Verbal and nonverbal predictors of early language problems: an analysis of twins in early childhood back to infancy*

BONAMY OLIVER

*Social, Genetic and Developmental Psychiatry Centre,
King's College, London*

PHILIP S. DALE

*Department of Communication Science and Disorders,
University of Missouri-Columbia, USA*

ROBERT PLOMIN

*Social, Genetic and Developmental Psychiatry Centre,
King's College, London*

ABSTRACT

We investigated infant precursors of low language scores in early childhood. The sample included 373 probands in 130 monozygotic (MZ) and 109 same-sex dizygotic (DZ) twin pairs in which at least one member of the pair scored in the lowest 15th percentile of a control sample on a general language factor derived from tester-administered tests at 4;6. From data at 2;0, 3;0 and 4;0 the antecedents of poor language performance at 4;6 for these probands were compared to 290 control children. As early as 2;0, language measures substantially predicted low-language status at 4;6, with predictions increasing at 3;0 and 4;0. Nonverbal cognitive development at 3;0 and 4;0 was nearly as predictive of low language at 4;6 as were the language measures. Behaviour problems were also significant predictors of low language status although the associations were only about half as strong. Bivariate genetic analyses indicated that these predictions are mediated by both genetic and shared environmental links.

INTRODUCTION

The definition and interpretation of language problems in early childhood can be greatly illuminated by examining prediction from infancy. There are several questions to be asked. For example, to what extent can language problems in early childhood be predicted by language problems in infancy? What other aspects of development can be predicted by language problems in infancy? Do other aspects of development in infancy – such as non-verbal cognitive development and behaviour problems – also predict language problems in early childhood? What are the genetic and environmental origins of these links between aspects of infant development and language problems in early childhood?

A strategy that has been widely used to address these issues is a 'prospective' analysis in which children with early language problems are followed longitudinally to examine factors in early development that may be predictors of their later problems. For example, in a previous report from the Twins' Early Development Study (TEDS), on which the present study is based, children with low vocabulary scores at 2;0 were followed at 3;0 and 4;0 (Dale, Price, Bishop & Plomin, 2003). Although fewer than half of the children with early low vocabulary scores met criteria for low-language at 3;0 and 4;0, as a group they performed less well than controls on language at 3;0 and 4;0. Importantly, children whose language delays persisted during this period were not necessarily those with the most severe initial difficulties (Dale *et al.*, 2003).

The results of this prospective analysis of poor vocabulary in infancy are similar to those of two other studies of late-talkers that showed that a large proportion of language delays in infancy are transient (Paul, 2000; Rescorla, 2002). Prospective studies of language problems that begin in early childhood rather than infancy also show some transience (Stothard, Snowling, Bishop, Chipchase & Kaplan, 1998; Johnson, Beitchman, Young, Escobar, Atkinson, Wilson, Brownlie, Douglas, Taback, Lam & Wang, 1999), although such language problems are more stable, for example, in predicting long-term difficulties in fluency (Boscolo, Ratner & Rescorla, 2002).

There is a substantial body of research that has aimed to determine which aspects of nonverbal development are predicted by early language difficulties. For example, the aforementioned TEDS study (Dale *et al.*, 2003) showed that children with low vocabulary scores at 2;0 performed almost as poorly on nonverbal cognitive measures at 3;0 and 4;0 as they did on language measures. Other studies have shown that by early childhood, language difficulties predict long-term difficulties such as reading problems (Catts, 1997) and other academic difficulties (Snowling, Adams, Bishop & Stothard, 2001), as well as behaviour problems such as school truancy, low self esteem and other adjustment problems (Hinshaw, 1992).

Much less is known about non-language predictors of childhood language problems. One study found that, of several measures gathered on late talkers at about 2;0, the only nonverbal predictor of expressive language outcome at around 8;0 was socioeconomic status (Paul & Fountain, 1999). Others suggested that non-language predictors of childhood language difficulties include behavioural style (Sajaniemi, Hakamies-Blomqvist, Maekelae, Avellan, Rita & von Wendt, 2001), middle ear disease (Lonigan, Fischel, Whitehurst, Arnold & Valdez-Menchaca, 1992), and the quality and quantity of external verbal input (Hammer, Tomblin, Zhang & Weiss, 2001).

In contrast to the prospective approach, another longitudinal analytic strategy, which could be called a 'retrodictive' analysis, examines the early antecedents of later language problems. The two strategies, although complementary, can yield different results. The major distinction between the two is that the prospective strategy includes many children whose early language problems are transient, whereas the targets of the retrodictive strategy are children with language problems later in development, some of whom may not have shown earlier problems. The retrodictive strategy used in the present study – which differs from a 'retrospective' approach in that the data are collected longitudinally, rather than using data collected retrospectively – can provide important information about antecedents of later language problems. We do not claim that it is entirely novel, as it is similar to the case-control design in medical research. Nevertheless, the strategy has rarely been used to examine precursors of early language development. One such study, which focuses on later childhood rather than infancy, supports the prospective findings mentioned previously in that language scores at 7;0, and to some extent nonverbal ability, were shown to be good predictors of language outcome at 11;0 (Botting, Faragher, Simkin, Knox & Conti-Ramsden, 2001).

We are unaware of research that addresses the genetic and environmental mediation of the links between infant development and language problems in early childhood. A genetic analysis of the prospective study by Dale *et al*. (2003) found modest genetic influence on both persistent and transient language delay for children with low vocabulary at 2;0 (Bishop, Price, Dale & Plomin, 2003). As mentioned previously, results could differ for a retrodictive analysis.

The present analysis had two main aims. First, using a large community sample of twins with language scores in the lowest 15th percentile at 4;6, we investigated language, nonverbal and behaviour problem predictors at 2;0, 3;0 and 4;0. Although at 4;6 there is still some recovery from language difficulties, as discussed, predictors of outcome are more stable by this age. Secondly, using a twin design, we examined the extent to which genetic and environmental influences mediate these associations. To this end, we developed new techniques that facilitate comparisons between phenotypic

and genetic analyses of both individual differences data and extremes. Specifically, in terms of individual differences, we developed a 'cross-twin' discriminant analysis that provides a genetic analogue to phenotypic discriminant analysis. In terms of the analysis of extremes, we developed a 'phenotypic group correlation', which provides a phenotypic analogue to the genetic extremes analysis.

METHOD

Sample and procedure

The sampling frame for the present study was the Twins' Early Development Study (TEDS), a study of 6963 pairs of twins born in the UK in 1994 and 1995 with data at 2;0, 3;0 and 4;0. Despite some attrition, the TEDS sample has been shown to be reasonably representative of the population. TEDS is described in full elsewhere (Trouton, Spinath & Plomin, 2002; Spinath, Ronald, Harlaar, Price & Plomin, 2003).

From the main study, a sub-sample of twins was selected from the parent data completed at 4;0 to be assessed at home on a battery of standard tests of language and non-language ability. All twin pairs selected for in-home testing were white to avoid issues with ethnic stratification in molecular genetic research also being carried out on this sample. English was their first and only language. After medical and perinatal exclusions, 300 twin pairs were selected so as not to fall into the extreme low ability groups (lowest 5%) for verbal and nonverbal ability, as informed by the parental assessments at 4;0; they comprised a representative 'control' sample. In addition, we selected a sample of 627 probands who were selected to be scoring in the lowest 5% on parental assessments of language at 4;0, and who scored at or below the level of those at the bottom 15% of the 'control' sample distributions on the general language factor from in-home testing at 4;6. These children are referred to as 'low-language' for the purposes of this report, although the authors acknowledge that this is not a clinically defined low-language sample. Nevertheless, to give some idea of the clinical relevance of the sample, by 4;0, 36·1% of the probands had seen a specialist for advice because of parental concerns about language, while 7·3% of the controls had been seen.

The representativeness of the control and low-language samples has been described in previous publications (Colledge, Bishop, Koeppen-Schomerus, Price, Happé, Eley, Dale & Plomin, 2002; Viding, Spinath, Price, Bishop, Dale & Plomin, 2004). Mothers in the control sample were found to be somewhat more highly educated than mothers reported in 1994 UK census data (45% vs. 34% A levels). However, in the control sample, the means and standard deviations (S.D.) on the McCarthy Scales of Children's Abilities (MSCA) were highly similar to MSCA norms (McCarthy, 1972) with

means of 49 vs. 50 respectively on the MSCA verbal index (*S.D.*s: 11 vs. 10) and means of 104 vs. 100 on the MSCA general cognitive index (*S.D.*s of 15 for both). As would be expected given the association between maternal education and children's language, mothers in the low-language sample were less well educated than mothers of the control children (28% vs. 45% A levels).

For the purposes of the present analysis, we excluded twin pairs selected to be in the 'control' group from parental assessments at 4;0 but who were found to reach criteria for the 'low language' group when assessed at 4;6. We also excluded opposite-sex twins because the bivariate DF extremes analysis technique (see Analyses section) allows for same-sex pairs only. Thus, after exclusions, the entire sample for the current analysis was 403 pairs of twins, 218 boy pairs and 185 girl pairs. The control sample consisted of 151 same-sex pairs of twins, 76 monozygotic (MZ) and 75 same-sex dizygotic (DZ) pairs. The age range for the control sample was 3;11 to 4;10 (mean 4;5). The 'low' sample consisted of 252 families (402 low-language probands), 134 MZ and 118 same-sex DZ pairs. The age range for the low group was 4;1 to 4;11 (4;6).

Measures

Predictors: Parental assessment at 2;0, 3;0 and 4;0. Near their second, third and fourth birthdays, the twins were assessed using parental measures of verbal and nonverbal ability and behaviour problems. Parents were asked to complete booklets which included an age appropriate version of the Parent Report of Children's Abilities which includes a parent report and a parent administered component (PARCA; Saudino, Dale, Oliver, Petrill, Richardson, Rutter, Simonoff, Stevenson & Plomin, 1998; Oliver, Dale, Saudino, Pike & Plomin, 2002) to assess nonverbal ability, and the UK short form version of the MacArthur Communicative Development Inventories (MCDI:UKSF; Dale, Simonoff, Bishop, Eley, Oliver, Price, Purcell, Stevenson & Plomin, 1998), which allowed parents to assess their children's vocabulary, grammar and semantic/pragmatic ability (Dale *et al.*, 2003). Behaviour problems were similarly assessed using the Revised Rutter Parent Scale for Preschool Children (RRPSPC, behaviour problems; Plomin, Price, Eley, Dale & Stevenson, 2002).

Reliability and validity of the language and nonverbal measures. Evidence for traditional indices of reliability such as internal consistency and test–retest reliability is limited for our measures, in part because their application and interpretation for parent report measures can be questioned (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994). However, long-term stability provides conservative estimates of test–retest reliability. The best available evidence for reliability of the measures is validity, since reliability sets an

upper bound for validity in the sense that reliability is a necessary but not sufficient condition for validity. That is, if a validity coefficient such as the correlation between a target test and another test is high, reliability must be at least as high as the validity coefficient.

For the MacArthur Communicative Development Inventory: Words and Sentences (MCDI:WS) on which the MCDI:UKSF at 2;0 in this study are based, validity results include a correlation of 0·73 with a standard tester-administered measure of expressive vocabulary with children aged 2;0 (reviewed in Fenson *et al.*, 1994) and 0·85 with a sample of language-impaired children at 3;0 (Thal, O'Hanlon, Clemmons & Fralin, 1999). The 100 vocabulary items and 12 grammar items used in the present study were selected to have high predictive validity (above 0·90 for both) to the corresponding full list of the MCDI:WS (see Fenson, Pethick, Renda, Cox, Dale & Reznick, 2000, for vocabulary; Dionne, Dale, Boivin & Plomin, 2003, for grammar). Moreover, in a sample of 107 children (all twins or triplets), we have previously found a correlation of 0·58 between the MCDI:UKSF and the language subscale of the Bayley Scales of Infant Development – II Language scale at 2 years (Saudino *et al.*, 1998).

The language measures at 3;0 and 4;0 were developed for TEDS as an extension of the MCDI. In a study of 85 British children at 2;8–3;4, the correlations between the vocabulary measure at 3;0 administered by mail and the McCarthy Scales of Children's Abilities Verbal Score administered by testers in the home were 0·68 and 0·48 for first- and second-born twins respectively (birth order within the twin pair) (Oliver *et al.*, 2002). The design similarity of the measure at 3;0 to the measure at 2;0, and the substantial stability of vocabulary and grammar scores from 2;0 to 3;0 (Dionne *et al.*, 2003), suggest face validity for the specific measures at 3;0.

The language measure at 4;0 is the newest measure. Evidence for the validity of the measure for selecting children with low language development is provided by the results of two related TEDS studies. A representative sub-sample of 600 twins (300 pairs) from the TEDS sample was administered nine language measures in the children's home. The nine measures were aggregated into a composite and standardized, which yielded a strong general factor (Colledge *et al.*, 2002). The current sub-sample of TEDS assessed at home were selected from the lowest 5·7% on the basis of parent report on the language measure at 4;0, using a composite of the vocabulary, grammar and semantic/pragmatic measures. For the twins thus selected for low parent-reported language, the mean of the tester-administered composite (before the additional exclusions for the current study, but after medical and perinatal exclusions) was −1·29, approximately the 12th 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.

Similar validity information exists for the nonverbal measure. A correlation of 0·55 was obtained between the PARCA at 2;0 administered by post and the Mental Development Index of the Bayley Scales of Infant Development-II administered in home by testers for a group of 107 twins at 2;0 (Saudino *et al.*, 1998). The correlation rose to 0·66 when the vocabulary at 2;0 and grammar scores were also used as predictors. This validity coefficient is more appropriate, because the PARCA was specifically designed to assess non-verbal skills, whereas the Bayley is a broad measure of cognitive development. Similarly, PARCA at 3;0 scores obtained by mail were compared with in-home testing on the McCarthy Scales of Children's development, and yielded a correlation of 0·46 (Oliver *et al.*, 2002). The correlation rose to 0·63 when vocabulary at 3;0 was also used as a predictor.

Outcome: In-home testing at 4;6. Two testers visited each family so that members of the twin pairs could be tested simultaneously. The children were assessed on a diverse battery of nonverbal and verbal tests. The verbal measures included three subtests (word knowledge, verbal fluency, and opposite analogies) from the Verbal Index of the McCarthy Scales of Children's Abilities (MSCA; McCarthy, 1972), Action Pictures tasks from the Renfrew Language Scales, (Renfrew, 1997*a*), Verbal Comprehension sub-test from the British Ability Scales (Elliot, Smith & McCulloch, 1996), the Bus Story from the Renfrew Language Scales (Renfrew, 1997*b*), Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986), Non-Word Repetition (Gathercole, Willis, Baddeley & Emslie, 1994), and a Phonological Awareness task (Bird, Bishop & Freeman, 1995). Results of a principal component analysis indicated a strong general language factor that accounted for 41% of the variance of the language measures. A composite language factor score was created by averaging across z-scores of each language subtest. These measures are described in more detail elsewhere (Colledge *et al.*, 2002).

Thus, we used each of the five parental measures at 2;0, 3;0 and 4;0 as predictors of later language ability as defined by the composite language factor score at 4;6.

Analyses

All measures were corrected for the effects of sex and age at testing by regressing out these effects and using standardized residuals. As socio-economic status (SES) is known to correlate with language ability as well as nonverbal cognitive ability and behaviour problems, we repeated all analyses correcting for age, sex and an SES composite. The results remained very similar. For this reason, and because regressing out SES significantly reduces the variance we can explain, we report data for the sex and age correction only. The data from booklets at 2;0, 3;0 and 4;0 were standardized on the

basis of the entire TEDS sample. For all of the language and nonverbal cognitive measures, lower scores indicate children with more difficulties; for the behavioural problems measures, higher scores indicate more problems.

Phenotypic analyses. To give an overview of our results, and to provide an indication of how well the predictor variables from data at 2;0, 3;0 and 4;0 discriminate between the low-language and control groups, a discriminant function analysis was first conducted. This analysis generates a discriminant function based on linear combinations of the predictor variables that give the best discrimination between the two groups. In addition, the analysis provides a canonical correlation, the square of which represents the proportion of between-group variance explained by the predictors.

However, the main phenotypic (non-genetic) comparison involved the mean scores of the low-language probands at 4;6 at 2;0, 3;0 and 4;0. An innovative index, called the phenotypic 'group' correlation (PGC), indicates the magnitude of the phenotypic relationship between the selection variable (in this case being a low-language proband at 4;6) and the earlier quantitative variables obtained at 2;0, 3;0 and 4;0. An advantage of the PGC as an index of longitudinal stability is that it can be decomposed in a very direct way into genetic and environmental components of variance, as described below. The PGC is calculated by dividing the standardized proband mean for the measures at 2;0, 3;0 and 4;0 by the standardized proband mean score for the selection variable. For example, as illustrated in Figure 1, if the proband *mean* $(P(L))$ is two standard deviations (*S.D.*) below the sample mean on the language factor at 4;6 and only one *S.D.* below the mean on vocabulary at 2;0 $(P(V))$, the PGC is 0·5. A PGC of 0·5 indicates that the mean vocabulary score at 2;0 for the low-language probands selected at 4;6 regresses halfway back to the population mean. If the probands are also two *S.D.* below the mean on vocabulary, the PGC is 1·0. If the probands regress all the way back to the population mean on vocabulary, the PGC is 0·0.

Genetic analyses. We conducted two types of analyses in order to examine the extent to which the phenotypic associations are mediated by genetic, shared environmental and non-shared environmental influences on the retrodictive prediction from data at 2;0, 3;0 and 4;0 of low language status at 4;6. Firstly, we conducted a novel type of discriminant analysis using cross-twin canonical correlations, in which the discrimination between the low-language twins and the control twins was made, not from their own data at 2;0, 3;0 and 4;0, but rather from their co-twins' scores. The MZ cross-twin canonical correlations can be compared to those for DZ pairs in order to estimate genetic and environmental influences on the overall discrimination of the low-language and control twins from the data at 2;0, 3;0 and 4;0.

The second and primary genetic analyses for the current study involved DF extremes analysis (DeFries & Fulker, 1988) extended to the bivariate
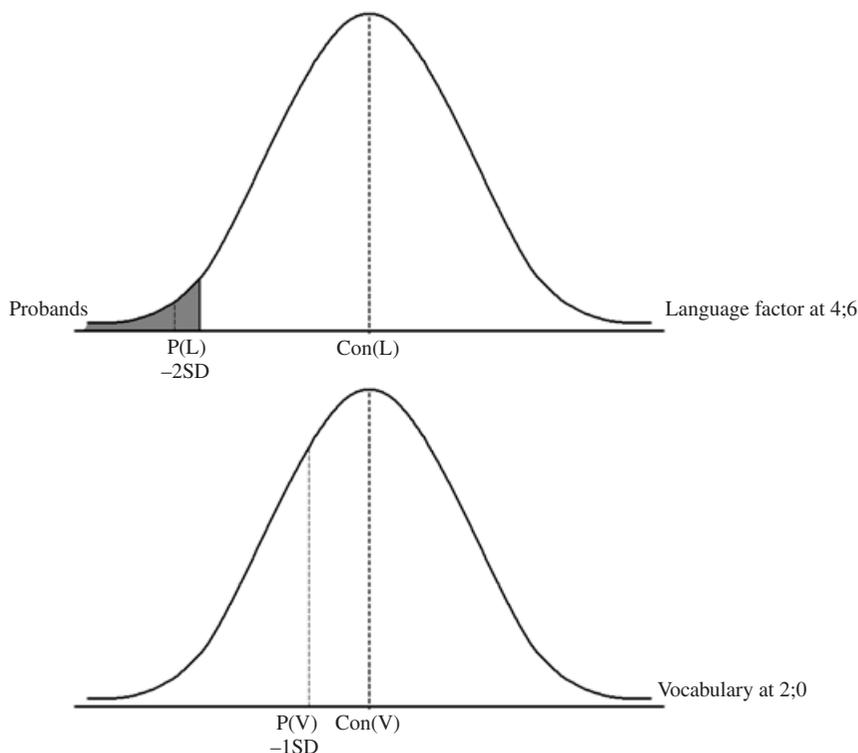
Fig. 1. Representation of the phenotypic group correlation (PGC) between low language (L) at four-and-a-half years and vocabulary (V) at two years. PGC is the ratio of the proband mean for vocabulary ($P(V)$) to the proband mean for the language factor ($P(L)$). Con= control mean.

condition (Gillis, Gilger, Pennington & DeFries, 1992). Bivariate DF extremes analysis is based on cross-trait cross-twin 'group' correlations (CGC) which are similar to the PGC described above, except that they compare probands and their co-twins. That is, the CGC divides the standardized co-twin vocabulary mean ($C(V)$) at 2;0, 3;0 and 4;0 by the standardized proband vocabulary mean ($P(V)$). Figure 2 illustrates a hypothetical example in which the CGC is 0·5 because the co-twins of the low-language probands at 4;6 have a mean vocabulary score of −0·5, half as great as the probands' mean vocabulary score (−1·0). If the co-twins were not at all similar to the probands, their vocabulary scores would regress to the population mean (0·0) and thus the CGC would be 0·0.

Genetic influence is suggested to the extent that the MZ CGC exceeds the DZ CGC. The effect size indicator, bivariate 'group heritability' (h^2g), can be estimated by doubling the difference between the MZ CGC and the
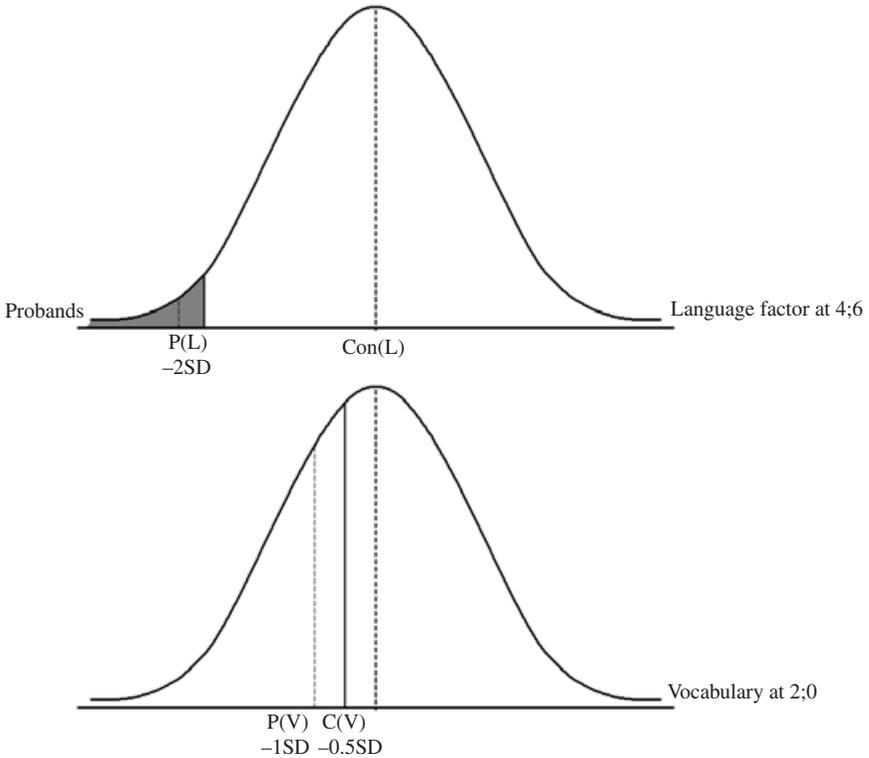
Fig. 2. Representation of the cross-trait cross-twin group correlation (CGC) for vocabulary at two years. CGC is the ratio of the cotwin mean for vocabulary ($C(V)$) to the proband mean for vocabulary ($P(V)$).

DZ CGC. In practice, genetic and environmental parameters are estimated from DF extremes analysis from a regression-based model represented as the regression, $C(V) = B_1 P(L) + B_2 R + A$, where the co-twin's scores at 2;0, 3;0 and 4;0 (e.g. $C(V)$) are predicted from the proband mean score on the language factor ($P(L)$), and the coefficient of relatedness (R) which is 1·0 for MZ and 0·5 for DZ twins, since MZ twins are genetically identical while DZ twins share on average half of their genes. The regression weight B_2 estimates bivariate group heritability. Bivariate group heritability is a measure of the extent to which the mean difference between the low-language (4;6) probands and the population on the variables at the earlier years can be explained by genetic factors. In other words, bivariate group heritability indicates how much of the longitudinal relationship (PGC) between low-language status at 4;6 and the earlier measures can be explained by genetic factors. Similarly, we can estimate the extent to which the mean

TABLE 1. *Discriminant function analysis discriminating low-language twins and control twins from measures of language, cognition and behaviour problems at 2;0, 3;0 and 4;0*

Age	Canonical correlation	Wilks' Lambda	Chi-square	df	*p*
2;0	0·60	0·64	218·82	5	<0·001
3;0	0·61	0·63	186·51	5	<0·001
4;0	0·68	0·54	430·77	5	<0·001

difference between the probands and the rest of the distribution can be attributed to group shared environmental influences (c^2g) by subtracting the group heritability from the MZ CGC. The DF extremes analysis yields standard errors for the estimates of group heritability and shared environmental influences.

RESULTS

Discriminant function analysis

Phenotypic analysis. A discriminant function analysis was conducted to describe the joint prediction of all of the measures at 2;0, 3;0 and 4;0 in discriminating low-language and control twins. The results from this analysis are shown in Table 1.

The phenotypic canonical correlations indicate that there is significant and substantial prediction of the differences between low-language and control groups from the data at 2;0, 3;0 and 4;0. The overall prediction from 2;0 is only marginally less than the prediction from 3;0 and 4;0. The accuracy of prediction is moderate from data at 2;0, 3;0 and 4;0, and increases somewhat, as expected, with age; 76·6% of original grouped cases are correctly classified at 2;0, 77·6% at 3;0 and 82·3% at 4;0.

Genetic analyses. Table 2 presents cross-twin canonical correlations (see Analyses). The MZ cross-twin canonical correlations at 2;0, 3;0 and 4;0 shown in Table 2 are nearly as great as the phenotypic canonical correlations shown in Table 1. Since MZ twins are genetically identical, the fact that the cross-twin canonical correlation for MZ twins is greater than that of the DZ twins is evidence for genetic influence on the discrimination between the low-language group and the control group from measures at 2;0, 3;0 and 4;0. A rough index of the extent to which genetic factors mediate the phenotypic canonical correlation can be estimated by doubling the difference between the MZ and DZ cross-twin canonical correlations: 0·10 at 2;0, 0·18 at 3;0 and 0·44 at 4;0. Dividing these genetic contributions by the phenotypic canonical correlations indicates the proportion of the phenotypic canonical correlation accounted for genetically. This ratio

TABLE 2. *Cross-twin discriminant function analysis discriminating low-language twins and control twins from co-twins' scores on measures of language, cognition and behaviour problems at* 2;0, 3;0 *and* 4;0

	MZ					DZ				
Age	Canonical correlation	Wilks' Lambda	Chi-square	df	*p*	Canonical correlation	Wilks' Lambda	Chi-square	df	*p*
2;0	0·57	0·68	103·15	5	<0·001	0·52	0·73	70·31	5	<0·001
3;0	0·57	0·67	83·88	5	<0·001	0·48	0·77	49·06	5	<0·001
4;0	0·65	0·58	201·96	5	<0·001	0·43	0·81	65·96	5	<0·001

TABLE 3. *Discriminant function analysis: standardized canonical discriminant function coefficients*

Age	Measure	Standardized canonical discriminant function coefficient
2;0	Vocabulary	0·48
	Grammar	−0·18
	Semantic/pragmatic language	0·47
	PARCA	0·36
	Behaviour problems	−0·22
3;0	Vocabulary	0·45
	Grammar	0·28
	Semantic/pragmatic language	0·18
	PARCA	0·32
	Behaviour problems	−0·10
4;0	Vocabulary	0·23
	Grammar	0·26
	Semantic/pragmatic language	0·30
	PARCA	0·50
	Behaviour problems	−0·08

can be interpreted as a 'canonical heritability', and it increases steadily with age, estimated as 0·17, 0·30 and 0·65 at 2;0, 3;0 and 4;0 respectively.

Although the discriminant function analyses provide information about the joint prediction of the measures at each age, we are also interested in the phenotypic discrimination offered by the individual measures.

Table 3 presents the standardized canonical discriminant function coefficients, which indicate the independent prediction of each variable in the company of all the other variables at each age. These coefficients indicate that, at 2;0, the prediction seems to be driven by semantic/pragmatic language and vocabulary. At 3;0, vocabulary seems to take charge, and by 4;0 the PARCA is the driving force of the prediction.

TABLE 4. *Bivariate DF extremes analysis: proband means, standard deviations* (S.D.) *and phenotypic group correlations* (*PGC*) *for all measures*

Measure	Age	All probands Mean (*S.D.*) (*N* = 145–373)	All probands PGC	MZ mean (*S.D.*) (*N* = 95–219)	MZ PGC	DZ mean (*S.D.*) (*N* = 50–154)	DZ PGC
Language composite	4;6	−2·18 (0·87)		−2·18 (0·80)		−2·18 (0·97)	
Vocabulary	2;0	−1·02 (0·73)	0·47	−1·02 (0·77)	0·47	−1·03 (0·66)	0·47
	3;0	−1·37 (1·01)	0·63	−1·34 (1·02)	0·61	−1·42 (1·01)	0·65
	4;0	−1·58 (1·46)	0·72	−1·60 (1·38)	0·73	−1·55 (1·56)	0·71
Grammar	2;0	−0·68 (0·83)	0·31	−0·67 (0·84)	0·31	−0·69 (0·83)	0·32
	3;0	−1·19 (0·89)	0·55	−1·21 (0·84)	0·56	−1·16 (0·97)	0·53
	4;0	−1·33 (1·14)	0·61	−1·36 (1·13)	0·62	−1·29 (1·16)	0·59
Semantic/ language pragmatic	2;0	−1·05 (0·89)	0·48	−1·09 (0·89)	0·50	−1·00 (0·88)	0·46
	3;0	−1·23 (0·97)	0·56	−1·29 (0·97)	0·59	−1·15 (0·99)	0·53
	4;0	−1·60 (1·17)	0·73	−1·70 (1·17)	0·78	−1·46 (1·15)	0·67
PARCA	2;0	−0·80 (0·95)	0·37	−0·88 (0·91)	0·40	−0·67 (1·01)	0·31
	3;0	−1·12 (1·09)	0·51	−1·12 (1·03)	0·51	−1·12 (1·17)	0·51
	4;0	−1·48 (1·05)	0·68	−1·55 (1·01)	0·71	−1·37 (1·10)	0·63
Behaviour problems	2;0	0·45 (1·06)	−0·21	0·30 (1·03)	−0·14	0·67 (1·07)	−0·31
	3;0	0·56 (1·09)	−0·26	0·46 (1·06)	−0·21	0·70 (1·14)	−0·32
	4;0	0·67 (1·10)	−0·31	0·49 (1·00)	−0·22	0·91 (1·18)	−0·42

Although the cross-twin versions of these coefficients for the individual tests are not suitable for estimating genetic and environmental contributions to the discrimination between the low-language and control groups, we can use a well documented behavioural genetic technique to do this, DF extremes analysis.

DF extremes analysis

Phenotypic analysis. As an initial phenotypic comparison for DF extremes analysis, we calculated phenotypic group correlations as described above (see Analyses).

The proband means and *S.D.* for each of the language measures are presented in Table 4. The mean standardized score of the probands was −2·18 on the language composite measure at 4;6, the measure on which they were selected. The mean vocabulary score of the probands at 2;0 was −1·02. Thus, the probands, who were about 2 *S.D.*s below the mean on the language composite at 4;6, were about 1 *S.D.* below the mean on the vocabulary score at 2;0. This finding is reflected in the phenotypic group correlation (PGC) of 0·47.

For the language measures, PGCs are substantial; the lowest PGC for all probands is for grammar at 2;0 (0·31) and the highest for abstract language at 4;0 (0·73). Without exception, these correlations indicate an increase in

TABLE 5. *Bivariate DF extremes analysis: co-twin means, standard deviations* (S.D.) *and cross-twin group correlations* (CGC) *for all measures*

Measure	Age	All co-twins Mean (S.D.) (N=53–105)	CGC	MZ mean (S.D.) (N=95–219)	MZ CGC	DZ mean (S.D.) (N=50–154)	DZ CGC
Language composite	4;6	−1·81 (1·12)	0·83	−1·98 (1·00)	0·91	−1·60 (1·23)	0·73
Vocabulary	2;0	−0·95 (0·75)	0·93	−0·98 (0·78)	0·96	−0·91 (0·70)	0·88
	3;0	−1·22 (1·07)	0·89	−1·29 (1·05)	0·96	−1·12 (1·09)	0·79
	4;0	−1·43 (1·44)	0·91	−1·55 (1·35)	0·97	−1·28 (1·53)	0·82
Grammar	2;0	−0·62 (0·87)	0·91	−0·64 (0·85)	0·96	−0·59 (0·91)	0·86
	3;0	−1·03 (0·93)	0·87	−1·15 (0·89)	0·95	−0·87 (0·96)	0·75
	4;0	−1·19 (1·19)	0·89	−1·31 (1·15)	0·96	−1·04 (1·22)	0·81
Semantic/ pragmatic language	2;0	−1·00 (0·91)	0·95	−1·03 (0·92)	0·94	−0·96 (0·89)	0·96
	3;0	−1·15 (0·99)	0·93	−1·25 (0·98)	0·97	−1·00 (0·99)	0·87
	4;0	−1·43 (1·25)	0·89	−1·65 (1·20)	0·97	−1·13 (1·26)	0·77
PARCA	2;0	−0·71 (0·94)	0·89	−0·85 (0·88)	0·97	−0·51 (0·98)	0·76
	3;0	−1·01 (1·07)	0·90	−1·09 (1·01)	0·97	−0·89 (1·15)	0·79
	4;0	−1·34 (1·06)	0·91	−1·54 (0·99)	0·99	−1·07 (1·11)	0·78
Behaviour problems	2;0	0·37 (1·05)	0·80	0·27 (1·01)	0·90	0·52 (1·08)	0·78
	3;0	0·46 (1·06)	0·82	0·42 (1·05)	0·91	0·51 (1·08)	0·73
	4;0	0·55 (1·03)	0·82	0·47 (0·99)	0·96	0·65 (1·08)	0·71

predictiveness from 2;0 to 3;0 to 4;0. For the nonverbal measure (PARCA), PGCs are almost as great as for the language measures, ranging from 0·37 at 2;0 to 0·68 at 4;0, and they too show a steady increase in predictiveness with age. PGCs are lower for behaviour problems as expected, but are still noteworthy, increasing in magnitude from −0·21 at 2;0 to −0·31 at 4;0. The PGCs for behaviour problems are negative because, unlike the other measures, a high score on the behaviour problems measure indicates problems, but they also show an increasing pattern of predictiveness from 2;0 to 3;0 to 4;0. PGCs were also generated separately for MZ and DZ twins in order to compare them to the MZ and DZ co-twin means and S.D.s as discussed later in relation to Table 5. Although there were slight differences in PGCs for MZ and DZ twins because of MZ–DZ mean differences on these variables, the pattern of predictiveness is similar for MZ and DZ probands.

Genetic analyses. Table 5 presents co-twin means and cross-trait cross-twin group correlations (CGCs) for MZ and DZ co-twins which are used in DF extremes analysis to estimate genetic and environmental mediation of the links between low-language scores at 4;6 and earlier development. As described earlier, the CGC is the ratio of the co-twin mean on the earlier quantitative measure to the low-language proband's mean on the same measure. It is identical to the transformed co-twin mean in DF analysis and

can be interpreted like other MZ and DZ twin correlations although it refers to mean group differences rather than individual differences (Plomin, DeFries, McClearn & McGuffin, 2001).

For example, Table 5 shows that the MZ co-twin mean for vocabulary at 2;0 was −0·98, almost as low as the probands' mean of −1·02 (Table 4). In other words, the vocabulary score at 2;0 for the MZ co-twins of the probands is nearly as low as the probands themselves. The corresponding CGC for MZ twins (the transformed co-twin mean in the DF analysis) is 0·96 (i.e. −0·98/−1·02), indicating substantial MZ twin resemblance. Although the DZ proband mean for vocabulary at 2;0 is only slightly lower than the MZ proband mean (see Table 4), the ratio of the DZ co-twin mean to the DZ proband mean yields a DZ CGC of 0·88, indicating somewhat less resemblance for DZ twins than for MZ twins. Doubling the difference between the MZ CGC of 0·96 and the DZ CGC of 0·88 suggests a bivariate group heritability of 0·16. As explained in the Analyses section, this bivariate group heritability estimate indicates that 16 percent of the mean difference between low-language probands and vocabulary at 2;0 is mediated by genetic factors. Bivariate shared environmental influence can be estimated as MZ twin resemblance that is not accounted for by heritability, that is 0·96−0·16=0·80. In other words, the mean vocabulary difference between the low-language probands and the population is to some extent due to genetic factors but is largely due to shared environmental influences.

Table 5 shows that the MZ CGCs are all extremely high, above 0·90 for all measures, whereas the DZ CGCs are lower, from 0·71 to 0·96. This pattern of results suggests moderate bivariate group heritability and substantial bivariate shared environment. Across the three language measures, the average MZ and DZ CGCs are, respectively, 0·96 and 0·83 at 2;0, 0·96 and 0·81 at 3;0, and 0·96 and 0·87 at 4;0. Doubling these differences between MZ and DZ CGCs yields estimates of bivariate group heritability of 0·26 at 2;0, 0·30 at 3;0 and 0·18 at 4;0. Comparing the MZ and DZ CGCs for PARCA suggests bivariate group heritabilities of 0·42 at 2;0, 0·36 at 3;0 and 0·42 at 4;0. Thus the overall magnitude of genetic mediation of the nonverbal measure's prediction of low language is similar to the prediction from language measures at 3;0, but at 2;0 and 4;0 the PARCA's prediction appears to be more heritable than the prediction from the language measures.

Finally, the total behaviour problems measure yields bivariate group heritabilities of 0·24 at 2;0, 0·36 at 3;0, and 0·50 at 4;0. That is, although the PGCs for behaviour problems are only half those for the language and cognitive measures, the longitudinal associations with behaviour problems are similarly mediated genetically at 2;0 and 3;0 and more strongly mediated genetically at 4;0.

TABLE 6. *Bivariate DF extremes analysis: model-fitting results*

	Age	R	SE	h^2g	SE	p	c^2g	SE	p
Vocabulary	2;0	0·18	0·14	0·18	0·17	0·152	0·79	0·20	<0·001
	3;0	0·38	0·13	0·38	0·17	0·013	0·59	0·29	0·021
	4;0	0·33	0·16	0·33	0·20	0·051	0·64	0·32	0·023
Grammar	2;0	0·22	0·27	0·22	0·27	0·208	0·74	0·25	0·002
	3;0	0·40	0·21	0·40	0·21	0·026	0·56	0·30	0·032
	4;0	0·35	0·16	0·35	0·16	0·016	0·62	0·26	0·009
Semantic/pragmatic language	2;0	0·00	0·16	0·00	0·20	0·542	0·97	0·25	<0·001
	3;0	0·35	0·18	0·35	0·23	0·061	0·64	0·33	0·025
	4;0	0·39	0·13	0·39	0·17	0·010	0·59	0·30	0·026
PARCA	2;0	0·40	0·24	0·40	0·30	0·091	0·56	0·27	0·020
	3;0	0·37	0·16	0·37	0·21	0·039	0·60	0·30	0·021
	4;0	0·39	0·11	0·39	0·14	0·003	0·59	0·23	0·005
Behaviour problems	2;0	0·19	0·54	0·19	0·69	0·389	0·68	0·30	0·011
	3;0	0·35	0·37	0·35	0·47	0·228	0·57	0·28	0·023
	4;0	0·52	0·28	0·52	0·36	0·072	0·44	0·23	0·027

Note: R = overall multiple regression. SE = standard error. h^2g = bivariate heritability. c^2g = bivariate shared environment.

As described in the Analyses section, DF extremes analysis provides a model-fitting approach that yields standard errors of estimate for genetic and environmental 'group' parameters. As shown in Table 6, model-fitting results of the DF extremes analysis are very similar to the results inferred from the comparison of MZ and DZ CGCs in Table 5. For example, the bivariate group heritability estimate for vocabulary at 2;0 obtained by DF extremes analysis is 0·18, similar to the estimate of 0·16 derived by doubling the difference between the MZ and DZ CGCs in Table 5. The standard error of this estimate is 0·17 and its p value is 0·152. The shared environmental estimate in Table 6 is 0·79 (SE = 0·20, p = 0·000), again similar to the estimate of 0·80 derived from the CGCs in Table 5. The bivariate group heritability and shared environmental estimates for the other measures, shown in Table 6, are also closely matched to those suggested by the respective CGCs in Table 5.

The results in Table 6 indicate that the PGCs between low language status at 4;6 and the measures at 2;0, 3;0 and 4;0 are due to both genetic and shared environmental influences. Across the three language measures, the average bivariate group heritability estimates are 0·13 at 2;0, 0·38 at 3;0 and 0·36 at 4;0. Results are roughly similar for the vocabulary and grammar. For semantic/pragmatic language the results are similar at 3;0 and 4;0, although the bivariate group heritability estimate at 2;0 is 0·00. However, the average bivariate group heritability estimates across the three years for each of the language measures are similar: 0·30 for vocabulary, 0·32 for grammar, and 0·25 for semantic/pragmatic language.

For the nonverbal measure, bivariate group heritabilities were equally strong: 0·40 at 2;0, 0·37 at 3;0 and 0·39 at 4;0. That is, low language status at 4;6 is at least as related genetically to low nonverbal cognitive development at 2;0, 3;0 and 4;0 as they are to low-language scores at 2;0, 3;0 and 4;0.

For behaviour problems at 2;0, 3;0 and 4;0, even though the PGCs in Table 4 are lower than for the language and cognitive measures, bivariate group heritabilities are similar at 2;0 (0·19) and at 3;0 (0·35) and somewhat greater at 4;0 (0·52). In other words, although behaviour problems at 2;0, 3;0 and 4;0 are not as strongly predictive phenotypically of low language at 4;6, the associations are also moderately mediated genetically.

Bivariate group shared environmental estimates in Table 6 are substantial across the language measures, 0·67, 0·64 and 0·73 at 2;0, 3;0 and 4;0 respectively. Bivariate group shared environment is also substantial for PARCA (0·58 across the three ages) and for behaviour problems (0·56).

Figure 3 summarizes visually these model-fitting estimates of bivariate heritability and bivariate shared environment (Table 6) in relation to the PGCs (Table 4). The PGCs indicate the extent to which low language at 4;6 can be predicted by measures at 2;0, 3;0 and 4;0. The bivariate heritability and shared environment estimates indicate the extent to which these PGCs are mediated by genetic factors and by shared environment. The figure shows that low-language status at 4;6 is increasingly predicted by language measures at 2;0, 3;0 and 4;0 and that this prediction is mediated by genetic and shared environmental factors.

DISCUSSION

The present study had two main objectives. First, we used a large community twin sample to examine the longitudinal relationships between low-language status at 4;6 and verbal abilities, nonverbal abilities and behaviour problems at 2;0, 3;0 and 4;0. Our results indicate that low language scores at 4;6 can be predicted by language measures as early as 2;0. This is consistent with previous studies that show that language delay in infancy is a risk factor for later language impairment. Moreover, vocabulary, grammar, and semantic/pragmatic language are all important developmentally. Other studies have found vocabulary to be especially important in prognosis (Fischel, Whitehurst, Caulfield & DeBaryshe, 1989), and the results at 2;0 and particularly 3;0 support this, as well as highlighting the importance of semantic/pragmatic language at 2;0 and non-verbal development (PARCA) as the child gets older (4;0).

The results confirm those found in the previous prospective analysis of the TEDS data at 2;0, 3;0 and 4;0 (Dale *et al.*, 2003). Unlike results from the prospective study, however, which are somewhat disappointing in terms
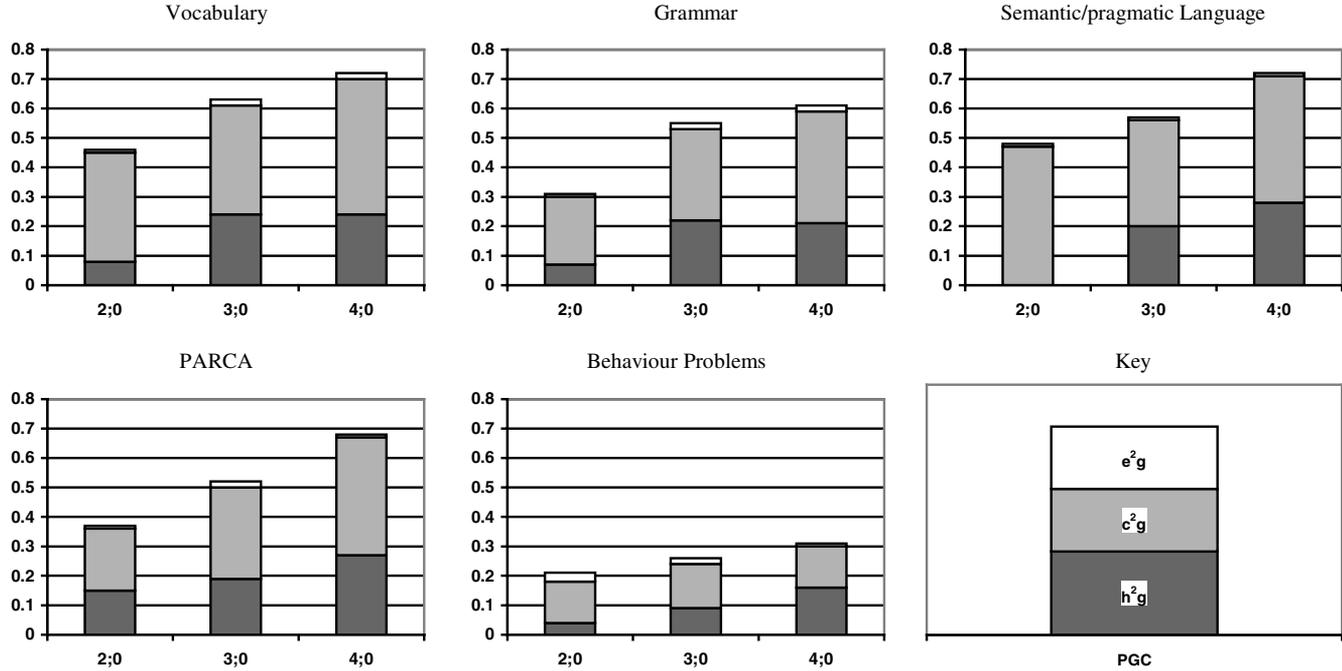
Fig. 3. Representation of the model-fitting estimates of bivariate heritability and bivariate shared environment (Table 6) in relation to the PGCs (Table 4).

of prediction because so many children recover from their early language difficulties, the present findings provide stronger evidence for infant antecedents for those children with problems in early childhood. Moreover, the prediction increases from 2;0 to 3;0 to 4;0. The longitudinal nature of TEDS will allow us to continue to look at the strength of prediction as the children age and language difficulties become more stable and predictive of long-term problems in spoken and written language.

What may be more surprising about our results is the finding that the phenotypic association between having later low language and earlier nonverbal cognitive ability as assessed by the PARCA is very nearly as strong as the association with language ability. Indeed, although it may be argued that no nonverbal ability test can be language-free, the PARCA is designed to be as nonverbal as possible, and yet it is the strongest predictor at age 4;0. This finding is also consistent with previous findings from the TEDS study (e.g. Plomin *et al.*, 2002; Dale *et al.*, 2003), which suggest the importance of general cognitive ability in addition to specific language abilities. Associations with behaviour problems are also significant, although much more modest than those with the language and cognitive variables.

One of the major points of consideration in the previous prospective analysis of the TEDS data was that it relied solely on parental report measures. Despite the well-established general validity of the parental report measures (Dale, 1991; Saudino *et al.*, 1998; Oliver *et al.*, 2002), there may be limitations especially in our sample of families with low language children because of the familial component to language difficulties (Dale *et al.*, 2003). The present study adds weight to the previous conclusions now that we include data not only from parent report as before, but also from tester-administered measures, which anchor our selection of low-language children at 4;6. Indeed, the strength of prediction to standard tester assessments across the parent report measures as well as the difference in predictive power for each measure lends confidence to our use of parent measures. A particularly interesting finding in the current study is that the PGCs for all five measures increase steadily from 2;0 to 3;0 to 4;0. It might be argued that the prediction of language status at 4;6 from data at 4;0 is less interesting than the prediction from 2;0 and 3;0 because the children who were tested at home at 4;6 were selected on the basis of the data just six months earlier (at 4;0), so that we might expect at least the language measures to be predictive. However, this argument does not render prediction from the data at 4;0 uninteresting, for several reasons. First, the substantial change in measurement between the parent measures at 4;0 and the tester-administered measures at 4;6 means that the strength of prediction supports the validity of the parent measures. Moreover, the relationship between low-language status from tester assessment at 4;6 and the nonverbal ability PARCA scores assessed by parents at 4;0 is

almost as strong as the relationship with the verbal measures at 4;0, and the association with behaviour problems remains stable from 2;0 to 3;0 to 4;0, even though neither the PARCA nor behaviour problems were used in the selection of the children for testing in the home at 4;6. Finally, for each measure there is the same robust pattern of an increasing association with age – the pattern does not falter at 4;0, nor does it increase disproportionately, as might be expected if the design of the study is exaggerating associations at 4;0. We are currently collecting data when the twins are 7;0 and 9;0 which will make it possible to test whether this steady increase in PGCs for every measure from 2;0 to 4;0 continues into middle childhood.

The second objective of the current study was to use the twin design to examine the extent to which genetic and environmental influences mediate the phenotypic relationships between low-language status at 4;6 and earlier language and nonverbal factors. Our results indicate that the role of genetic factors in the phenotypic relationships between low-language status at 4;6 and all the earlier language and nonverbal parental measures is moderate, with genetic factors explaining around a quarter of the association in each case. Interestingly, although the phenotypic relationship between low-language at 4;6 and behaviour problems is smaller, the role of genetic influences is similar, if a little stronger proportionately. The current analysis suggests that, when predicting language outcomes in children it may be just as important to examine nonverbal cognitive ability as well as verbal ability in the family history and, possibly behavioural difficulties in family members.

Despite the influence of genetic factors, the greatest effect on the associations described here is that of shared environment, explaining well over half of the longitudinal associations. That is, environmental factors responsible for resemblance between twins growing up in the same family similar in terms of language development are the greatest source of stability. Previous work on the aetiology of low language in TEDS also suggested that early language delay has largely shared environmental origins (Bishop *et al.*, 2003). However, this previous study relied solely on parental report, which can potentially inflate estimates of the shared environment component of behaviours because the same parent completed the measures for both twins and thus may provide artificially similar scores for the twins. Unlike the previous study, the substantial role of shared environment found in the current analysis cannot be explained methodologically in terms of a rater effect, because the outcome measures were administered by testers rather than parents, and each member of a twin pair was assessed individually by a different tester. Finding such strong shared environmental mediation of the prediction of low language at 4;6 should encourage research aimed at identifying specific environmental predictors of language problems, although it is likely that twins share such environmental influences to

628

a greater extent than non-twin siblings who differ in age (Koeppen-Schomerus, Spinath & Plomin, 2003). The influence of shared environmental factors over time is another avenue which TEDS will allow us to explore as we continue to follow the sample into middle childhood and early adolescence.

Although the predictiveness of our measures is not sufficiently strong to be of use for clinical decisions about a particular child, these results have important implications for understanding change and continuity in the early development of language problems. The strength of prediction of low language scores from nonverbal ability is of particular interest, and examining its genetic and environmental mechanisms, as well as that of other predictors, will contribute to understanding the roots of language difficulties in childhood and beyond.

REFERENCES

Bird, J., Bishop, D. V. M. & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech & Hearing Research* **38**, 446–62.

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**(3), 561–75.

Boscolo, B., Ratner, N. B. & Rescorla, L. (2002). Fluency of school-aged children with a history of specific expressive language impairment: an exploratory study. *American Journal of Speech Language Pathology* **11**(1), 41–9.

Botting, N., Faragher, B., Simkin, Z., Knox, E. & Conti-Ramsden, G. (2001). Predicting pathways of Specific Language Impairment: what differentiates good and poor outcome? *Journal of Child Psychology & Psychiatry & Allied Disciplines* **42**(8), 1013–20.

Catts, H. W. (1997). The early identification of language-based reading disabilities. *Language, Speech & Hearing Services in the Schools* **28**(1), 86–9.

Colledge, E., Bishop, D. V. M., Koeppen-Schomerus, G., Price, T. S., Happé, F. G. E., Eley, T. C., Dale, P. S. & Plomin, R. (2002). The structure of language abilities at 4 years: a twin study. *Developmental Psychology* **38**(5), 749–57.

Dale, P. S. (1991). The validity of a parent report measure on vocabulary and syntax at 24 months. *Journal of Speech & Hearing Research* **34**(3), 565–71.

Dale, P. S., Price, T. S., Bishop, D. V. M. & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient delay at three and four years. *Journal of Speech, Language, and Hearing Research* **46**, 544–60.

Dale, P. S., Simonoff, E., Bishop, D. V. M., Eley, T. C., Oliver, B., Price, T. S., Purcell, S., Stevenson, J. & Plomin, R. (1998). Genetic influence on language delay in two-year-old children. *Nature Neuroscience* **1**(4), 324–8.

DeFries, J. C. & Fulker, D. W. (1988). Multiple regression analysis of twin data: etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologiae* **37**(3–4), 205–16.

Dionne, G., Dale, P. S., Boivin, M. & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development* **74**(2), 394–412.

Elliot, C. D., Smith, P. & McCulloch, K. (1996). *Verbal Comprehension Scale from British Ability Scales, 2nd edn.* Windsor: NFER-Nelson.

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), v-173.

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**, 218–27.

Gathercole, S. E., Willis, C., Baddeley, A. D. & Emslie, H. (1994). The children's test of non-word repetition: a test of phonological working memory. *Memory* **2**, 103–27.

Gillis, J. J., Gilger, J. W., Pennington, B. F. & DeFries, J. C. (1992). Attention deficit disorder in reading-disabled twins: evidence for a genetic etiology. *Journal of Abnormal Child Psychology* **20**(3), 303–15.

Goldman, R. & Fristoe, M. (1986). *Goldman Fristoe test of articulation*. Circle Pines, MN: American Guidance Service Inc.

Hammer, C. S., Tomblin, J. B., Zhang, X. & Weiss, A. L. (2001). Relationship between parenting behaviours and specific language impairment in children. *International Journal of Language & Communication Disorders* **36**(2), 185–205.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. *Psychological Bulletin* **111**(1), 127–55.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., Brownlie, E. B., Douglas, L., Taback, N., Lam, I. & Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: speech/language stability and outcomes. *Journal of Speech Language & Hearing Research* **42**(3), 744–60.

Koeppen-Schomerus, G., Spinath, F. M. & Plomin, R. (2003). Twins and non-twin siblings: different estimates of shared environmental influence in early childhood. *Twin Research* **6**(2), 97–105.

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**(3), 430–40.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. San Antonio, TX: The Psychological Corporation.

Oliver, B., Dale, P. S., Saudino, K. J., Pike, A. & Plomin, R. (2002). The validity of a parent-based assessment of cognitive abilities in three-year olds. *Early Child Development and Care* **172**(4), 337–48.

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*. Hove, UK: Psychology Press.

Paul, R. & Fountain, R. (1999). Predicting outcomes of early expressive language delay. *Infant Toddler Intervention* **9**(2), 123–35.

Plomin, R., DeFries, J. C., McClearn, G. E. & McGuffin, P. (2001). *Behavioral genetics, 4th edn.* New York, Worth.

Plomin, R., Price, T. S., Eley, T. C., Dale, P. S. & Stevenson, J. (2002). Associations between behaviour problems and verbal and non-verbal cognitive abilities and disabilities in early childhood. *Journal of Child Psychology and Psychiatry* **43**(5), 619–33.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech Language & Hearing Research* **45**(2), 360–71.

Renfrew, C. E. (1997*a*). *Action Picture Test* (4th edn.). Bicester: Winslow Press Ltd.

Renfrew, C. E. (1997*b*). *Bus Story Test – a test of narrative speech* (4th edn.). Bicester: Winslow Press Ltd.

Sajaniemi, N., Hakamies-Blomqvist, L., Maekelae, J., Avellan, A., Rita, H. & von Wendt, L. (2001). Cognitive development, temperament and behavior at 2 years as indicative of language development at 4 years in pre-term infants. *Child Psychiatry & Human Development* **31**(4), 329–46.

Saudino, K. J., Dale, P. S., Oliver, B., Petrill, S. A., Richardson, V., Rutter, M., Simonoff, E., Stevenson, J. & Plomin, R. (1998). The validity of parent-based assessment of the

cognitive abilities of two-year-olds. *British Journal of Developmental Psychology* **16**(3), 349–63.

Snowling, M. J., Adams, J. W., Bishop, D. V. M. & Stothard, S. E. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *International Journal of Language and Communication Disorders* **36**(2), 173–83.

Spinath, F. M., Ronald, A., Harlaar, N., Price, T. S. & Plomin, R. (2003). Phenotypic 'g' early in life : on the etiology of general cognitive ability in a large population sample of twin children aged 2–4 years. *Intelligence* **31**(2), 195–210.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B. & Kaplan, C. A. (1998). Language-impaired preschoolers : a follow-up into adolescence. *Journal of Speech, Language and Hearing Research* **41**(2), 407–18.

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**(2), 482–96.

Trouton, A., Spinath, F. M. & Plomin, R. (2002). Twins Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behaviour problems in childhood. *Twin Research* **5**, 444–48.

Viding, E., Spinath, F. M., Price, T. S., Bishop, D. V. M., Dale, P. S. & Plomin, R. (2004). Genetic and environmental influence on language impairment in 4 year old same-sex and opposite sex twins. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **45**(2), 315–325.