Outcomes of Early Language Delay: I. Predicting Persistent and Transient Language Difficulties at 3 and 4 Years

Philip S. Dale
University of Missouri–Columbia

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

Robert Plomin
Institute of Psychiatry, London

Parent-based assessments of vocabulary, grammar, nonverbal ability, and use of language to refer to past and future (displaced reference) were obtained for 8,386 twin children at 2 years of age. Children with 2 year vocabulary scores below the 10th centile were designated the early language delay (ELD) group, and their outcomes at 3 and 4 years were contrasted with the remainder of the sample, the typical language (TL) group. At 3 and 4 years old, children were designated as language impaired if their scores fell below the 15th centile on at least 2 of the 3 parent-provided language measures: vocabulary, grammar, and use of abstract language. At 3 years, 44.1% of the ELD group (as compared to 7.2% of the TL group) met criteria for persistent language difficulties, decreasing slightly to 40.2% at 4 years (as compared to 8.5% of the TL group), consistent with previous reports of frequent spontaneous resolution of delayed language in preschoolers. Although relations between language and nonverbal abilities at 2 years and outcome at 3 and 4 years within the ELD group were highly statistically significant, effect sizes were small, and classification of outcome on the basis of data on 2-year-olds was far too inaccurate to be clinically useful. Children whose language difficulties persisted were not necessarily those with the most severe initial difficulties. Furthermore, measures of parental education and the child's history of ear infections failed to substantially improve the prediction.

KEY WORDS: language assessment, language delay, preschoolers, longitudinal outcome

Recent advances in our ability to assess the language development of young children have had the effect of sharpening a dilemma long familiar to pediatricians, speech-language pathologists, and other professionals. How much concern and/or intervention is warranted when a child's early language development is notably slow? Although most children have a substantial vocabulary and have started putting words together by their second birthday, there is considerable variation in the age at which these milestones are achieved. Indeed, concern about slow language development is a common reason for parents to consult a pediatrician.

The availability of well-normed and valid parent report measures (Dale, 1996) has made it possible to determine with relative confidence whether a child's development falls, for example, in the lowest 10th percentile at 24 months. However, studies that follow late talkers over

time find that a high proportion of toddlers do appear to catch up in language development after a slow start and score within the normal range by the time they start school, though they are often still below norms for typically developing children or below a matched control group (Paul, 2000; Rescorla, 2002; Thal & Katich, 1996; Whitehurst, Fischel, Arnold, & Lonigan, 1992). Nevertheless, for some children, delayed acquisition of language milestones is the first indication of language impairment that will persist throughout childhood, interfering with everyday communication and academic attainment.

Thus, a question of considerable clinical importance is how to distinguish transient from persistent language difficulties in young children. Most clinicians would argue that early intervention is needed for children who are likely to go on to have long-term language difficulties. However, it would be inefficient to target intervention at all late-talking toddlers, given that we know a high proportion of these children will go on to develop normal language without any treatment. Intervention has both economic and psychological costs for families (Paul, 2000). Ideally, we would like to distinguish between those children whose language will not spontaneously improve and those whose language will improve without intervention.

This challenging problem is the clinical face of a larger question about human growth. Individuals differ not only in their level of development; they differ in the shape of their *trajectory*, or growth curve. The children under consideration have growth curves with a low initial slope, which may or may not be followed by acceleration. What factors explain the shape of trajectory? And even if we cannot explain the shape, are there factors that predict it?

Several studies have focused on the practical question of identifying predictors of outcome in young language-delayed children, but their conclusions have been limited by small sample sizes and reliance on volunteer samples. Three predictors have received attention: severity of initial impairment in expressive language, degree of impairment in receptive ability as well as expressive (this might be interpreted as another facet of severity), and degree of impairment in gestural communication. Rescorla and Schwartz (1990) found that severity of expressive language delay was a predictor of continued delay in a group of late talkers identified at 24–30 months. Fischel, Whitehurst, Caulfield, and Debaryshe (1989) found that within a sample selected for expressive but not receptive delays, children whose vocabularies were very restricted at 2 years of age (e.g., 8 words or less) had a worse prognosis than those who had a vocabulary of 20 or more words at that age. In Thal's studies (Thal & Katich, 1996), only receptive language and gestural communication had significant predictive value within the preschool period. Some of the variation in conclusions reflects differences in early identification criteria for delay. In general, however, these studies have demonstrated group differences in outcome based on these early predictors but have failed to demonstrate prediction at an individual level that would be adequate for clinical purposes.

Measures outside of language development proper have received limited attention in these predictive studies. Although there is a well-established sex difference in prevalence of early language delay, no studies have examined sex as a predictor of recovery from that delay. Various aspects of socioeconomic status, such as parental education and occupation, have been examined with inconsistent results. Paul (2000) noted a substantial prediction to later performance even within her largely middle class sample, whereas Thal and Katich (1996) reported little or no prediction. Socioeconomic status (SES) is particularly interesting as a potential predictor because of its demonstrated correlation with the quantity and quality of linguistic input provided to children (Hart & Risley, 1995).

Finally, although there is a large and inconsistent literature on the effects of otitis media on early language development, apparently only one study (Lonigan, Fischel, Whitehurst, Arnold, & Valdez-Menchaca, 1992) has examined this factor as a predictor within the early language delay group. They found that history of middle ear disease between 12 and 18 months predicted expressive language improvement in the ELD group. Lonigan et al. concluded that a subgroup of the ELD population has a language delay arising from early middle ear disease, and that this delay is likely to disappear once the transient effects on hearing have resolved.

In this article, we present parent-report data from the Twins Early Development Study (TEDS), a large-scale community study of U.K. twins, to address the question of whether it is possible to predict which 2-year-olds with early language delay will show significant language difficulties at 3 and 4 years of age.

Method

Participants

Parents of all twins born in England and Wales in 1994 and 1995 were contacted when the twins were 1 year old, after checking for infant mortality, and invited to participate in TEDS. A positive response was obtained from 11,352 families of the 15,906 who were contacted. These families were sent a booklet explaining the study in more detail and asking for background information about the twins. Background booklets were returned by 9,380 families. Further booklets (one for each twin) were sent shortly before the twins' second, third, and fourth

birthdays, including questions about language and cognitive development, medical history, and home background, as well as some tasks for parents to carry out with each twin (Trouton, Spinath, & Plomin, 2002).

The longitudinal analyses in this article require full parental report data on the measures described below at 2 years and at 3 years and/or 4 years. As shown in Table 1, a total of 5,208 families provided such data. For the analyses reported in this article, we excluded those with uncertain zygosity; those with major medical disorders, including serious perinatal hazard and genetic syndromes; and those whose primary language was not English. After the exclusions, 8,386 twins from 4,193 families (see Table 1, "Pairs remaining" row) remained in the sample for analysis. As noted by Eley, Dale, Bishop, Price, and Plomin (2001), twins with full data at 2 and 3 years of age were of slightly higher SES than the remainder of the sample (mothers with A-levels, the age 18 exam generally required for university entrance, 38% vs. 34%) and had slightly higher language scores at age 2 years (vocabulary scores = 47.7 vs. 45.2 on the MCDI:UKSF described below); both differences were significant with this very large sample size.

The analyses reported in this article were conducted on the whole sample as defined above, including the opposite-sex twin pairs, whereas the behavioral genetic analyses reported in the companion article in this issue (Bishop, Price, Dale, & Plomin, 2003) are restricted to same-sex twin pairs. All analyses in this article were repeated with the sample restricted to same-sex pairs, and in no case did the results change significantly.

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

Sample	Cohort		Total
	1994	1995	
Total sample with complete data at 2 yr, and at 3 and/or 4 yr	2,865	2,343	5,208
Excluded pairs:			
Uncertain zygosity	73	117	190
Medical exclusion	190	186	376
Non-English-speaking home	203	246	449
Pairs remaining	2,399	1,794	4,193
MZ	836	553	1,389
DZ – same sex	797	615	1,412
DZ – opposite sex	766	626	1,392
Complete data at 2, 3, and 4 yr	1,896	1,145	3,041
Complete data at 2 and 3 yr only	354	509	863
Complete data at 2 and 4 yr only	149	140	289
Sample size at 3 yr	2,250	1,654	3,904
Sample size at 4 yr	2,045	1,285	3,330

Measures

Vocabulary

The age-appropriate version of the MacArthur Communicative Development Inventory: U.K. Short Form (MCDI:UKSF; Dionne, Dale, Boivin, & Plomin, 2003) was included in each test booklet. This instrument includes a list of words from which parents are asked to check those that they have heard their child say. The score is the total number of words checked. The 100 words on the 2-year version were selected (Fenson et al., 2000) to give good prediction of the total score from the longer MacArthur Communicative Development Inventory (MCDI; Fenson et al., 1994). The list of words was then "anglicized" for appropriate spelling in a United Kingdom setting. The 3-year vocabulary measure was developed in accordance with similar design principles. In order to have an appropriate difficulty level and range of variation, 45 words were selected from the original, full MCDI, and 55 new words were included based on literature review and pilot testing (Dale, Reznick, Thal, & Marchman, 2001). The 4-year vocabulary measure, developed for this project, included 48 words chosen on the basis of literature review and pilot testing. For certain analyses, scores were converted to z scores after transformation to improve the normality of the distributions.

Grammar

The 2- and 3-year versions of the MCDI:UKSF included a grammar scale, with different items being used at 2 and 3 years. The first question asked whether the child is combining words. For the remaining items the rater was asked to indicate which of two sentences is most like the way the child talks. Both sentences in each item expressed the same meaning, but the first was developmentally simpler. For instance, in the 2-year version, one item was *baby want eat* versus *baby want to eat*. The 12 items on the 2-year version were selected from the full MCDI on the basis of good prediction of the full set of 37 items. The 12 items on the 3-year version included some new and more advanced aspects of grammar, chosen on the basis of literature review and pilot testing.

At age 2, grammar was scored on a 3-point scale (0 = *not yet combining words*, 1 = *using only the simple sentences of the pairs presented*, 2 = *using at least one of the more complex sentence forms*). The 3-year scale included three items that used more developmentally advanced forms, allowing more sensitive differentiation into a 5-point scale (0 = *not yet combining words*, 1 = *combining words but never using more complex forms*, 2 = *using at*

least one of the more complex options but not using any developmentally advanced forms, 3 = *using one or two of the more advanced forms*, 4 = *using all three of the more advanced forms*). At age 4, parents were asked to judge which of six statements best described how their child talked. The list of options, given in Appendix A, was converted to a 6-point ordinal Grammar Rating, with 6 indicating the most complex use of language.

Other Language Items

In the 2-year booklet, a further five items from the original MCDI were included to assess the ability to use language to refer to contexts other than the "here and now" (see Appendix B). These were combined to give a 10-point Displaced Reference scale. Further language items were included in the 3- and 4-year booklets to explore the child's receptive and expressive use of more abstract concepts (see Appendixes C and D). The number of "yes" responses was totaled to give an index of Abstract Language. Scores on the Displaced Reference and Abstract Language scales were converted to *z* scores.

At ages 3 and 4, parents were asked to indicate if they had any concerns about their child's speech and language, and if so to indicate the nature of the problem (see Appendix E). A final set of 6 items (see Appendix F) was used to assess whether the child showed any indication of communication abnormalities, such as failure to show joint attention or the presence of echolalia or language regression. These items were intended to identify children who might show characteristics of pervasive developmental disorder. Although children with a diagnosis of autism had been excluded from the sample, it was anticipated that some of those with slow language development might show milder autistic symptomatology. These items were summed to give a Communicative Abnormality scale.

Nonverbal Ability

The Parent Report of Children's Abilities (PARCA; Oliver et al., 2002; Saudino et al., 1998) consists of both parent-report questions and parent-administered items (e.g., copying shapes, imitating actions). The original 2-year version was extended to include content suitable for 3- and 4-year-old children. Total scores were converted to *z* scores relative to the whole sample.

Reliability and Validity of the Language and Nonverbal Measures

Many of the measures reported here were developed for the present study. As Fenson et al. (1994) pointed out, conventional indices of reliability have limited application and interpretability for parent-report measures. For this reason, the best evidence for reliability is the validity of the measures, because reliability sets an upper bound for validity. The 2-year language measures in this study were based on the MacArthur Communicative Development Inventory: Words and Sentences (MCDI:WS). Research reviewed in Fenson et al. (1994) and more recent studies, including Thal, O'Hanlon, Clemmons, and Fralin's (1999) study of children with language delay, have confirmed the high internal consistency of the MCDI:WS, its substantial correlation with vocabulary and grammar as measured by tests and language sample measures, and its ability to discriminate between these two aspects of language development. The 100 vocabulary items and 12 grammar items were selected to have excellent predictive validity (above .9 for both) to the corresponding full list (Fenson et al., 2000, for vocabulary; parallel analyses for grammar were done in the context of the present study).

The 3- and 4-year measures were developed for the present study. In a study of 85 British children age 32–40 months, Oliver et al. (2002) observed correlations of .68 and .48 for first- and second-born twins, respectively (i.e., birth order within the twin pair), between the 3-year vocabulary measure and the McCarthy Scales of Children's Abilities Verbal Score. Thal (cited in Dale et al., 2001) has obtained a correlation of .63 between the 3-year vocabulary measure and the Preschool Language Scale–3 Total Score for 19 typically developing children at 36–37 months. Although no results are yet available examining the validity of the 3-year vocabulary and grammar scores to corresponding measures from a language sample, the design similarity of the 3-year measure to the 2-year measure, and the substantial stability of vocabulary and grammar scores from 2 to 3 (Dionne et al., 2003), suggest that the specific 3-year measures have reasonable validity.

The 4-year measure is the newest. Evidence for the validity of the measure for selecting children with low language development was provided by the results of two related studies by Colledge and colleagues. Colledge et al. (2002) selected 600 twins from the TEDS sample and administered nine measures in the children's home. The nine measures were aggregated into a composite and standardized. Viding et al. (in press) selected 1,025 twins, 5.7% of the available sample, for low language on the basis of parent report, using a composite of vocabulary, grammar, and abstract language use. The nine measures were administered to this sample, and the composite was calculated; for the twins selected for low parent-reported language, the mean of the tester-administered composite was –1.23, approximately the 11th percentile. In addition to regression to the mean, some of the discrepancy reflects the fact that the tester-administered battery included a wider range of language measures, including articulation, phonological awareness, and narrative skills.

Saudino et al. (1998) obtained a correlation of .55 between the 2-year PARCA and the Mental Development Index of the Bayley Scales of Infant Development–II for a group of 107 two-year-old twins. This correlation rose to .66 when the 2-year vocabulary and grammar scores were also used as predictors.

SES

SES was assessed in two ways: first by occupational status, utilizing the higher code of fathers' and mothers' occupation on a 6-point scale, and second by mothers' and fathers' highest educational qualification on an 8-point scale. Although all three measures were significantly associated with measures of children's language development, the relationship was consistently strongest for mother's highest educational qualification, and only those results are reported here. On this scale, 1 = *no qualifications*, 2 = *below standard for a pass on the school-leaving examination*, 3 = *O-levels* (passing score on school-leaving examination), 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*.

Assessment of Ear Infections

At child age of 18 months, and 3 and 4 years, parents responded to a set of questions concerning ear infections and hearing loss, such as earache, mucus effusion from the ear, heavy nasal discharge, mouth breathing, apparent difficulty hearing during a cold, and others. Parents selected from the responses *often*, *sometimes*, *occasionally*, and *never*. These responses were scored from 1 to 4, respectively, and totaled to derive a scale in which low scores indicated the presence of increased middle ear disease. Rovers, Haggard, Gannon, Koeppen-Schomerus, and Plomin (2002) have shown that a first principal component of these items at each age demonstrates reasonable coherence (all items loading above .45) and substantial heritability (.49, .66, and .71 at the three ages, respectively). However, it should be noted that this measure did not assess degree of hearing loss or chronicity of severe ear infections with great precision.

Results

Characteristics of Children Who Had Language Delay at 2 Years

Dale et al. (1998) identified as language delayed those 2-year-olds from the 1994 cohort who had vocabulary scores of 9 or less. For the 1994 and 1995 cohorts reported here, this criterion would identify 4.5% of children as language delayed. To obtain an adequate sample size for developmental genetic analysis of language delay (see Bishop et al., 2003—our companion article in this issue), we adopted a less stringent cutoff vocabulary score of 15 or less for the present study, a cutoff that was close to the 10th centile. The 802 children scoring below this cutoff, constituting 9.6% of the sample, are referred to as the early language delay (ELD) group, whereas the remainder are referred to as the typical language (TL) group. A vocabulary criterion for early delay was selected because, due to the slight delay typical of twins at this age, there is a substantial floor effect for grammar. More than 30% of the sample scored at 0 (not yet combining words) on the grammar scale, which rendered it impossible to select a lowest 10% group. For selected analyses, we also report results using a more stringent criterion for delay at 2 years, the lowest 5%, to evaluate the possibility that prediction is enhanced within this more extreme group.

Table 2 summarizes comparisons of the two groups on selected measures at 2 years. Boys (12.8%) were nearly twice as likely as girls (6.5%) to be in the ELD group. Monozygous twins (11.4%) were more likely to be delayed than dizygous twins (8.6%). On average, children in the ELD group had mothers with lower educational qualifications, though there was great variation within both groups; the effect size (η^2, the proportion of variance accounted for) was less than .01 (i.e., less than 1% of the variance). There was no relation between language delay at 2 years and ear infections as reported at 18 months.

Delay in vocabulary was generally accompanied by delay in grammar and displaced reference. There was a particularly close association between low vocabulary and failure to combine words at 2 years: 96% of the ELD group had grammar scores of 0 (no word combinations). This contrasts with the TL group where 29% obtained grammar scores of 0, 34% obtained scores of 1, and 37% obtained scores of 2 (see Dale, Dionne, Eley, & Plomin, 2000, and Dionne et al., 2003, for further discussion of the relation between lexical and grammatical development). In addition, the mean nonverbal ability *z* score was lower for the ELD group than for the TL group. However, there was substantial variation within the ELD group, and the effect size was relatively small (.058).

Outcome of Children Who Had Language Delay at 2 Years

Figures 1 and 2 show the distribution of vocabulary and Abstract Language scores at age 3 for ELD and TL groups. Figures 3 and 4 show the corresponding scores at 4 years. The distributions of grammar ratings are shown in Table 3. (In these and other analyses, the *N* varies slightly because of the pattern of missing responses.) Three conclusions are apparent from these outcome data: first, early vocabulary delay was

Table 2. Comparison of early language delay (ELD) and typical language (TL) groups on selected measures at 2 years.

Measure	ELD (N = 802)		TL (N = 7,584)		Significance testing and effect size for continuous measures
	M	*SD*	*M*	*SD*	
Gender					Fisher's exact test, $p < .001$
Males	525	12.8%	3,585	87.2%	
Females	277	6.5%	3,999	93.5%	
Zygosity					$\chi^2(4) = 123.2$, $p < .001$
MZM	207	16.2%	1,067	83.8%	
MZF	111	7.4%	1,393	92.6%	
DZ same sex M	180	12.5%	1,264	87.5%	
DZ same sex F	72	5.2%	1,308	94.8%	
DZ opposite sex	232	8.3%	2,552	91.7%	
Mother's educational qualification	3.30	1.71	3.85	1.99	$F(1, 8186) = 55.0$, $p < .001$, $\eta^2 = .007$
Ear infection score (18 mos)	23.57	3.15	23.48	2.93	$F(1, 7929) = 0.624$, *ns*, $\eta^2 = .00$
Language					
Vocabulary	9.37	4.08	51.78	22.34	[groups defined on this measure]
Grammar	0.05	0.26	1.08	0.81	$F(1, 7102) = 1{,}193.2$, $p < .001$, $\eta^2 = .144$
Displaced Reference	2.91	1.89	6.93	2.40	$F(1, 8217) = 2{,}043.1$, $p < .001$, $\eta^2 = .199$
Nonverbal score (PARCA)	–.69	.96	.08	.91	$F(1, 8260) = 513.2$, $p < .001$, $\eta^2 = .058$

Note. For gender and zygosity, percentages represent the proportion of children in the defined group who fall into the ELD or TL category. For all other measures, means and standard deviations (in parentheses) for the two groups are reported. MZM = monozygotic male; MZF = monozygotic female; DZ = dizygotic; PARCA = Parent Report of Children's Abilities.

a predictor of continuing low scores in all three language domains; second, there was overlap in the distributions for groups ELD and TL, reflecting great variability within both groups; and third, the degree of overlap was greater at 4 years than at 3 years.

Given that children with known diagnoses of autism had been excluded from the sample, it is perhaps not surprising that the measures of communicative abnormality, which had been included to identify autistic features, yielded very low scores for both groups. At 3 years, 85% of the TL group and 78% of the ELD group obtained scores of 0 or 1, 12% of the TL group and 17% of the ELD group obtained scores of 2, with only 3% of the TL group and 4% of the ELD group scoring 3 or

Figure 1. Vocabulary scores at 3 years for the TL and ELD groups.

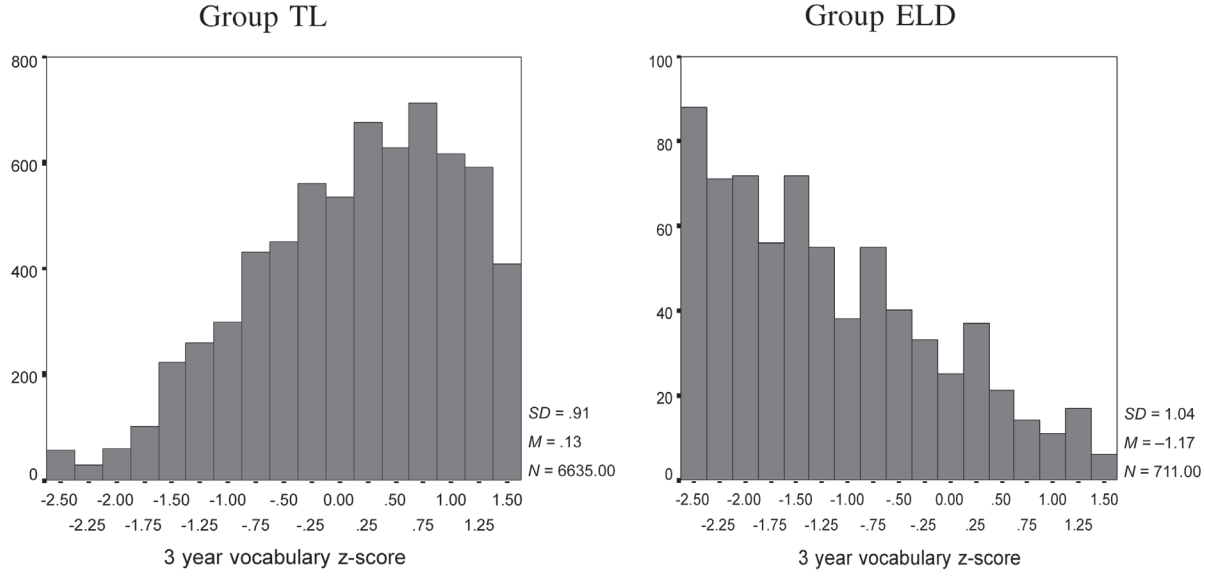
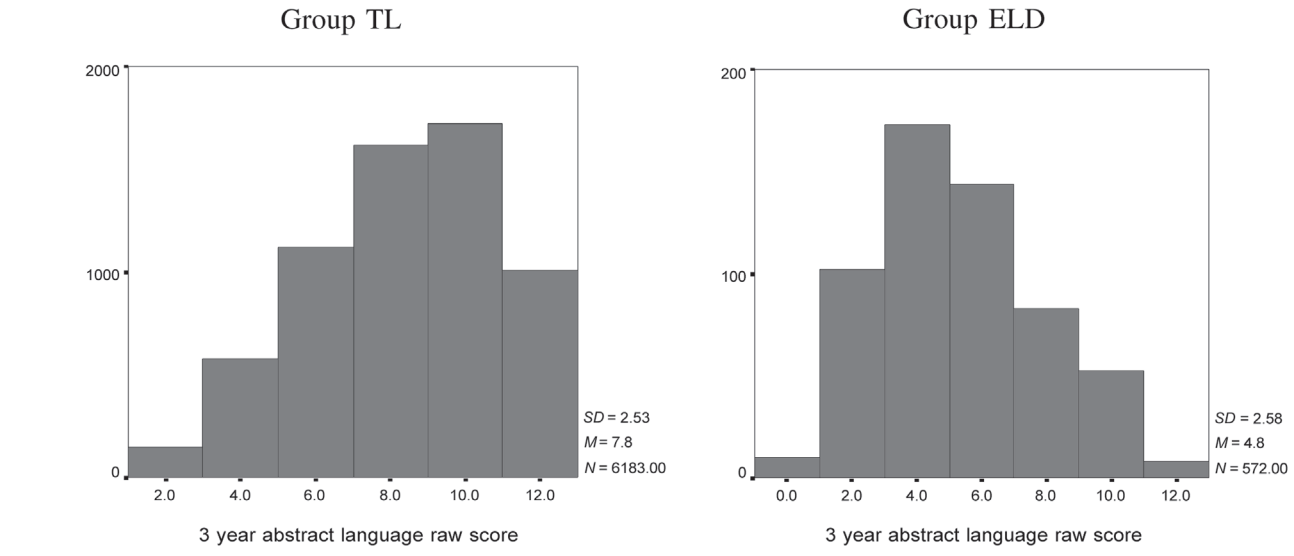


Figure 2. Abstract Language scores at 3 years for the TL and ELD groups.



more (out of a maximum score of 6). At 4 years, 5% of the TL group and 9% of the ELD group obtained a score of 2 or more.

Identification of Children With Language Difficulties at 3 Years and at 4 Years

Our primary interest in this article is distinguishing children with transient early language delay from those with more persistent difficulties. It would be possible to construct a criterion for language outcome based on the average score of the three language measures at each age (i.e., Vocabulary, Grammar, and Abstract Language). However, there are two difficulties with this approach. First, previous studies have found that vocabulary in late talkers is often substantially improved by 3 years, but grammar is more likely to remain impaired (Paul & Alforde, 1993; Rescorla, Roberts, & Dahlsgaard, 1997). Thus, an average measure might mask persisting difficulties. Second, our grammar measures at 3 and 4 gave an ordinal scale with few intervals, and so were not suitable for transformation to *z* scores as required for averaging. For these reasons, we

Figure 3. Vocabulary scores at 4 years for the TL and ELD groups.

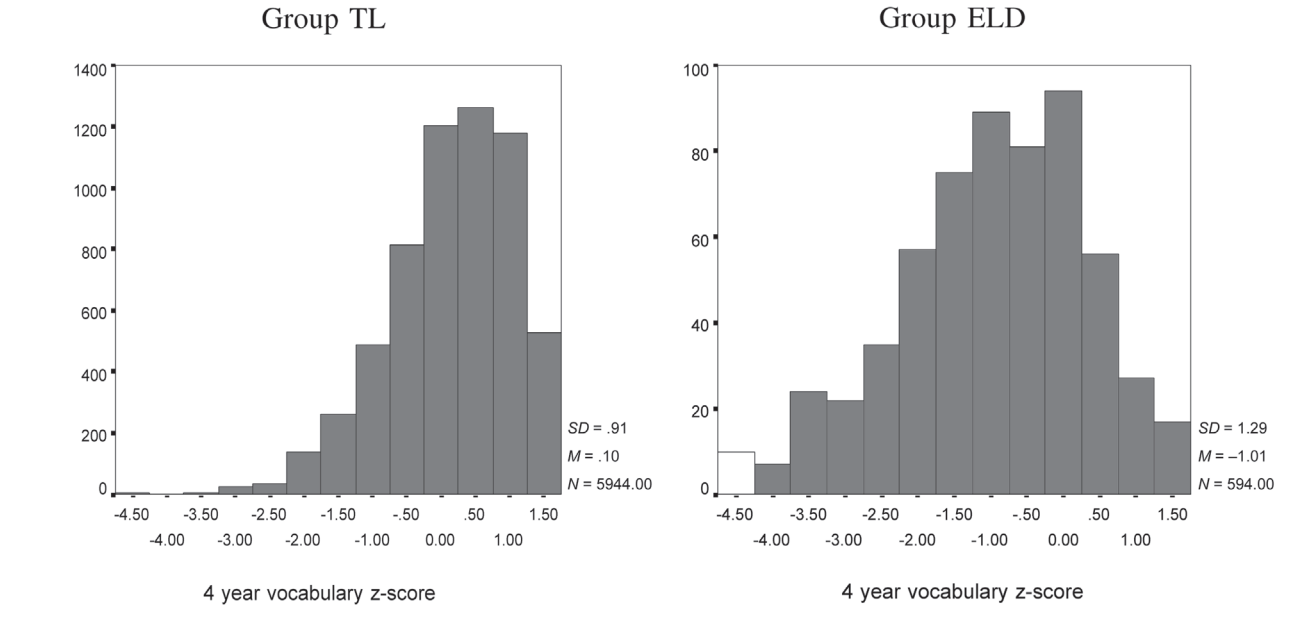
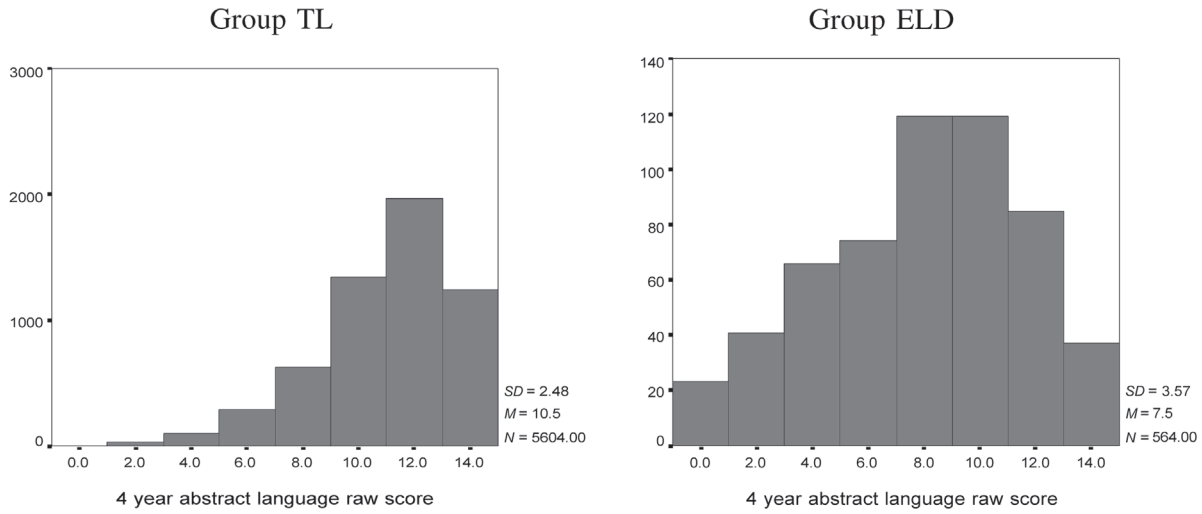


Figure 4. Abstract Language scores at 4 years for the TL and ELD groups.



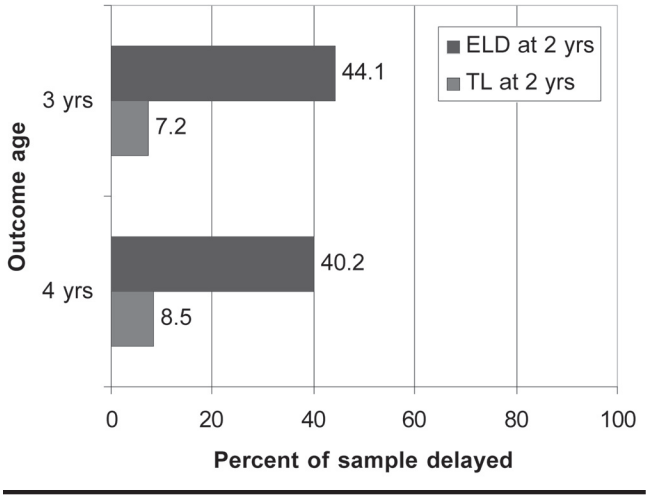
categorized children's outcome at 3 and 4 years as showing persistent difficulties if their scores on at least 2 of 3 language measures were at or below the 15th centile. In this way we avoided the need to use *z* scores while retaining sensitivity for detecting impairment in a child with an uneven language profile. The specific criteria for persisting language difficulties at 3 years were raw scores of less than 33 for Vocabulary, 2 for Grammar, and 5 for Abstract Language. A total of 835 children (10.7% of the total sample of 3-year-olds with available data) were impaired in at least two of the three domains, based on these criteria. At 4 years the criteria were less than 29 for Vocabulary, 6 for Grammar, and 8 for Abstract Language. A total of 764 children (11.5% of the 4-year-olds with available data) were impaired in at least two of the three domains, based on these criteria.

Figure 5 illustrates the relationship of difficulties at 3 and 4 years to classification as TL or ELD at 2 years. At both ages, the proportion was much higher for the ELD group, but less than 50%. Relative risk (or risk ratio) is a measure of the predictive value of one set of classification categories in relation to another, in this case 2 year classification and outcome classification. It is defined as the ratio of the probability that individuals in one category (ELD) will be in a specific outcome category (difficulties at 3 or at 4) to the probability that individuals in the other category (TL) will be in the same outcome category. The relative risk for

Table 3. Distribution of grammar ratings at ages 3 and 4 in relation to 2-year vocabulary.

	Grammar rating at 3 years						
Group	**0**	**1**	**2**	**3**	**4**		**Total**
TL							
N	11	532	2,158	2,426	1,273		6,400
%	0.2	8.3	33.7	37.9	19.9		100
ELD							
N	36	250	219	92	13		610
%	5.9	41.0	35.9	15.1	2.1		100
	Grammar rating at 4 years						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
TL							
N	11	7	1	37	623	5,296	6,589
%	0.2	0.1	0.0	0.6	10.4	88.6	100
ELD							
N	4	12	12	52	202	332	644
%	0.6	1.9	1.9	8.5	32.9	54.1	100

Figure 5. Proportion of children classified as delayed at 3 and 4 years as a function of classification at 2 years.



ELD as a predictor of persisting difficulties at 3 years was 6.1; that is, the probability of being in the persisting difficulties group at 3 years was six times greater for children in the ELD group than in the TL group. Similarly, the relative risk for ELD as a predictor of persisting difficulties at 4 years was 4.7.

It is possible that more extreme delays at 2 years have greater prediction of later language difficulties. To explore this possibility, the indices of the previous paragraph and Figure 5 were recomputed, using a criterion of lowest 5% at 2 years for early delay. As would be expected, the proportion of the ELD group classified as having difficulties at 3 and 4 years was somewhat higher, 47.3 and 46.3%, respectively. However, the proportion of the TL group classified as having difficulties at 3 and 4 years also increased, to 8.7% and 9.5%, respectively. Thus, there was little evidence for improved prediction using the more stringent criterion. This was confirmed by the calculation of the relative risk of ELD as a predictor of delay at 3 and 4 years; it was 5.4 and 4.9, respectively, similar in magnitude to the relative risk for the original, more liberal criterion. For readers more familiar with odds ratios in this context, the analogous figures for the previous paragraph are 10.2 and 7.2, respectively, and for the present paragraph, 9.4 and 8.2. Like relative risk, odds ratios measure the predictive value of a classification on the basis of the ratio of the likelihood of being in a specified outcome category if an individual is or is not in a specified predictor category, but in this case likelihood is measured by odds rather than probability. For example, if the probability of being in an outcome category is .80, corresponding to odds of 4:1, for individuals in the predictor category, and .20, corresponding to odds of 1:4, for individuals who are not, the relative risk would be .80/.20 = 4.0, whereas the odds ratio would be (4:1)/(1:4) = 4/.25 = 16.0.

Persistent and Transient Outcome at 3 and 4 Years for the ELD Group

For the remaining analyses, attention is focused on the ELD group, and the distinction within it between persistent and transient language difficulties. Table 4 shows the mean 2-year scores and other measures for ELD children who did and did not meet criteria for persisting language difficulties at 3 years, and Table 5 shows the corresponding data for children who did and did not meet criteria for persisting language difficulties at 4 years. Although all the differences at 2 years between

Table 4. Two-year measures for transient and persistent ELD subgroups, classified by 3-year outcome.

Measure	Transient group (*N* = 414) *M*	Transient group (*N* = 414) *SD*	Persistent group (*N* = 326) *M*	Persistent group (*N* = 326) *SD*	Significance and effect size
Vocabulary	9.72	4.07	8.90	4.11	$F(1, 738) = 7.3$, $p < .01$, $\eta^2 = .010$
Grammar	.06	.29	.03	.21	$F(1, 692) = 1.7$, *ns*, $\eta^2 = .002$
Displaced reference	3.21	1.94	2.53	1.71	$F(1, 714) = 24.1$, $p < .001$, $\eta^2 = .033$
PARCA (nonverbal) *z* score	−.59	.99	−.85	.90	$F(1, 784) = 14.4$, $p < .001$, $\eta^2 = .018$
Gender					
Males	260	53.7%	224	42.7%	
Females	154	60.2%	102	36.8%	Fisher's exact test, $p < .06$
Mother's educational qualification	3.52	1.76	3.07	1.58	$F(1, 718) = 12.5$, $p < .001$, $\eta^2 = .017$
Ear infection score					
At 18 months	23.75	3.18	23.56	2.95	$F(1, 697) = 0.67$, *ns*, $\eta^2 = .001$
At 3 years	18.64	3.28	18.73	3.07	$F(1, 697) = 0.14$, *ns*, $\eta^2 = .000$

Table 5. Two-year measures for transient and persistent ELD subgroups, classified by 4-year outcome.

Measure	Transient group (*N* = 372) *M*	*SD*	Persistent group (*N* = 250) *M*	*SD*	Significance and effect size
Vocabulary	9.94	4.04	8.38	4.08	$F(1, 622) = 22.0$, $p < .001$, $\eta^2 = .034$
Grammar	.04	.24	.04	.23	$F(1, 577) = 0.07$, *ns*, $\eta^2 = .000$
Displaced Reference	3.31	1.97	2.41	1.64	$F(1, 599) = 34.2$, $p < .001$, $\eta^2 = .054$
PARCA (nonverbal)	−.46	.96	−.95	.89	$F(1, 607) = 40.4$, $p < .001$, $\eta^2 = .062$
Gender					Fisher's exact test, $p < .06$
Males	230	57.4%	171	42.6%	
Females	142	64.3%	79	35.7%	
Mother's educational qualification	3.63	1.83	2.94	1.55	$F(1, 601) = 23.1$, $p < .001$, $\eta^2 = .037$
Ear infection score					
At 18 months	23.79	2.86	23.14	3.60	$F(1, 586) = 5.9$, $p < .05$, $\eta^2 = .010$
At 3 years	19.16	2.81	18.19	3.26	$F(1, 541) = 13.6$, $p < .01$, $\eta^2 = .025$
At 4 years	19.08	3.06	18.38	3.55	$F(1, 585) = 6.6$, $p < .01$, $\eta^2 = .011$

transient and persistent cases were highly statistically significant in this large sample, it is clear from the effect sizes that numerically they were very slight. To evaluate their combined ability to predict outcome, logistic regression analyses were carried out within the ELD sample to predict outcomes at 3 and 4 years from the 2-year data. Only those predictors that were significantly, or nearly significantly, related to the relevant outcome were included in the analysis. Logistic regression was favored over discriminant analysis because it makes weaker assumptions about the measures, and also because it permits the use of categorical predictors (in the present study, sex) as well as continuous predictors. Like discriminant analysis, it produces a predicted outcome classification for every participant, which can be compared with the actual classification.

The first two sections of Table 6 summarize the full set of analyses for the ELD sample. Table 7 provides the predicted × actual outcome cross-classification for one of these analyses, the fourth (and most accurate overall) of Table 6. The overall correct classification rate for these analyses was modest (60%–70%), but a more precise analysis is provided by four specific indices also included in Table 6. Sensitivity is the proportion of children with continuing language difficulties who had been predicted to have them; for the example of Table 7, this was 120/233 = 51.5%. Specificity is the proportion of children with only a transient delay who were predicted

Table 6. Results from logistic regression analyses, using selected 2-year measures as predictors and 3- or 4-year language status as outcome.

Sample/outcome variable/ predictors	Fit *df*, χ^2 [a]	% classified correctly	Sensitivity	Specificity	PPV	NPV
ELD/3-yr language status						
2-yr Vocabulary, Displaced Reference, PARCA	3, 31.2	59.6	38.5	76.2	56.1	61.1
Above + sex, mother's education	5, 38.6	60.3	42.0	74.8	56.9	61.9
ELD/4-yr language status						
2-yr Vocabulary, Displaced Reference, PARCA	3, 67.0	65.8	44.6	80.5	61.4	67.7
Above + sex, mother's education	5, 91.1	68.5	51.5	80.0	63.8	70.7
Above + ear infections at 18 mos, 3 & 4 yrs	8, 85.9	67.2	49.5	79.1	61.3	70.0
All/3-yr language status						
All predictors above	5, 1420.4	89.8	18.9	98.1	53.1	91.2
All/4-yr language status						
All predictors above	8, 939.0	89.5	19.4	98.3	58.1	90.7

Note. PPV = positive predictive value; NPV = negative predictive value. All accuracy of prediction measures are in percentages.

[a]All chi-square values significant at $p < .001$.

Table 7. Accuracy of prediction of 4-year language status within the ELD group from 2-year Vocabulary, Displaced Reference, PARCA, sex, and mother's education.

	Language in normal range at 4 years	Language impairment at 4 years	Row total
Language predicted in normal range at 4 years	273	113	386
Language impairment predicted at 4 years	68	120	188
Column total	341	233	574

to have normal development later; for the example, it was 273/341 = 80.5%. Positive predictive value is the proportion of children with a prediction of later delay who in fact had such a delay; for the example, it was 120/188 = 63.8%. (The term *positive* here refers to the category of special interest, namely, later language difficulty.) Negative predictive value is the proportion of children with a prediction of later normal development who in fact did have language in the normal range later; for the example, it was 273/386 = 70.7%. Although the specificity figures were relatively high, the other indices were all very low. In other words, these prediction equations fail to detect the majority of children who will have continuing language difficulties (low sensitivity), and a substantial minority of children who are predicted to "catch up" will not (low negative predictive value).

Because these analyses were based on a subsample with restricted range, and because some children who had difficulties at 3 and 4 were not in the ELD group, similar analyses were conducted on the entire sample, looking at 2-year predictors of language difficulties at 3 and 4. Although overall prediction was better due to the inclusion of the great majority of children who neither actually had nor were predicted to have difficulties, the sensitivity of prediction and the positive predictive value were even poorer (see final two rows of Table 6). That is, the prediction within the ELD group appeared to be somewhat better than across the entire distribution. Children outside the ELD group who had persisting difficulties at 3 or 4 were not picked up by the prediction equation (low sensitivity), and almost half of the children who were predicted to have later difficulties did not (low positive predictive value).

Parental Concern as an Alternative Outcome Measure at 3 and 4 Years

The analyses reported so far have looked at outcomes in terms of low scores on parental reports of vocabulary, grammar, and use of abstract language, primarily in expressive language. Because parents and others may be sensitive to delays and impairments in other areas, we explored an alternative indicator of outcome, the extent to which parents express concern about their child's language development. Parents who reported concern about slow language development or poor understanding in their child (see Appendix E) were coded as cases of parental concern. The size of these categories (for children in the ELD group) was comparable to those for the language measures: $N = 367$ for concern at 3 compared with 326 for the language measures; $N = 250$ for concern at 4 compared with 250 for the language measures. See our companion article (Bishop et al., 2003) for further discussion of the relation among these classifications. Logistic regression analyses were repeated using parental concern at 3 or 4 years as the outcome measure, and, for simplicity, just the full set of predictors identified above (see Table 8). The sensitivity of prediction, positive predictive value, and negative predictive value were comparable to the results of the previous analysis (see Table 6), whereas specificity was lower. That is, the major change is that an outcome of return to normal language

Table 8. Results from logistic regression analyses, using selected 2-year measures (Vocabulary, Displaced Reference, PARCA, sex, and mother's education) and relevant ear infection measures as predictors, and parental concern at 3 or 4 years as outcome.

Sample/outcome variable/ predictors	Fit *df*, χ^2 [a]	% classified correctly	Sensitivity	Specificity	PPV	NPV
ELD/3-yr parental concern	5, 38.8	56.7	59.5	53.6	57.6	55.5
ELD/4-yr parental concern	8, 43.7	62.4	42.3	77.7	59.0	64.0
All/3-yr parental concern	5, 896.2	82.7	10.8	98.5	60.5	83.4
All/4-yr parental concern	8, 519.7	84.7	4.6	99.4	57.8	85.1

Note. PPV = positive predictive value; NPV = negative predictive value. All accuracy of prediction measures are in percentages.

[a] All chi-square values significant at $p < .001$.

development was even less likely to be accurately predicted. When the analyses were repeated with the full sample (see final two rows of Table 8), again as in Table 6, sensitivity dropped considerably, but specificity and negative predictive value increased.

Prediction Within a More Stringently Defined Early Delay Group

In order to evaluate the possibility that prediction might be better within a more stringently defined group, the seven analyses of Tables 6 and 8 that focused on the ELD group were repeated using a criterion of 2-year vocabulary within the lowest 5% as defining delay. In none of these cases did the percentage of children correctly classified at 3 or 4 years change by more than 3% (see Table 9). As might be expected, sensitivity was improved by 10%–20%, but specificity decreased by nearly as great an amount, and positive and negative predictive values were approximately the same.

Discussion

This study confirmed findings from previous small-scale studies in showing that early language delay is a risk factor for language difficulties at 3 and 4 years. Of the 2-year-old ELD group, 44.1% had persisting language difficulties at 3 years (vs. 7.2% of the TL group), and 40.2% had persisting language difficulties at 4 years (vs. 8.5% of the TL group). These differences between the ELD and TL groups correspond to relative risk ratios greater than 4 for prediction of continuing language problems at 3 and 4 years. Also consistent with previous small-scale studies, however, was that a substantial proportion of children showed spontaneous improvement, with language development moving within the normal range by 3 or 4 years of age.

Given this pair of findings, we attempted to predict which ELD children would show persistent language problems and for which children language problems would resolve. We had anticipated that persistent difficulties might be related to smaller vocabulary size or other indicators of lower language and nonverbal ability at 2 years of age, consistent with the notion that persistent language delay involves children with more severe delays. With such a large sample size, even a very small effect will be statistically significant. And indeed, the relations between several 2-year measures and outcome, shown in Tables 4 and 5, were highly statistically significant, although the effect sizes were small, and the accuracy of prediction from 2-year measures was too poor to be of practical utility in discriminating persistent and transient difficulties. Adding the sociodemographic variables of sex and maternal education failed to substantially improve prediction. This failure of prediction is striking, because sex and maternal education are not only known to be correlated with rate of early language development overall (Fenson et al., 1994; Hart & Risley, 1995), but also are correlated with placement into the early delay group in the present study. In other words, boys are notably more likely to be delayed, but once in the delayed group, they are only slightly more likely to remain in it. Finally, a composite measure of ear infections at 18 months, 3 years, and 4 years failed to improve prediction.

Table 9. Results from logistic regression analyses, using a lowest 5% criterion for delay at 2 years.

Sample/outcome variable/ predictors	Fit *df*, χ^2 [a]	% classified correctly	Sensitivity	Specificity	PPV	NPV
ELD/3-yr language status						
2-yr Vocabulary, Displaced Reference, PARCA	3, 14.2	59.6	50.0	67.3	58.1	60.5
Above + sex, mother's education	5, 13.6	57.9	48.6	66.2	56.1	59.1
ELD/4-yr language status						
2-yr Vocabulary, Displaced Reference, PARCA	3, 51.4	67.1	63.9	70.0	65.6	68.3
Above + sex, mother's education	5, 64.5	69.8	66.0	73.1	68.7	70.6
Above + ear infections at 18 mos, 3 & 4 yrs	8, 57.9	68.2	66.1	70.0	66.1	70.0
ELD/3-yr parental concern						
All five predictors	5, 24.2	57.8	75.1	34.6	60.6	50.9
ELD/4-yr parental concern						
All five predictors	8, 32.4	63.3	61.8	64.8	63.3	63.3

Note. PPV = positive predictive value; NPV = negative predictive value. All accuracy of prediction measures are in percentages.

[a] All chi-square values significant at $p < .001$.

An analysis of the errors of prediction provided some additional clarification. As shown by the measures in Table 6, positive predictive value (PPV) was relatively low, in the range of 50%–60%. These figures reflect the finding previously discussed, that a substantial proportion of children with low language scores at 2 years do not in fact fall into the group with language difficulties at 4 years. Less expected, however, is the finding of very low sensitivity, which signifies that half or more of the children who will fall into the "difficulty" group later were not classified as delayed at 2 years. To some extent, this may be a statistical artifact of the methodology. Our classification of delay or other language difficulty was not based on an absolute cutoff, but on relative position within the sample (roughly, the lowest 10%). Thus, if some children move out of the extreme-low group, others must move into it. Inspection of individual cases confirms that although children with early delay often move into the average or even above-average range, those children who drop into the extreme-low group are most often just above it at 2 years. Nevertheless, all of these findings confirm the general lack of predictability based on early language delay.

Interestingly, prediction of outcome at 4 years, while still poor, was superior to prediction at 3, as shown by all measures in Table 6. The effect size figures in Table 4 suggest that vocabulary and nonverbal development have especially improved prediction, along with ear infection scores. Expressive vocabulary is an aspect of early language development that parents are highly capable of evaluating, as well as an aspect with continuing central significance to language. The PARCA assesses a diverse set of nonverbal abilities, and that breadth likely contributes to its predictive ability. It is likely that some portion of the 4-year-olds classified as showing persisting difficulties will move into the normal range in the next few years. Bishop and Edmundson (1987), for example, noted that nearly 40% of their sample of impaired 4-year-olds were in the normal range by age 5. The TEDS project is obtaining further assessment information on these children at 6 years, and the success of prediction to that age will be evaluated. As the classification of language impairment stabilizes at the end of the preschool period, prediction may improve.

Additional analyses, using the more stringent criterion of vocabulary in the lowest 5% at 2 years, demonstrated that the major findings of the study, namely, significant but relatively modest prediction from ELD status to later language status (less than 50%), and relatively poor prediction within the ELD group using additional measures, are consistent across criteria for early delay.

In interpreting these findings, it is important to note methodological factors that could influence results. First, this sample was unlike others in the literature, in that it was restricted to twins. Twinning is known to be associated with delay in early language milestones (Dale et al., 2000), raising the question of how far these results can be generalized to the general population. In fact, despite substantial differences in methodology, these results have many similarities to previous studies on single-born children. Exact comparison among various studies is not possible, because differing criteria for continuing delay have been used. Nevertheless, the proportion of late talkers whose difficulties resolved by 3 or 4 years in the present study (approximately 56% and 60%, respectively) can be compared to the results of several other studies. Rescorla, Mirak, and Singh (2000) followed 28 late talkers identified on the basis of low vocabulary by parent report at age 2. Between 24 and 31 months, 39% of the sample experienced a rapid vocabulary spurt; from Figure 3 of their article it appears that several additional children made a substantial advance by age 3. Whitehurst and Fischel (1994) identified 2-year-olds on the basis of low scores (2 *SD*s below the mean) on the Expressive One-Word Picture Vocabulary Test. At $3^1/_2$ years, 88% of the children had scores in the normal range on this test. In the domain of grammar, Rescorla, Dahlsgaard, and Roberts (2000) found that at 3 years, 41% of a group of late talkers had mean length of utterance values above the 10th percentile, and by 4 years, 71% did so. In Paul's (1996) study of toddlers identified as late talkers on the basis of vocabularies below 50 words by parent report between 20 and 34 months, grammatical development assessed with Developmental Sentence Score was in the normal range for 41% and 57% of the 3- and 4-year-olds, respectively. Thus the balance of persistent and transient problems in this sample of twins is quite similar to that observed in single-born children. Although the prognosis for children with early language delay is, on the whole, relatively good in this sample as it is in others, it is noteworthy that a substantial minority of late talkers did go on to have persisting problems at 4 years of age. Language delays in preschool twins are sometimes ignored by professionals, who assume they are a normal phenomenon. The current study cautions against such an attitude: Although many late-talking twins will show spontaneous improvement, the likelihood of persistent problems seems similar to that seen in single-born children.

A second point to consider is how far our findings are affected by reliance solely on parental report measures. Although most of these measures have been validated against other forms of assessments in other studies, it might plausibly be conjectured that their validity is most threatened in the very group we are most interested in, namely, parents of children with language impairments. We know there is a strong familial component to clinically significant specific language impairment, and so parents of children with language difficulties are,

on average, likely to be less articulate and literate than parents of unaffected children. However, as noted earlier, Thal et al. (1999) found that parents of children with early language delay provided information on the MacArthur CDI that had concurrent validity as high as that provided by parents of typically developing children. In current work with a subset of this sample, we are addressing this issue by considering how far parental report of language difficulties at 4 years of age is consistent with direct assessment of the child on a battery of language tests.

The primary language measures of this study—vocabulary, grammar, displaced reference, and use of abstract language—are measures of expressive language and not of receptive language. Previous studies (Thal & Katich, 1996) have suggested that receptive language level is relevant for prognosis, though far from sufficient for individual prediction. The MacArthur Communicative Development Inventories (Fenson et al., 1994), which provided the basis for our measures, include receptive measures only in the Words and Gestures instrument, designed for children between 8 and 16 months. By the middle of the second year of life, the child's receptive ability is so extensive that it is no longer possible for parents to monitor it accurately. Structured testing and/or observation is likely to be necessary in addition to parent report.

The poor prediction of outcome from 2-year measures is a disappointing result for those interested in identifying children at risk for continuing language difficulties. Although it is widely assumed that intervention is more effective when provided earlier rather than later, it has also been noted that it is inefficient, if not unethical, to provide speech-language treatment to young children whose problems are likely to resolve spontaneously (Paul, 2000). A parental report screening measure that could distinguish transient from persistent delays would greatly enhance the efficacy of preschool speech and language treatment. The fact that our measures did not fulfill this goal does not, of course, mean that it cannot be attained with improved parent report measures. Alternatively, it may be that a two-stage process will be most effective, with parent report identifying a high-risk group of children to be further screened professionally. It is also possible that the search for predictors could productively focus on the role of even more internal, organismic factors. We turn to genetic factors in our companion article (Bishop et al., 2003).

Acknowledgments

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 Research Fellowship.

References

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired four-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156–173.

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.

Colledge, E., Bishop, D. V. M., Koeppen-Schomerus, G., Price, T. S., Happé, F. G. E., Eley, T. C., et al. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology, 38*, 749–757.

Dale, P. S. (1996). Parent report assessment of language and communication. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (pp. 161–182). Baltimore: Paul H. Brookes.

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., Reznick, J. S., Thal, D. J., & Marchman, V. A. (2001). *A parent report measure of language development for three-year-olds.* Unpublished manuscript, University of Missouri–Columbia.

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.

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.

Eley, T. C., Dale, P. S., Bishop, D. V. M., Price, T. S., & Plomin, R. (2001). Longitudinal analysis of components of cognitive delay: Examining the aetiology of verbal and performance aspects of cognitive delay. *Journal of Educational Psychology*, *93*, 698–707.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5, Serial No. 242).

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.

Fischel, J. E., Whitehurst, G. J., Caulfield, M. B., & Debaryshe, B. (1989). Language growth in children with expressive language delay. *Pediatrics, 83*, 218–227.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brookes.

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.

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. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5*(2), 5–30.

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.

Paul, R., & Alforde, S. (1993). Grammatical morpheme acquisition in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech and Hearing Research*, *36*, 1271–1275.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research*, *45*, 360–371.

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

Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: Lexical development from 2;0 to 3;0. *Journal of Child Language, 27*, 293–311.

Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at 3. *Journal of Speech and Hearing Research*, *40*, 556–566.

Rescorla, L., & Schwartz, E. (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics, 11,* 393–407.

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.

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.

Thal, D. J., & Katich, J. (1996). Predicaments in early identification of specific language impairment: Does the early bird always catch the worm? In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (pp. 1–28). Baltimore: Paul H. Brookes.

Thal, D. J., O'Hanlon, L., Clemmons, M., & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 482–496.

Trouton, A., Spinath, F., & Plomin, R. (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition, and behaviour problems in childhood. *Twin Research, 5,* 444–448.

Viding, E., Colledge, E., Spinath, F. M., Price, T. S., Bishop, D. V. M., Dale, P. S., & Plomin, R. (in press). Genetic and environmental influence on language impairment in 4-year-old same-sex and opposite-sex twins. *Journal of Child Psychology and Psychiatry.*

Whitehurst, G. J., & Fischel, J. E. (1994). Early developmental language delay: What, if anything, should the clinician do about it? *Journal of Child Psychology and Psychiatry, 35,* 613–648.

Whitehurst, G. J., Fischel, J. E., Arnold, D. S., & Lonigan, C. J. (1992). Evaluating outcomes with children with expressive language delay. In S. F. Warren & J. E. Reichle (Eds.), *Causes and effects in communication and language intervention* (pp. 277–313). Baltimore: Paul H. Brookes.

Received December 7, 2001

Accepted December 17, 2002

DOI: 10.1044/1092-4388(2003/044)

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

Appendix A. Grammar rating at 4 years.

Parents were asked to select one of the following in response to the instruction, "On the whole, which of the following best describes the way your child talks?"

1. not yet talking
2. s/he is talking, but you can't understand him/her
3. talking in one-word utterances, such as "milk" or "down"
4. talking in 2 to 3 word phrases, such as "me got ball" or "give doll"
5. talking in fairly complete sentences, such as "I got a doll" or "can I go outside?"
6. talking in long and complicated sentences, such as "when I went to the park, I went on the swings," or "I saw a man standing on the corner."

Appendix B. Displaced reference at 2 years.

Each item is coded as "often," "sometimes," or "not yet."

1. Does your child ever talk about past events of people who are not present? For example, a child who saw a carnival last week might later say "carnival," "clown," or "band."
2. Does your child ever talk about something that is going to happen in the future, for example saying "choo-choo" or "aeroplane" before you leave the house on a trip, or saying "swing" when you are going to the park?
3. Does your child talk about objects that are not present, such as asking about a missing toy, referring to a pet out of view, or asking about someone not present?
4. Does your child understand if you ask for something that is not in the room? For example, would s/he go to the bedroom to get a teddy bear when you say "where's the bear"?
5. Does your child ever pick up or point to an object and name an absent person to whom the object belongs? For example, a child might point to Mummy's shoe and say "Mummy."

Appendix C. Abstract language at 3 years.

1. Does your child understand the meaning of "one"? If you ask for just one smartie or raisin, will your child give you only one and then stop?
2. Does your child ask questions (with more than one word) that begin with "what" or "where"?
3. Does your child ask questions (with more than one word) that begin with "why" or "how"?
4. Does your child give reasons for things, using the word "because"?
5. If you asked your child "What is a horse"?, could he answer "an animal"?
6. Can your child name simple shapes with the words "circle," "square," and "triangle"?
7. Does your child talk about things that "could" or "might" happen, such as "he could hurt himself if he's not careful"?
8. Does your child ever ask what a word means?
9. Can your child tell you which of two objects is larger if they were not present, for example when asked "which is bigger, a horse or a dog"?
10. Does your child know his/her right hand from his/her left hand?
11. Does your child use words that end in *-est* like "biggest" or "tallest"?
12. Can your child answer questions such as "what do you do when you are hungry?" or "what do you do when you are tired?" with answers that fit, such as "get food," "eat," "go to sleep," or "go to bed"?

Appendix D. Abstract language at 4 years.

1. Can your child say how old s/he is?
2. Can your child say the month and day of his/her birthday when asked?
3. Can your child tell you what happened at a past event (such as a birthday party or holiday) as if s/he were telling a story from beginning to end?
4. Can your child talk clearly about what s/he will do later on, such as tomorrow or next week?
5. Can your child tell a fairy tale, joke, or television show story completely from beginning to end and in the correct order?
6. Does your child know his/her right hand from his/her left?
7. Does your child use *-est* words, like "biggest," "strongest," or "greatest"?
8. Does your child use the word "today" correctly?
9. Does your child use the word "yesterday" correctly?
10. Does your child understand the difference between "accident" and doing something "on purpose"?
11. Does your child ever ask you what a word means?
12. Does your child use phrases or sentences containing "but"?
13. Does your child talk about the order of events by using words like "before" and "after"?
14. Does your child "play" with language by making jokes about words and their sounds, such as words that rhyme?

Appendix E. Parental concerns at 3 and 4 years.

Do you have any concerns about your child's speech and language? YES / NO

If YES, what are your concerns?

his/her language is developing slowly

it is hard for other people to understand him/her

s/he doesn't seem to understand other people

s/he pronounces words poorly

s/he doesn't hear well

s/he stutters

other

Appendix F. Communicative abnormality at 3 and 4 years.

1. Does your child use his/her index finger to show INTEREST in something, not just to ask for something (for example, pointing to an aeroplane, animal, or something on TV)?
2. If you suddenly look up at something interesting, does your child follow your line of gaze to see what caught your attention?
3. Does your child produce meaningless but fluent and "tuneful" speech, so that it sounds a bit like a foreign language?
4. Does your child like to "echo" what other people say (for example, if you say "where is your coat?" s/he might say "your coat," rather than answering the question)?
5. Does your child recite nursery rhymes or "jingles" from advertisements?
6. Has your child's language ever seemed to go into decline so that s/he no longer knows words that s/he used before?