1994 whose zygosity could not be assigned at the time of the previous report.

Measures

At 2, 3, and 4 years of age, parents completed the age-appropriate version of the MacArthur Communicative Development Inventory UK Short Form (MCDI: UKSF; Dionne, Dale, Boivin, & Plomin, 2003). Three measures obtained from the MCDI:UKSF were used to quantify language outcomes at 3 and 4 years of age. The first was a measure of expressive vocabulary. The second was a grammar rating, based on 12 forced choice items at 3 years and a single scale at 4 years. The final measure was a set of items used to evaluate the child's use of abstract language. Two additional, more global outcome measures were obtained. First, parents were asked whether they had any concern about their child's language development—those who reported slow language development or comprehension problems in their child were coded as 1, the remainder as 0. Parents were also asked if they had consulted a professional (speech-language pathologist, pediatrician, etc.) because of concern about their child's speech or language development. Data on reliability and validity of language measures are presented in the companion article.

Socioeconomic status was assessed in terms of mother's educational qualifications on an 8-point scale (1 = *no qualifications*, 2 = *below standard for a pass on the school-leaving examination*, 3 = *O-levels* (passing score on school-leaving examination at age 16), 4 = *A-levels* (age 18 exam, generally required for university entrance), 5 and 6 = *tertiary vocational qualifications*, 7 = *an undergraduate degree*, and 8 = *a postgraduate degree*). Middle ear status was assessed at each age using a brief questionnaire asking about relevant symptoms at 18 months (see Rovers, Haggard, Gannon, Koeppen-Schomerus, & Plomin, 2002). In this article we focus on the seven items used at 2 years, namely catarrh during cold, hearing difficulty, mouth-breathing, snoring, leaking ears, pulling/scratching ears, and red/sore ears. Parents rated each symptom as 1 = *often*, 2 = *sometimes*, 3 = *occasionally*, and 4 = *never*, and these ratings were summed.

Categorization of Language Status

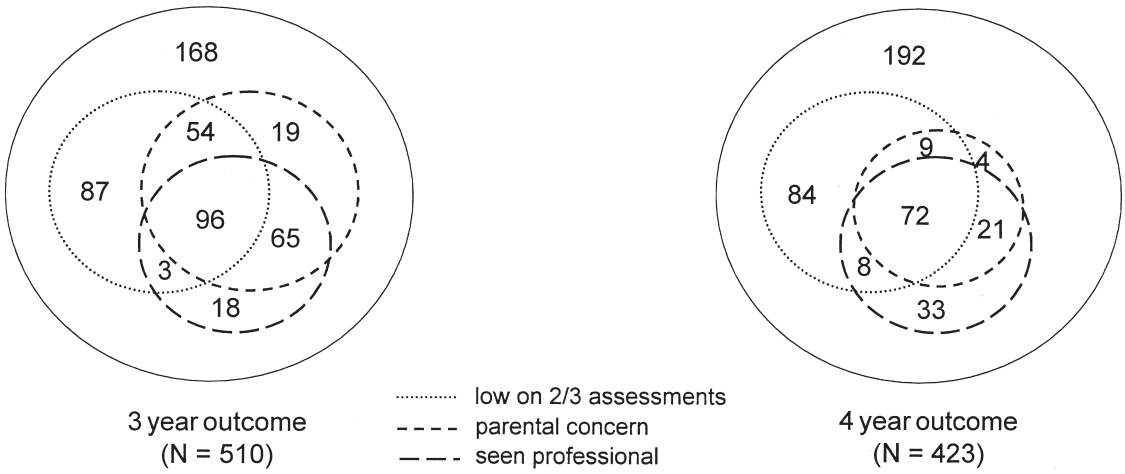
At 2 years of age, children were categorized as having ELD if their vocabulary score was below the 10th centile (15 words or less), using as the normative base the whole twin sample who did not have any exclusionary factors (i.e., medical problems or non-English background). We used the 10% cutoff rather than the 5% cutoff previously used by Dale et al. (1998) in order to have sufficient numbers for assessment of heritability in subgroups. Our cutoff score is lower than the 10th centile figure of 30 words computed from U.S. norms (Fenson et al., 2000) on the MacArthur Communicative Development Inventory Short Form (MCDI:SF). There are two possible reasons for this. First, early language development is often a little slower in twins than in single-born children (Rutter & Redshaw, 1991). Second, early vocabulary development, as assessed by parental report, is slower in U.K. than in U.S. samples (Hamilton, Plunkett, & Schafer, 2000).

At 3 and 4 years of age, children were coded as having language difficulties if their language attainments fell below the 15th centile on at least two out of three of the parent report measures (vocabulary, grammar, and abstract language); this cutoff was chosen because it selected about 10% of the sample (see our companion article). In separate analyses, language outcomes were categorized in terms of whether or not the child's parent reported concern about language development and whether professional advice had been sought for speech and/or language difficulties. These different methods of classifying outcome identified partially overlapping samples of children. Figure 2 shows outcomes for 2-year-old children with ELD according to each of the three criteria, at 3 and 4 years. It can be seen that 168 of 510 children (33%) were not impaired by any of the three criteria at 3 years of age, and 192 of 423 children (45%) were not impaired by any of the criteria at 4 years of age. The next most common outcome at 3 years was to be impaired on all three criteria—language ratings, parental concern, and professional involvement. However, overlap between the outcome categories at this age was relatively low. By 4 years of age, there was greater overlap between parental concern and professional involvement, but it is noteworthy that a substantial number of children (84 of 423) did not arouse parental or professional concerns, despite parental report of low language attainments.

Analytic Approach

Resemblance between members of a twin pair can arise because of genetic similarity or because of environmental influences shared by the two twins, either pre- or postnatally. Differences between twins reflect unique, nonshared environmental effects. The goal of behavior genetic analysis is to quantify the relative contribution of genes, shared environment, and unique environment to a trait. Twins provide a useful natural experiment, because it is possible to contrast MZ twins,

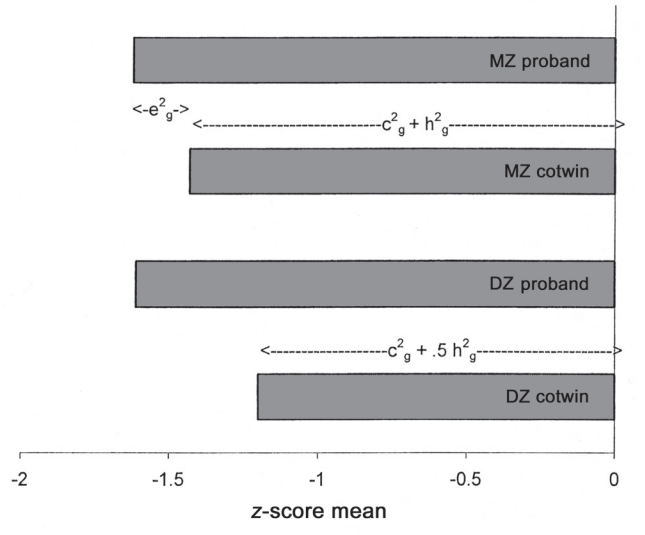
Figure 2. Venn diagrams showing numbers of children in different language outcome groups at age 3 and 4 years. The outmost circle includes all children with early language delay who had relevant follow-up data; those falling within the inner circles have language impairment at age 3 or 4 years according to the different definitions shown in the legend.



who share all their genes, with DZ twins, who share half their segregating genes on average. If MZ twins resemble one another more closely than DZ twins, this provides evidence of genetic etiology.

In this article, our focus is on estimating heritability of low vocabulary at 2 years of age. Although we use data from follow-up assessments at 3 and 4 years to classify children's outcomes, the dependent measure in all the genetic analyses is the 2-year vocabulary score. We used DF analysis, which is a method for analyzing twin data when the condition of interest is defined in terms of an extreme score on a quantitative scale. For this procedure, one first identifies individuals who are affected cases. These individuals, known as probands, are defined in terms of scores below some specified cutoff on the dimension of interest. Our probands were those meeting criteria for ELD (i.e., with 2-year-old vocabulary scores of 15 or less). The extent of genetic influence on ELD can then be determined by considering the scores of co-twins of the probands in relation to zygosity. Consider the situation depicted in Figure 3. The MZ and DZ probands have been selected because their *z* scores fall below a cutoff of 1 *SD* below the population mean of zero. The means of their co-twins will depend on the etiological factors involved in causing impairment. If impairment were caused solely by random events specific to the individual (nonshared environment), then the expectation would be that the co-twins' scores would be at the population mean (zero), with no impairment. If impairment is influenced by environmental factors that are common to both twins, then the co-twins should resemble the probands and also have low *z* scores, irrespective of zygosity. If, however, genes are implicated, MZ twins should resemble one another more closely than DZ twins, as shown in the figure. As DeFries and Fulker (1985) noted, with significant genetic influence the prediction is that the average DZ co-twin score will regress further to the population mean than the average MZ co-twin score. One can get estimates of the relative importance of genes (h^2_g), shared environment (c^2_g), and nonshared environment (e^2_g) from the means shown in Figure 3, if the data are rescaled so that the proband mean is 1.0. MZ twins share all their genes and many environmental experiences: The difference between MZ

Figure 3. Illustrative diagram showing how means for MZ and DZ probands and co-twins can be used to estimate relative contribution of genes (h^2_g), shared environment (c^2_g), and nonshared environment (e^2_g) to a condition defined in terms of a score below cutoff on a continuum.



probands and their co-twins thus provides an estimate of e^2_g. Similarity between MZ probands and co-twins reflects influence of genes (100% in common) and shared environment, whereas in DZ twins, similarity between reflects influence of 50% genetic similarity plus shared environment. Thus, by doubling the difference in means for MZ and DZ co-twins one obtains an estimate of h^2_g. Because the contributions of h^2_g, c^2_g, and e^2_g sum to 1.0, c^2_g can be obtained by subtraction. Furthermore, this logic can be implemented in a multiple regression analysis in which the scores of co-twins are the dependent variable, with the scores of probands and the degree of genetic relationship, *R* (1.0 for MZ and 0.5 for DZ), as predictors. When data are appropriately transformed, the regression coefficient for the genetic relationship term can be interpreted directly as indicating group heritability, h^2_g, an index of the extent to which genes cause impairment. When using the DF method, the question arises as to how to categorize children when both twins meet criteria as probands. We followed the customary procedure of double entry, whereby both children are included in the analysis as probands, with the standard error of the heritability estimate then being adjusted to take into account the inflation of degrees of freedom that this entails.

Our main aim was to see whether group heritabilities for late-talkers at 2 years differed for those with transient or persistent difficulties. We know there are strong genetic influences on SLI in older children, and therefore we would expect to find significant heritability in children with persistent difficulties. The question of interest is whether the same picture is seen in those whose ELD resolves in the preschool period. To address this, we divided our sample into outcome subgroups and estimated h^2_g separately for each of these. However, we cannot tell whether estimates of h^2_g are significantly different for the outcome groups, because they are not independent (i.e., pairs where one twin is classified as persistent and one as transient will be represented in both subgroups). To do this, we used an extension of DF analysis (see Castles, Datta, Gayan, & Olson, 1999). As with standard DF analysis, the problem is recast in terms of multiple regression, with the goal being to predict the co-twin's 2-year vocabulary score (C) from the 2-year vocabulary score of the proband (P) and the coefficient of relationship (R; .5 for DZ and 1.0 for MZ). In addition, a term corresponding to outcome status (S: 1 for transient difficulties, 0 for persisting difficulties) is incorporated in the regression, to give the equation:

$$C = b_1.P + b_2.R + b_3.S + b_4.P.S + b_5.R.S.$$

In this equation, b_1 through b_5 are regression coefficients. If the regression coefficient for the R.S term, b_5, is significant, this indicates there is a reliable interaction, that is, h^2_g depends on the level of S.

This augmented analysis was conducted separately using each of our three different methods of assessing outcome: (a) ratings of language attainment, (b) parental concern, and (c) professional involvement, first for 3-year outcomes and then for 4-year outcomes. Thus, we ran a total of six augmented DF analyses on 2-year vocabulary scores, using outcome at 3 or 4 years as the interaction term.

Results

Heritability of ELD

As a preliminary step, we computed group heritability of 2-year language delay regardless of outcome. Probands were those meeting our criterion of ELD (i.e., those falling in the lowest 10% on the 2-year vocabulary measure), and standard DF analysis was used to predict co-twins' scores from the scores of the proband and the coefficient of relationship. There were 318 MZ probands and 252 DZ probands, and h^2_g was .240 (95% confidence interval [CI] = .12–.36). This analysis gives a different picture from the report by Dale et al. (1998), which was based on the bottom 5% of 2-year-olds in the 1994 cohort. In that report, a value of h^2_g of .73 (95% CI = .38–1.00) was found for the 2-year-old vocabulary measure.

There were several differences between the dataset used by Dale et al. (1998) and that used here. First, we used a less extreme cutoff for abnormality (bottom 10% rather than bottom 5%; thus, twice as many children were selected as probands). Second, we excluded children who did not have follow-up data at 3 or 4 years, whereas all those with data for age 2 were included by Dale et al. (1998). Third, a small number of children who had been included in Dale et al. (1998) were excluded here because exclusionary medical factors were discovered after 2 years of age. And fourth, we included data from both 1994 and 1995 cohorts, whereas Dale et al. (1998) included data only from those born in 1994. Additional analyses were conducted to see which of these factors might be responsible for the lower heritability values that we observed in this analysis.

To see whether the use of a less extreme cutoff was responsible for the change in heritability estimates, the data for both 1994 and 1995 cohorts were reanalyzed in the current sample using the more extreme 5% cutoff adopted by Dale et al. (1998). This gave a heritability estimate that was only slightly and nonsignificantly higher than the value using a 10% cutoff. In the current sample, there were 127 MZ and 104 DZ probands, and h^2_g was .318 (95% CI = .11–.53).

We next looked at those twins born in 1994 who were excluded from the current analysis but were included in Dale et al. (1998). These included 37 twin pairs with

missing or incomplete data at 3 and 4 years, 39 pairs who had ambiguous vocabulary scores at 2 years (maternal report of zero words, but evidence from other items that the child did produce speech), 8 pairs whose questionnaires were not returned within 3 months of the 2nd birthday, and 9 pairs in whom an exclusionary medical diagnosis was made after 2 years of age. The concordance rates seen in these excluded pairs for MZ and DZ twins were comparable to those for the current sample, and heritability estimates for the 1994 cohort barely changed when they were excluded.

This suggested that the difference between analyses must arise because of inclusion of the 1995 cohort, and this indeed proved to be the case. Using just those children in the current sample, with a 5% cutoff, the 1994 cohort gave h^2_g of .583 (based on 179 probands), whereas for the 1995 cohort h^2_g was only .173 (based on 173 probands). Thus, it appeared that sampling variation played the major role in accounting for the difference between the current estimates and those previously reported by Dale et al. (1998). This is a surprising and counterintuitive finding, and we can think of no good reason why heritability should differ from one year to another. Because heritability is a function of the difference in concordance between MZ and DZ twins, the drop in heritability from 1994 to 1995 could reflect either an increase in 1995 in the proportion of discordant MZ twins or an increase in the proportion of concordant DZ twins. The latter appeared to be the case. We computed pairwise concordance (i.e., the number of concordant twin pairs expressed as a percentage of all pairs containing a proband). For MZ twins in the 1994 cohort this was 73/97 = 75% and for DZ twins it was 30/71 = 42%. In 1995, pairwise concordance for MZ twins was closely comparable at 51/69 = 74% but for DZ twins it was 42/79 = 53%. The reasons for the difference between cohorts are unclear, but we assume that the current results, using both cohorts combined, should be accepted as the more reliable.

Heritability in Relation to Outcome

To give a preliminary impression of the data, Table 2 shows categorical concordance for twin pairs in relation to outcome status (based on parental report of language attainments). When concordance is high and similar for both MZ and DZ twins, this points to a shared environmental basis for ELD, whereas a higher concordance for MZ than DZ twins is consistent with a genetic etiology. At both 3 and 4 years of age there is a significant difference between the categorization for MZ and DZ twin pairs; for 3-year outcome (columns 3 and 4), $\chi^2(4, N = 319) = 26.7$, $p < .001$; for 4-year outcome (columns 5 and 6), $\chi^2(4, N = 257) = 22.2$, $p < .001$. It is evident that this reflects higher levels of concordance in the MZ pairs, both with respect to classification at age 2 (compare rows 1–2 with rows 3–5) and with respect to outcome at 3 or 4 years (see rows 3–5). Data from Table 2 can also be used to evaluate specifically how far twins with ELD tend to have the same type of outcome—transient or persistent. To look at this question, we restricted attention to pairs who are concordant for ELD (i.e., rows 3–5), comparing the distribution of concordant outcome pairs (rows 3 and 5 combined) and discordant outcome pairs (row 4) for MZ and DZ pairs. MZ twins are significantly more likely than DZ twins to be in the same outcome subgroup, both at 3 years, $\chi^2(1, N = 191) = 11.5$, $p < .01$, and at 4 years, $\chi^2(1, N = 166) = 8.49$, $p < .01$. This suggests that there is a genetic influence on categorization as persistent or transient difficulties within the ELD category.

We then went on to obtain quantitative estimates of heritability of 2-year vocabulary impairment, repeating the DF analysis as previously, but first subdividing the sample according to outcome. Tables 3 and 4 show mean 2-year vocabulary *z* scores for probands and cotwins in relation to language outcomes at 3 and 4 years, together with estimates from DF analysis. Influences on ELD can be partitioned into those due to genes (h^2_g), environmental factors shared by both twins (also known

Table 2. Numbers (%) of twin pairs categorized by zygosity and language outcome at 3 or 4 years, including all pairs in which at least one twin was categorized as ELD at 2 years.

Pairwise concordance for ELD	Composition of twin pair	3-year outcome		4-year outcome	
		MZ	DZ	MZ	DZ
Discordant	One twin transient, other twin not ELD	32 (19.3%)	47 (30.7%)	24 (16.9%)	34 (29.6%)
Discordant	One twin persistent, other twin not ELD	17 (10.2%)	32 (20.9%)	12 (8.5%)	21 (18.3%)
Concordant	Both twins transient	51 (30.7%)	27 (17.6%)	58 (40.8%)	26 (22.6%)
Concordant	One twin transient, other twin persistent	13 (7.8%)	22 (14.4%)	9 (6.3%)	15 (13%)
Concordant	Both twins persistent	53 (31.9%)	25 (16.3%)	39 (27.5%)	19 (16.5%)
Total *N*		166	153	142	115

Note. Outcome at 3 or 4 years was categorized as persistent if the child scored below the 15th centile on at least 2 out of 3 of the parental reports of language attainments (vocabulary, grammar, and abstract language).

Table 3. Mean (*SD*) 2–year vocabulary *z* scores for probands selected on the basis of early language delay and their co–twins, in relation to proband outcome at 3 years.

Zygosity/ outcome	*N*	Proband		Co-twin		h^2_g (95% CI[a])	c^2_g	e^2_g
		M	*SD*	*M*	*SD*			
1. Outcome in terms of 3-year language (low on 2 of 3 measures); transient = 0, persistent =1								
MZ transient	147	−1.55	0.162	−1.45	0.349			
DZ transient	123	−1.54	0.174	−1.27	0.510			
MZ persistent	136	−1.57	0.156	−1.54	0.207	.22 (.07 to .37)	.71	.06
DZ persistent	104	−1.58	0.179	−1.35	0.613	.25 (.08 to .42)	.72	.02
2. Outcome in terms of 3-year parental concern; transient = none, persistent = concern								
MZ transient	146	−1.53	0.149	−1.46	0.248			
DZ transient	130	−1.52	0.164	−1.38	0.386	.10 (–.01 to .21)	.85	.04
MZ persistent	137	−1.59	0.165	−1.52	0.332			
DZ persistent	97	−1.61	0.183	−1.21	0.724	.41 (.20 to .62)	.55	.04
3. Outcome in terms of seeing professional at 3 yr; transient = no, persistent = yes								
MZ transient	183	−1.53	0.145	−1.47	0.240			
DZ transient	145	−1.53	0.159	−1.36	0.383	.14 (.03 to .24)	.82	.04
MZ persistent	100	−1.61	0.171	−1.53	0.368			
DZ persistent	82	−1.62	0.193	−1.21	0.776	.40 (.15 to .65)	.55	.05

[a]Corrected for double entry (cases where both twins are probands) by multiplication of *SE* by $((D - 3)/(S - 3))^{1/2}$, where *D* is number of probands and *S* is number of twin pairs contributing to probands.

as common environment, or c^2_g), and other sources of variance unique to the individual (e^2_g). The analytic method scales these so that they add to one. The values reported in these tables were obtained by conducting standard DF analysis for each outcome subgroup separately, using all available children at each age with relevant data (i.e., the samples in Tables 3 and 4 overlap substantially, but not completely, because some children had data for only one follow-up time point. Note also that the samples included in analyses of transient difficulties and persistent difficulties within the numbered sections of each table are not entirely independent, as pairs with one twin with transient difficulties and one twin with persistent difficulties will be included in both sets of analyses). In all the analyses, probands are defined in terms of having a low vocabulary *z* score at 2 years, and the analysis involves predicting the co-twin's 2-year vocabulary score from that of the proband and from the coefficient of relationship. If the MZ co-twins obtain lower vocabulary *z* scores than the DZ co-twins, then this is evidence of genetic influence. The h^2_g coefficient, a measure of group heritability, can be interpreted as the extent to which the average difference between the probands and the population on the quantitative trait measure can be attributed to genetic influence. A group heritability of 0 indicates no genetic influence and a value of 1.0 would indicate that genes are the only factor influencing impairment. The significance of h^2_g can be estimated by taking the CI around the estimate to see if it is significantly greater than 0.

Inspection of these data suggests three interim conclusions. First, although most of the estimates of h^2_g are significantly greater than 0, they are much lower than heritability estimates obtained with older samples of children with SLI (e.g., Bishop et al., 1995). Second, shared environmental factors exert greater influence on ELD than do genes: All the estimates of c^2_g are larger than those of h^2_g. Third, the pattern of results suggests that the heritability of ELD does vary according to outcome, but only when outcome is measured in terms of parental concern (at 3 years) or professional involvement (at 3 or 4 years), and not when outcome is considered in terms of parental report of language attainments. For children whose parents report professional involvement at 3 or 4 years, heritability of ELD is around .4, whereas for those with no professional involvement at 3 or 4 years, heritability is nonsignificant at 3 years and very low at 4 years.

Direct Evaluation of Heritability Differences for ELD in Relation to Outcome

To test whether values of h^2_g are significantly different in relation to outcome, we used the augmented DF analysis described earlier. For each definition of outcome, transient and persistent subgroups were combined, and terms representing the proband's outcome subgroup and its interaction with other terms were

Table 4. Mean (*SD*) 2–year vocabulary *z* scores for probands selected on the basis of early language delay and their co–twins, in relation to proband outcome at 4 years.

Zygosity/ outcome	N	Proband		Co-twin				
		M	SD	M	SD	h^2_g (95% CI[a])	c^2_g	e^2_g
1. Outcome in terms of 4-year language (low on 2 of 3 measures); transient = 0, persistent = 1								
MZ transient	149	−1.54	0.157	−1.48	0.295			
DZ transient	101	−1.54	0.168	−1.29	0.543	.24 (.10 to .38)	.72	.04
MZ persistent	99	−1.59	0.240	−1.53	0.240			
DZ persistent	74	−1.60	0.183	−1.44	0.376	.13 (–.03 to .29)	.84	.03
2. Outcome in terms of 4-year parental concern; transient = none, persistent = concern								
MZ transient	189	−1.54	0.156	−1.48	0.273			
DZ transient	128	−1.55	0.175	−1.33	0.510	.20 (.06 to .33)	.76	.04
MZ persistent	59	−1.61	0.154	−1.55	0.279			
DZ persistent	47	−1.60	0.180	−1.40	0.408	.17 (–.21 to .56)	.80	.04
3. Outcome in terms of seeing professional at 4 yr; transient = no, persistent = yes								
MZ transient	171	−1.54	0.147	−1.49	0.270			
DZ transient	118	−1.54	0.168	−1.42	0.316	.08 (–.02 to .18)	.88	.03
MZ persistent	77	−1.60	0.175	−1.53	0.286			
DZ persistent	57	−1.61	0.187	−1.21	0.701	.41 (.16 to .66)	.55	.04

[a]Corrected for double entry (cases where both twins are probands) by multiplication of *SE* by $((D-3)/(S-3))^{1/2}$, where *D* is number of probands and *S* is number of twin pairs contributing to probands.

added to the regression. The regression coefficients in Table 5 test the significance of the interaction (i.e., a significant coefficient indicates that group heritability differs reliably depending on the child's status regarding the outcome category). A positive value of *b* indicates that persistent difficulties are more heritable than transient, and a negative value indicates the converse. Because our a priori prediction was that any differences would reflect higher heritability for persistent difficulties, we adopted a Bonferroni-corrected *p* value of .1/6 = .017. At neither age was the interaction significant when outcome was coded in terms of language attainment. However, when outcome was coded in terms of parental concern the interaction was significant for the 3-year outcome. Finally, when outcome was coded in terms of whether or not the child had seen a professional because of concerns about speech and language, the interaction was significant at 4 years, with a trend in the same direction at 3 years. As can be seen from Tables 3 and 4, these interactions reflect higher group heritability for children who saw a professional and for those who aroused parental concern at 3 years.

Characteristics of ELD Children Whose Parents Sought Professional Help at 3 or 4 Years

Given that the highest estimates of heritability were found for children whose parents had sought professional

Table 5. Regression coefficient for interaction term in DF analyses shown in Tables 3 and 4.

Outcome category used in interaction	b	SE	P
3 year: low on 2 of 3 language measures	0.062	0.111	.575
3 year: parental concern	0.308	0.117	.009
3 year: professional involvement	0.277	0.122	.024
4 year: low on 2 of 3 language measures	–0.186	0.116	.108
4 year: parental concern	–0.017	0.129	.895
4 year: professional involvement	0.332	0.119	.005

Note. SE is corrected for double entry (cases where both twins are probands) by multiplication by $((D-K-1)/(S-K-1))^{1/2}$, where *D* is the number of probands, *K* is the number of independent variables in the regression equation, and *S* is number of twin pairs contributing probands.

advice about speech and language difficulties, it is of interest to consider the characteristics of such children. Although children in this subgroup overlap with those in the parental concern category (see Figure 2), this overlap is not complete, for two reasons. First, some 33% of ELD 3-year-olds and 13% of ELD 4-year-olds whose parents expressed concern about language development had not sought professional advice. Second, among those who did seek professional advice were some cases (11% at 3 years and 25% at 4 years) whose concerns were with speech rather than language (as evidenced by reported concerns: "it is hard for other people to understand him/her," or "s/he pronounces words poorly"). These children were not included in our parental concern category, which was restricted to those reporting slow language development or poor comprehension.

Table 6 shows the characteristics of ELD children at 2, 3, and 4 years, respectively, on measures of language and nonverbal ability, in relation to whether or not their parents sought professional advice at 4 years. (The 2-year grammar measure is excluded because virtually all ELD children scored 0. Children with incomplete data on the Parent Report of Children's Abilities [PARCA; Oliver et al., 2002; Saudino et al., 1998] are also excluded). The effect size statistic, η^2, indicates the proportion of variance explained. There were no significant differences between groups on the measures for a 2-year-old. These results are in line with those reported in our companion article for other outcome measures and suggest that at 2 years it would not be possible to distinguish accurately between ELD children who would and would not be receiving professional help at 4 years. For the 3-year data the trend is for means of those with subsequent professional contact to be lower, but there is substantial variation and only the abstract language measure resulted in a significant group difference. At 4 years, however, ELD children receiving professional help are clearly distinguished from those who do not have contact with services (multivariate analysis of variance $\eta^2 = .152$), with the most striking difference appearing on the grammar measure. It is noteworthy that when one compares mean nonverbal PARCA *z* scores at 3 and 4 years, the ELD children without subsequent professional involvement show similar levels of score, whereas the trend is for a decline with age in those with professional involvement. This could, however, reflect greater implicit verbal content in the 4-year PARCA, which includes tests of semantic categorization.

Additional analyses were conducted to compare these two groups on three background variables: sex, socioeconomic status (as indexed by mother's highest educational qualification), and the index of middle ear disease at 2 years (see Table 7). Only the middle ear disease rating gave a significant group difference ($\eta^2 = .017$). This index is scaled so that a low score indicates higher rates of disease. The direction of effect is opposite to that predicted by Lonigan et al. (1992), who suggested that children whose language delay resolves have higher rates of middle ear disease early in life. However, this study differed from that of Lonigan et al. in two important respects. We relied on parental report of symptoms, whereas the diagnosis was made by a physician in their study. Further, they found that otitis media during 12–18 months of age was the critical period

Table 6. Language and nonverbal scores of children with ELD in relation to whether or not parents sought professional advice at 4 years (children with full data on all variables).

	Professional advice at 4 yr						
	No (*N* = 182)		Yes (*N* = 81)				
Measure	*M*	*SD*	*M*	*SD*	*F*(1, 261)	*p*	η^2
2-year measures							
Vocabulary *z* score	−1.54	0.15	−1.56	0.18	0.59	.441	—
Displaced reference, *z* score	−1.29	0.71	−1.43	0.57	2.34	.127	—
Nonverbal *z* score (PARCA)	−0.45	0.94	−0.54	1.04	0.51	.477	—
3-year measures							
Vocabulary *z* score	−0.91	0.75	−1.02	0.8	1.24	.267	—
Grammar rating (raw)	1.73	0.83	1.57	0.77	2.10	.148	—
Abstract language, *z* score	−1.05	0.97	−1.34	0.97	4.71	.031	.018
Nonverbal *z* score (PARCA)	−0.47	1.01	−0.73	1.19	3.31	.070	—
4-year measures							
Vocabulary *z* score	−0.62	0.93	−0.97	0.78	8.39	.004	.031
Grammar rating (raw)	5.70	0.48	5.20	0.73	44.12	<.001	.145
Abstract language, *z* score	−0.63	1.1	−1.06	1.37	7.14	.008	.027
Nonverbal *z* score (PARCA)	−0.52	1.04	−0.82	1.13	4.62	.032	.017

Table 7. Background characteristics of children with ELD in relation to whether or not parents sought professional advice at 4 years.

	Professional advice					
	No (*N* = 289)		Yes (*N* = 134)			
Background measures	*M*	*SD*	*M*	*SD*	Test statistic	*p*
% male	65.4		70.9		$\chi^2 = 1.25$	.263
SES (mother's highest qualification)	3.55	1.83	3.16	1.69	$t(409) = 1.92$	.056
Middle ear disease score, 2 yr	23.62	3.07	22.66	4.14	$t(385) = 2.41$	.017

Note. SES = socioeconomic status.

for seeing group differences, whereas we had only a global rating at 18 months, and so our measures may have been insensitive to such an effect.

Discussion

This analysis of genetic influences on ELD leads to three main conclusions. First, when the whole sample is treated together, genetic effects on low vocabulary at 2 years are statistically significant, but much smaller than those previously reported for older samples of children with SLI. Second, environmental influences shared by both twins are more substantial than genetic factors in accounting for ELD. Third, when the ELD sample is subdivided according to outcome at 3 or 4 years, heritability is significantly higher in those with persisting difficulties, but only when outcome is assessed in terms of parental concern at 3 years or professional involvement at 4 years. Differences in heritability of ELD were *not* seen when we considered outcome purely in terms of children's language attainments in vocabulary, grammar, and abstract language, as reported by parents.

Why Are Genetic Effects Relatively Small in This Sample?

Both Whitehurst et al. (1991) and Lyytinen et al. (2001) have proposed that samples of children selected on the basis of small vocabulary at 2 years will contain a heterogeneous mixture of etiologies. On this view, although we would expect any vocabulary screening of 2-year-olds to net potential cases of heritable SLI, it may also pick up substantial numbers of children with other, environmental reasons for language failure, and so dilute heritability estimates. Our data are compatible with this viewpoint: Heritability of 2-year vocabulary is moderate and significant if we restrict attention to those late-talking 2-year-olds who subsequently go on to merit professional concern at 4 years. In contrast, heritability is nonsignificant for late-talking 2-year-olds who do not have contact with professional services at 4 years.

This interpretation of the data must, however, be tempered by methodological considerations. We found unexpected differences between our heritability estimates for 2-year vocabulary delay and that reported by Dale et al. (1998), which partially overlapped with this sample. Supplementary analyses indicated that this was due to the inclusion in the current study of an additional cohort of children born in 1995, who had lower heritability estimates than those born in 1994. The reasons for this remain unclear.

Another point to note is that estimates of shared environmental influences on ELD in our sample will be influenced by reporting bias, because the same parent completed the questionnaires for both twins. Any general tendency to exaggerate or minimize language difficulties will affect ratings of both members of a twin pair, regardless of whether they are MZ or DZ, and this will be included in the *shared environment* term. We aimed to minimize such effects as far as possible by asking parents to complete questionnaires for the 2 children at different time points and by making ratings concrete rather than impressionistic, but clearly we cannot discount a role for rater bias.

A further point we need to bear in mind is that heritability estimates are specific to the population from which they were derived and may not generalize to a population with different environmental influences. For instance, quality and quantity of intervention may affect outcomes. In an epidemiological study of this kind, we were not in a position to evaluate the effects of intervention by manipulating the kinds of treatment children received. Children who received speech and language treatment would be counted in the category of "seen professional" (though this category will also include those who were reviewed but not treated and those seeing other types of professionals). These constituted 36% of ELD children at 3 years and 32% at 4 years (see Figure 2). One can also see from Figure 2 that most

children who had professional contact (90% at age 3 years and 27% at 4 years) had evidence of persisting difficulties (either on the basis of language ratings or in terms of parental concern); this is a not uncommon finding, given that scarce intervention resources tend to be directed toward the children with the most severe and intractable problems.

Why Do Different Outcome Criteria Give Different Results?

We had expected that parental report of the child's vocabulary, grammar, and use of abstract language would provide a better measure of outcome than more global ratings of parental concern or professional involvement. As can be seen from Figure 2, these ratings did identify persisting language difficulties in many ELD children who had not aroused parental or professional concern. This is consistent with epidemiological data from Tomblin et al. (1997), who found that a substantial proportion of 6-year-olds who met criteria for SLI on language assessment had not been identified by clinical services.

Intriguingly, in this sample, if we identified persisting language difficulties at 4 years purely in terms of parental report of vocabulary, grammar, and abstract language, the heritability of ELD was nonsignificant (see Table 4, section 1). This contrasts with the heritability estimate of .4 for children who had involvement of clinical services, which approaches values reported for older samples with SLI. This suggests that there may be phenotypic differences, not captured by our parental report measures, which lead parents to seek professional help, and which distinguish heritable from nonheritable language problems. These may include difficulties with language comprehension and with speech production: for a substantial minority of children who merited professional attention at 4 years, the main concern was with speech rather than language. In studies of older children with heritable SLI, both receptive language impairments and early articulation difficulties are common (Bishop et al., 1995): it may be that these, rather than rate of vocabulary development, will prove to be the best indicators of persistent, heritable communication problems. Another factor that may become increasingly evident as children grow older is grammatical impairment. Research has shown that children with SLI have problems with some aspects of grammatical morphology that are disproportionate relative to their difficulties with other aspects of language, such as vocabulary (see Leonard, 1998, for review). It is noteworthy that the simple grammar rating used at 4 years was the best measure for distinguishing between those who did and did not receive professional attention. It seems likely that as they mature, some of the ELD group with persisting difficulties will show characteristics of typical SLI. In work currently in progress, we are directly assessing language skills in a subset of these children, with the aim of specifying more precisely the nature of the phenotype of heritable language impairment. A further possibility is that children whose parents seek professional help may be those with significant difficulties outside the domain of language (e.g., behavioral and attentional difficulties).

Another reason why parental ratings of vocabulary, grammar, and abstract language may have been ineffective in picking up heritable impairments is that these require parents to act as reporters on differentiated aspects of their child's development. It is possible that parents who can accurately state that they are concerned that their child is having communication difficulties are nevertheless relatively poor at evaluating aspects of language such as words or constructions used by the child. Ability to make such judgments is likely to vary from one parent to another, and it may be particularly compromised where parents themselves have a history of language difficulties. In the future, we plan to compare results of parental report with direct assessment of a subset of children from this sample so that we can determine how well parental report agrees with language test data. These data will also allow us to address another question raised by these results: Are the higher concordance rates in MZ twins versus DZ twins with persistent difficulties influenced by a tendency of parents to evaluate and treat their MZ twins more similarly than their DZ twins? If such a tendency is operating, direct assessment should not find such an MZ–DZ difference.

A related point arising from this study is whether simply asking about parental concern would add useful information when identifying children with language difficulties at earlier ages. This is not usually done, perhaps because it is assumed that parents might be unaware of language impairments in very young children, especially if they have no older children. Expectations of children's language at 2 years are likely to be more variable than at 3 or 4 years. Furthermore, if "concern" is simply rated in a binary fashion, it does not allow us to assess the severity of problems in the way that a vocabulary scale does, and it would not be suitable for the kinds of quantitative analysis adopted here. We did not include a parental concern item at 2 years of age, but in future studies this would be worth doing to see how far it picks up children who (a) are not identified by a measure such as the MCDI or (b) have a poor prognosis.

In sum, when we include data from three time points, we obtain a clearer picture of etiological factors for language delay, with evidence that environmental

factors shared by both twins play the major role in causing early language delay. There was little or no evidence for genetic effects on 2-year-old vocabulary in children who were not subsequently referred for services, but there was evidence of modest genetic effects in those whose parents did seek professional help for them by 4 years of age. Our data suggest that in predicting language outcomes of children with early delay, family history of speech and language impairment should be taken into account. This concurs with the study of late talkers by Lyytinen et al. (2001). They concluded that a 2-year-old with low expressive language skills should not prompt concern if there are no other endogenous or exogenous risk factors present. However, if the child has a family history of language difficulties, then there is a greater likelihood of persisting difficulties.

Finally, our findings lead us to recommend that molecular studies of genetics of language impairment should focus on persistent language impairment, rather than transient early delay. We note that the present results for persistent delay are compatible with a range of possible genetic mechanisms, including multiple quantitative trait loci of small effect as well as a rare gene of large effect. Molecular genetic studies in progress should help clarify the mechanism.

Acknowledgements

We thank the parents of the twins in the Twins Early Development Study (TEDS) for making this study possible. The TEDS project is supported by a program grant from the U.K. Medical Research Council. Dorothy Bishop is supported by a Wellcome Trust Principal Research Fellowship.

References

Bishop, D. V. M. (1994). Is specific language impairment a valid diagnostic category? Genetic and psycholinguistic evidence. *Philosophical Transactions of the Royal Society of London, Series B, 346*, 105–111.

Bishop, D. V. M., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: Evidence from a twin study. *Developmental Medicine and Child Neurology, 37*, 56–71.

Castles, A., Datta, H., Gayan, J., & Olson, R. K. (1999). Varieties of developmental reading disorder: Genetic and environmental influences. *Journal of Experimental Child Psychology, 72*, 73–94.

Dale, P. S., & Cole, K. N. (1991). What's normal? Specific language impairment in an individual differences perspective. *Language, Speech, and Hearing Services in Schools, 22*, 80–83.

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., Eley, T. C., Oliver, B., Price, T. S., et al. (1998). Genetic influence on language delay in two-year-old children. *Nature Neuroscience, 1*, 324–328.

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

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74,* 394–412.

Fenson, L., Pethick, S. J., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21*, 95–115.

Gopnik, M. (1999). Familial language impairment: More English evidence. *Folia Phoniatrica et Logopaedica, 51*, 5–19.

Gopnik, M., & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition, 39*, 1–50.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory: Lower scores in the UK than the USA. *Journal of Child Language, 27*, 689–705.

Haynes, C., & Naidoo, S. (1991). *Children with specific speech and language impairment* (Clinics in Developmental Medicine, Vol. 119). London: MacKeith Press.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*, 519–523.

Leonard, L. B. (1987). Is specific language impairment a useful construct? In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (Vol. 1, pp. 1–39). Hillsdale, NJ: Erlbaum.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Lewis, B. A., & Thompson, L. A. (1992). A study of developmental speech and language disorders in twins. *Journal of Speech and Hearing Research, 35*, 1086–1094.

Lonigan, C. J., Fischel, J. E., Whitehurst, G. J., Arnold, D. S., & Valdez-Menchaca, M. C. (1992). The role of otitis media in the development of expressive language disorder. *Developmental Psychology, 28*, 430–440.

Lyytinen, P., Poikkeus, A., Laakso, M., Eklund, K., & Lyytinen, H. (2001). Language development and symbolic play in children with and without familial risk for dyslexia. *Journal of Speech, Language, and Hearing Research*, *44*, 873–885.

Neligan, G. A., & Prudham, D. (1969). Norms for four standard developmental milestones by sex, social class and place in family. *Developmental Medicine and Child Neurology*, *11*, 413–422.

Oliver, B., Dale, P. S., Saudino, K. J., Petrill, S. A., Pike, A., & Plomin, R. (2002). The validity of a parent-based assessment of the nonverbal cognitive abilities of three-year-olds. *Early Child Development and Care, 172*, 337–348.

Paul, R. (1991). Profiles of toddlers with slow expressive language development. *Topics in Language Disorders, 11*, 1–13.

Paul, R. (2000). Predicting outcomes of early expressive language delay: Ethical implications. In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 195–209). Hove, U.K.: Psychology Press.

Price, T. S., Freeman, B., Craig, I. W., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3*, 129–133.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers at 2: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27,* 643–664.

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

Rovers, M., Haggard, M., Gannon, M., Koeppen-Schomerus, G., & Plomin, R. (2002). Heritability of symptom domains in otitis media: A longitudinal study of 1,373 twin pairs. *American Journal of Epidemiology, 155,* 958–964.

Rutter, M., & Redshaw, J. (1991). Growing up as a twin: Singleton–twin differences in psychological development. *Journal of Child Psychology and Psychiatry, 32*, 885–895.

Saudino, K. J., Dale, P. S., Oliver, B., Petrill, S. A., Richardson, V., Rutter, M., et al. (1998). The validity of parent-based assessment of the cognitive abilities of two-year-olds. *British Journal of Developmental Psychology*, *16*, 349–363.

SLI Consortium. (2002). A genome-wide scan identifies two novel loci involved in specific language impairment (SLI). *American Journal of Human Genetics, 70*, 384–398.

Stromswold, K. (1998). Genetics of spoken language disorders. *Human Biology, 70*, 293–320.

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

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 40*, 1245–1260.

Van der Lely, H. K. J., Rosen, S., & McClelland, A. (1998). Evidence for a grammar-specific deficit in children. *Current Biology, 8*, 1253–1258.

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain, 125*, 452–464.

Whitehurst, G. J., Arnold, D. S., Smith, M., Fischel, J. E., Lonigan, C. J., & Valdez-Menchaca, M. C. (1991). Family history in developmental expressive language delay. *Journal of Speech and Hearing Research, 34*, 1150–1157.

Received December 7, 2001

Accepted December 17, 2002

DOI: 10.1044/1092-4388(2003/045)

Contact author: Philip S. Dale, PhD, Communication Sciences and Disorders, 303 Lewis Hall, University of Missouri, Columbia, MO 65211.
E-mail: dalep@health.missouri.edu