
Kovas, Yulia, Hayiou-Thomas, Marianna E., Oliver, Bonamy, Dale, Philip S., Bishop, Dorothy V. and Plomin, Robert

Genetic influences in different aspects of language development: The etiology of language skills in 4.5 year-old twins

Originally published in Child Development Copyright Society for Research in Child Development (SRCD). The publisher's version is available at: http://www3.interscience.wiley.com/journal/118733913/abstract Please cite the publisher's version.

You may cite this version as: Kovas, Yulia, Hayiou-Thomas, Marianna E., Oliver, Bonamy, Dale, Philip S., Bishop, Dorothy V. and Plomin, Robert, 2005. Genetic influences in different aspects of language development: The etiology of language skills in 4.5 year-old twins. Child Development, 76 (3), pp. 632-651. ISSN 00093920 [Article]: Goldsmiths Research Online.

Available at: http://eprints.gold.ac.uk/471/

Genetic influences in different aspects of language development: The etiology of language skills in 4.5 year-old twins

Yulia Kovas[1], Marianna E. Hayiou-Thomas[1], Bonamy Oliver[1], Philip S. Dale[2], Dorothy V. M. Bishop[3], and Robert Plomin[1]

[1] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London

[2] Department of Communication Science and Disorders, University of Missouri-Columbia

[3] Department of Experimental Psychology, University of Oxford

Address for Correspondence: Yulia Kovas, SGDP Centre, Institute of Psychiatry, De Crespigny Park, London SE5 8AF.

Acknowledgement: We gratefully acknowledge the ongoing contribution of the parents and children in the Twins' Early Development Study (TEDS). TEDS is supported by a programme grant (G9424799) from the UK Medical Research Council.

Abstract

The genetic and environmental etiologies of diverse aspects of language ability and disability, including articulation, phonology, grammar, vocabulary, and verbal memory, were investigated in a UK sample of 787 pairs of 4½ year-old same-sex and opposite-sex twins. Moderate genetic influence was found for all aspects of language in the normal range. A similar pattern was found at the low end of the distribution with the exception of two receptive measures. Environmental influence was mainly due to nonshared factors, unique to the individual, with little influence from shared environment for most measures. Genetic and environmental influences on language ability and disability are quantitatively and qualitatively similar for males and females. (109)

Introduction

Language acquisition is remarkable both for its universal patterns and the substantial individual differences that occur. These are different but complementary aspects of development. A general pattern of development and the variations around it may stem from different sources. Much of the theorizing about child language development has focused on the species-universal level of analysis, involving heated discussions on the extent of innate and environmental influences, with a consensus gradually emerging that the relevant issue is not nature versus nurture but rather the process of interaction between the two (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996). This issue is relevant not only for language as a whole, but for different components of language, and it is possible that there are linguistic features that are more or less environmentally sensitive (Gleitman, 1988). For example, the use of closed-class morphology in English, which is relatively impoverished but highly irregular, may depend on the caregiver's input more than phonology or syntax which may be more highly constrained by the maturation of a biologically based innate language acquisition device (Cazden, 1966; Gleitman, 1988; Goad & Ingram, 1987). Here again, both theoretical and empirical enquiry into this question have been almost exclusively couched in terms of the commonalities in children's development, that is, the means level of analysis. The same question could - and should - be posed at the individual differences level of analysis: not only does a large amount of variation exist in children's language development, but it is this very variation that presents practical challenges to speech pathologists and educators (Bates, Dale & Thal, 1995).

Within the domain of individual differences in development, another interesting question is whether common developmental disabilities should be conceptualized as the extreme end of a normal distribution, or as qualitatively different categories. According to one view, at least some forms of language impairment constitute a qualitatively distinct condition specifically associated with the abnormal development of syntax (e.g. Gopnik & Crago, 1991; Van der Lely, Rosen, & McClelland, 1998). Recent evidence from molecular genetics demonstrates that at least one qualitatively distinct subtype of language impairment does have a distinct etiology. A single mutation in the FOXP2 gene accounts for language impairment in the famous KE family: all affected members have the same mutation in this gene, while none of the unaffected members do. However, it is unlikely that FOXP2 is involved in more common cases of language impairment, as the mutation has not been found in two large-scale samples of language-impaired children (Meaburn, Dale, Craig, & Plomin, 2002; Newbury, Bonora, Lamb et al., 2002). It is possible that other single gene mutations will be found which underlie other – qualitatively different - cases of language impairment.

An alternative mechanism for language disability is the existence of multiple risk factors of varying effect sizes, some genetic and some environmental, which probabilistically combine to produce language disorder. Different permutations of these multifactorial risks may account for heterogeneity within developmental language disorders. According to this Quantitative Trait Locus (QTL) perspective, different alleles of the same set of genes are responsible for variation across the full range of the distribution (Plomin, DeFries, McClearn, & McGuffin, 2001; Plomin, Owen, & McGuffin, 1994). Common disorders are thus hypothesized to be the quantitative extremes of the normal range of the trait, rather than qualitatively different. Phenotypically, there is support for this position from findings showing that many of the linguistic characteristics of children with language impairment can also be seen to some extent in children with typically developing language (e.g. Bishop, 1997; Leonard, 1998; Bishop, Bright, James, Bishop & van der Lely, 2000).

The discussion on whether common language impairments are qualitatively or quantitatively different from the normal range of language variation has primarily centered on specific language impairment (SLI), which by most definitions requires at least some discrepancy between language and nonverbal ability (Rice, 2003). How SLI should be defined – and how 'specific' language impairment differs from non-specific language impairment – are challenging and important but currently unresolved questions. The jury is still out as to whether a language-nonverbal discrepancy is a stable and inherent part of the disorder (Cole, Dale & Mills, 1990; Cole, Dale & Mills, 1992), and one that is reflected in the etiology (Bishop, North, & Donlan, 1995; Hayiou-Thomas, Oliver & Plomin, in press). Nonetheless, the causes of language difficulties in children are worth considering, regardless of whether the children fulfill clinical diagnoses for SLI. Our focus in the present paper is on these children with low language abilities. The results are relevant for understanding SLI, but only in the context of the resolution of a different debate – namely, whether 'specific' language impairment is a valid and separate category.

Genetic methods such as twin studies that compare identical and fraternal twins provide a powerful tool for understanding individual differences by estimating the proportion of variance that can be attributed to genetic, shared, and nonshared-environment factors (Plomin et al., 2001). Only genetic and environmental factors that vary in the population are assessed by this method. In the case of complex traits that are likely to be influenced by multiple factors, the genetic component of variance refers to the influence of alleles at all gene loci that affect the trait. The similarity between twins on any particular trait can be due to genetic influences that they have in common, twice as much for identical (monozygotic, MZ) twins as for fraternal twins (dizygotic, DZ). On the other hand, the similarity may be due to the 'shared environment', which refers to any environmental influences that contribute to the similarity between co-twins. For example, twins experience similar conditions during gestation, have the same socio-economic status, live in the same family, and usually go to the same school. With respect to language, to the extent that both members of the twin pair are exposed to the same quality and similar quantity of linguistic input, and share the linguistic traditions and conventions practiced in their home, this might reasonably be expected to increase similarity between them. 'Nonshared environment' refers to any aspect of environmental influence that makes co-twins different from each other. The differences due to measurement error are also included in this term. Intuitively, nonshared environmental influences are likely to stem from aspects of environment that are specific to an individual, such as traumas and diseases, idiosyncratic experiences, different peers, and differential treatment by the parents.

We know relatively little about the specific environmental influences that contribute to similarities or differences between siblings. Although the examples we gave above for candidate 'shared environments' seem intuitively to be shared by co-twins, this must be empirically tested. Influences that appear to be objectively similar for two children could plausibly be experienced differently due to interactions with the children's genetic predispositions and unique experiences, and thus have a differential effect on their development (Plomin et al., 2001).

In a review of nearly a hundred genetic studies of normal variation in language skills, Stromswold (2001) concluded that almost all aspects of language ability, from syntax and semantics to phonology and articulation, are influenced by genetic factors to some extent. However, the studies differ greatly in the extent to which shared environmental influence appears to be important. For example, in the case of articulation, modest genetic influence and substantial shared environmental influence were found in two studies of young children (Matheny and Bruggemann, 1973; Mather and Black, 1984). Genetic influence was substantial and shared environmental influence was negligible for speech production and a nonword repetition task (Bishop, Bishop, Bright, James, Delaney, & Tallal, 1999). Genetic and shared environmental influences were both moderate for phonological awareness (Hohnen & Stevenson, 1999), vocabulary and verbal fluency (Thompson, Detterman, & Plomin, 1991), and grammar (Mitler, 1969; Mather & Black, 1984; Dale, Dionne, Eley, & Plomin, 2000). However, not all measures that appear to assess similar abilities have resulted in similar estimates. For example, in contrast to the moderate genetic effects on syntax reported by Dale et al. (2000), the 'wug' elicitation task of inflectional morphology (Berko, 1958) yielded a very modest estimate of genetic influence and substantial influence of shared environment (Mather & Black, 1984). This could be due to the different areas of morphosyntax measured by the two tasks, but could equally be due to differences in the samples or procedures in the two studies. One reason why the estimates of shared environment have differed so greatly across studies might be the use of the same versus different testers to assess members of twin pairs. For example, if the same parent assesses both twins (Dale et al., 2000), the estimate of shared environment can be inflated. Unfortunately, not all studies provided this information, so it is difficult to judge whether this methodological difference produced the inconsistencies in results. Furthermore, tester bias cannot be the sole explanation of different results across studies, since there are cases where the influence of shared environment was moderate even when members of the twin pair were assessed by different testers (e.g. Hohnen & Stevenson, 1999).

Estimates of heritability and environmentality can be used to address the second question raised above, namely, whether the genetic and environmental etiology of low performance differs from the etiology of normal variation (Plomin et al., 2001). In the case of language acquisition, there is some indication that genes may play a more important role in low-language performance than in the normal range (Spinath, Price, Dale, & Plomin, 2004). In the Twins Early Development Study (TEDS), parent-reported vocabulary and grammatical ability at 2, 3, and 4 years of age yielded slightly higher heritability estimates for the bottom 5% and 10% of the distribution, than for the whole sample. Although the pattern appears consistent across these ages, it is important to emphasize that these differences in heritability estimates were very small.

As with the results for the etiology of language development in the normal range, there is considerable variation in results for different aspects of language at the low end of ability, with heritability estimates ranging from 25% to 100%, depending on the measures, samples, and methods used (Stromswold, 2001).

It is therefore difficult to reach any reliable conclusions about the etiology of the early development of different domains of language, either for the normal range of ability or the low end of performance. The use of different methods of ascertainment and measures, and children of different and widely varying ages makes comparisons between studies difficult. As discussed earlier, another possible source of variation in results is whether the same person assesses both members of a twin pair, which may lead to overestimation of twin similarity (particularly for identical twins), and thus distort genetic and environmental parameter estimates. For most previous studies, however, the main limitation is small sample size, which results in large standard errors for genetic and environmental parameter estimates. For example, power analysis reveals that sample sizes of 180 pairs of each type of twin are required to attain 80% power simply to detect a typical heritability of .40 based on correlations of .60 for MZ twins and .40 for DZ twins (Neale, 1997). Point estimates of heritability based on twin samples of this size have large 95% confidence intervals and much larger sample sizes are required to compare estimates across measures. Estimates of shared environment require even greater sample sizes (Neale, 1997).

These considerations highlight the need for large twin studies of children of the same age with multiple measures of language and different testers for each twin. A recently initiated study that uses multiple measures and different testers – but does not yet have a large sample – focuses on pre-reading and related cognitive skills (Byrne et al., 2002). In a sample of 109 MZ and 106 DZ pairs, preliminary findings showed moderate genetic influences and negligible shared environmental influences on phonological awareness, and the opposite pattern for vocabulary, grammar and morphology in preschool same-sex twins (mean age 58.9 months). Again there was variability in genetic and environmental influences within what could be considered to be a single aspect of language. For example, productive grammar as assessed with a test of grammatical closure showed negligible genetic and moderate shared environmental factors, whereas it showed moderate genetic influences when assessed with a test of productive morphology. The sample size was too small to compare these estimates, although the authors report that further data collection is in progress; additionally, the focus is on individual differences in the normal range and there are no results reported for low performance.

The results reported in the present paper are from a sub-sample of the Twins' Early Development Study, the largest twin study to date to investigate diverse aspects of language, including articulation, phonology, grammar, vocabulary, and verbal memory, in a group of children of the same age (4½ years old), with each co-twin tested by a different tester. We have previously reported results of genetic analyses of a general language composite based on these measures that yielded estimates of moderate heritability and shared environmental influence for the normal range of variation in a control sample of 310 twin pairs (Colledge et al., 2002) and for low performance in a sample of 393 twin pairs in which at least one twin was in the lowest 15% of the distribution (Viding, Spinath, Price, Bishop, Dale, & Plomin, 2004). However, the general language factor, as indexed by a first unrotated principal component, accounted for less than half of the variance of the language measures (Hayiou-Thomas, Kovas, Harlaar, Bishop, Dale, & Plomin, submitted), which warrants an examination of the individual language measures.

The present paper focuses on univariate analyses of nine diverse measures of language, in order to investigate the relative contributions of genetic and environmental factors to individual differences, for both the normal range of variation, and the low end of ability.

In addition to comparing the etiology of normal variation and the low extremes in our sample, we also asked whether there are sex differences in the etiology of ability and disability in diverse language domains. Phenotypically, there is a well-established average male disadvantage for many areas of language. For example, girls' articulation was advanced relative to boys' in a study of young twins (Matheny & Bruggemann, 1973). Girls also tend to produce language earlier than boys (Karmiloff & Karmiloff-Smith, 2001) and to acquire vocabulary somewhat faster than boys between the ages of about 14 to 20 months, after which time boys begin to catch up (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). At the low end of performance, too, boys are more likely to have vocabulary delay at 2 years of age, but are nearly as likely as girls to catch up to normal vocabulary levels by 4 (Dale, Price, Bishop & Plomin, 2003). In addition, the average male: female ratio in SLI has been estimated as 2.8:1 (Robinson, 1987). However, these findings refer to mean differences between the sexes; our focus is on the etiology of individual differences within the sexes – that is, whether the same genetic and environmental factors influence differences in language ability for boys and girls. These etiological factors may be completely different from those that affect mean sex differences (for further discussion of this distinction, see Viding et al., 2004).

Previous analysis of the general language composite derived from our measures yielded little evidence for differences between boys and girls for the estimates of genetic and environmental influence on language impairment (Viding et al., 2004). However, previous studies have not investigated the possibility that sex differences might emerge for some linguistic skills but not others. In the current study we include opposite-sex twins in order to explore sex differences in genetic and environmental influences for each measure, which will test the possibility that such differences exist for some aspects of language but not others.

There are three possibilities with respect to the causes of individual differences in boys and girls, regardless of mean differences between the sexes (Neale & Maes, 2003). The first possibility is that different genetic and environmental factors are responsible for individual differences in language for boys and girls – these are called qualitative differences. An example of qualitative sex differences is sex-specific genetic influences that contribute to individual differences in one sex but not the other. Such sex-specific effects are not limited to genes on the X chromosome but can also involve genes on the autosomal chromosomes that affect boys and girls differently, for example, because the genes interact with sex hormones. The second possibility, not mutually exclusive with the first, is that the same etiological influences affect individual differences in boys and girls, but that they do so to a different extent. For example, the same genes may play a greater role in individual differences for boys than girls – these are known as quantitative differences. The third possibility is that even if there are mean differences, there are no differences in the etiology of individual differences for boys and girls, the same genes and environments operate to the same extent in both sexes. That is, boys as a group may exhibit a disadvantage in language, but the factors that make one boy different from another are the same as those that make one girl different from another girl.

In summary, the main objectives of this study are: (1) to use the twin method to estimate the extent to which genes and environment influence different aspects of language; (2) to investigate genetic and environmental influences on the low-language extremes of language development as well as on the normal range of variation; and (3) to test whether the same genetic and environmental factors and the same magnitude of genetic and environmental effects influence individual differences in boys and girls, both for the whole sample and the low-language extremes.

Method

Participants and Procedure

All participants were part of the Twins' Early Development Study (TEDS). This longitudinal study involves a representative sample of all twins born in England and Wales in 1994, 1995, and 1996 whose language, cognitive and behavior development has been assessed by parental questionnaires at 2, 3 and 4 years of age (Trouton, Spinath, & Plomin, 2002).

A sub-sample of TEDS was selected for an in-depth, in-home assessment at 4½ years of age (herein referred to as the 'inhome sample'). The mean age of the inhome sample was 4;6 (years;months; *SD* = 2 months, range from 4;0 to 4;11). All participants in the inhome sample were selected to be ethnically white in order to control for the effect of ethnic stratification in future molecular genetic studies; however, over 94% of the population of England and Wales is white. The following exclusion criteria were also used: specific medical syndromes such as Down's syndrome and other chromosomal anomalies, cystic fibrosis, cerebral palsy, hearing loss, autism spectrum disorder, organic brain damage, extreme outliers for birth weight, gestational age, maternal alcohol consumption during pregnancy, and special care after birth. Only participants for whom English was the first language spoken at home were selected. Both same- and opposite-sex twin pairs were included in order to investigate potential sex differences. As an indication of representativeness of this sub-sample, maternal education levels were assessed: 32% of mothers with A-levels (university entrance qualifications) - were comparable both to the overall TEDS sample (39%), and UK Office of National Statistics census data (32%).

The purpose of the inhome sample was to investigate impairments in language and cognitive development. In order to ensure that the sample would contain both a sufficient number of controls (with typical development in these areas) and a substantial number of children with impairment, we used the parental reports obtained at 4 years for the whole TEDS sample as screening tools. 516 twin pairs were selected for the inhome study, in which at least one twin's scores on the screening measures suggested the twin was at risk for low language or non-verbal ability.

Children were considered at risk for low language ability if they met any of the following three criteria: (1) lowest 5% of the vocabulary distribution as assessed by an upward extension of the MacArthur Communicative Development Inventory, a parent report measure of vocabulary (MCDI; Fenson et al., 1994; Dionne, Dale, Boivin, & Plomin, 2003); (2) parents selected one or more of the following descriptions of their child’s language: ‘not yet talking’, ‘talking in one-word utterances’, ‘talking in 2 or 3 word phrases’, in response to the question ‘which of the following best describes the way your child talks?’ (from the TEDS parent report booklet); (3) parents indicated concern about their child’s speech and language development, by selecting the following item: ‘his/her language is developing slowly’ (from the TEDS parent report booklet), and the child scored 4 or less on a parent-administered receptive picture vocabulary test (8 items based on the Peabody Picture Vocabulary Test - Revized (PPVT-R), Dunn & Dunn, 1981). The children were considered at risk of poor nonverbal development if they scored in the lowest 5% of the nonverbal ability distribution as assessed by the Parent Report of Children’s Abilities (PARCA; Saudino et al., 1998; Oliver, Dale, Saudino, Pike, & Plomin, 2002). All the twin pairs in the TEDS sample, in which at least one of the children met any of the above criteria, and who had data available at the start of the inhome study were selected for the in-home testing. In addition, a control sample of 310 twin pairs was randomly selected from all TEDS twins who did not meet any of the above criteria.

49 pairs were excluded from analyses because one or both twins in a pair had missing data or it was discovered at the time of the home visit that they met the exclusion criteria. All analyses described below were performed on data from the remaining sample of 1574 children who were from 281 MZ (monozygotic) pairs, 275 DZ (dizygotic) same sex pairs, and 231 DZ opposite-sex pairs.

Testing procedures

Informed consent was obtained in writing from all of the families who agreed to take part in the study. The sessions took approximately 1hr 30 min during which the children were assessed on a battery of verbal and non-verbal tests (the full battery is described in Colledge et al., 2002). A different tester assessed each member of a twin pair.

Measures

The verbal battery was chosen with the following criteria in mind: tests should be suitable for 4½ year olds, should show variation across the range of ability at this age, should have established psychometric properties, and should differ from each other with respect to the main putative source of variation. Without subscribing to a particular theoretical position on the structure of the language domain, we aimed to choose tests that would between them cover a wide range of the linguistic abilities of 4½ -year-old children, including phonology, semantics, and grammar. In addition, the measures differ according to whether they primarily assess expressive or receptive ability, and the demands they make on memory (either working or semantic memory) and metalinguistic awareness. Some overlap in what these tests measure is inevitable, as they each make demands on overlapping cognitive and performance factors (attention, motivation, memory) and it is never possible to get a completely 'pure' measure of one language component. Consider, for example, Berko's famous "Wug" test that was devised to measure children's knowledge of morphological rules (Berko, 1958). Children are asked to name pictures of strange animals and people performing unusual actions, using nonsense words. A classic example is the experimenter pointing to a picture saying: " This is a wug. This is another one. Now there are two

__." The children are then expected to fill the gap with the appropriate plural ending. Although good performance on this test provides evidence of knowledge of formation rules for verb morphology, poor performance could reflect non-morphological factors, such as difficulty in remembering the nonword that needs to be inflected (weak phonological short-term memory) or expressive phonological impairment. Although one can never completely control for the multiple verbal and nonverbal influences on performance of a language test, it is nevertheless possible to choose tests that stress one component of language more than another. This was the aim in the current study, and evidence that it was achieved can be seen from the fact that, for the nine measures we used, the phenotypic inter-correlations are moderate (see the Results section), accounting for approximately 16% of the variance between them, suggesting that the tests do measure diverse abilities.

The test battery consisted of the following:

Expressive Semantics

Three tests were used to index the child's semantic skills, while minimizing the role of syntax and phonology:

MSCA Word Knowledge (McCarthy Scales of Children's Abilities; McCarthy, 1972) is an expressive test of semantic knowledge. The Picture Vocabulary subtest requires the child to point to the picture corresponding to the word said by the examiner. The Oral Vocabulary subtest requires the child to give an oral definition of ten words: 2 points are awarded for including utility, salient characteristics or a good synonym; 1 point for describing a word incompletely or vaguely; 0 points when no knowledge of the word is indicated. For example, 'towel' would receive 2 points for a response which included 'to dry', but only 1 point for 'use in bathroom'. The maximum raw score for this subtest is 20. Only the oral vocabulary subtest was used because of a ceiling effect in the picture vocabulary subtest. Syntactic complexity and phonological accuracy of responses are not taken into account when scoring the Word Knowledge subtest.

MSCA Verbal Fluency (McCarthy, 1972) is a test of word generation and semantic knowledge. The child is asked to name as many examples of items as possible in a given category within 20 seconds. There are four categories, namely 'things to eat', 'animals', things to wear' and 'things to ride'. 1 point is awarded for each acceptable response, with a maximum score of 9 for each category imposed; the maximum possible raw score is therefore 36. This test, unlike MCSA Word Knowledge, stresses speed and flexibility in retrieving lexical items from memory.

The Renfrew Bus Story Test (Renfrew, 1997a) assesses ability to give a coherent description of a continuous series of events. The experimenter reads a story from a book with pictures, and the child is then asked to retell the story while looking at the pictures. We used the information score suggested by Renfrew et al. (1997a), which reflects the story content included in the re-telling. For example, in the story, a policeman blows his whistle and says: "Stop, bus!" to a runaway anthropomorphic bus. The child would receive one point for mentioning the policeman, an additional one for mentioning the whistle, and yet another for mentioning that the policeman said 'Stop'. The information score disregards the grammatical complexity of the child's narrative, and is concerned only with the content. Although it is possible to obtain an index of syntactic complexity from the Bus Story, we did not include this in the current analysis, as it was felt that results might be biased in favor of finding commonalities between semantics and syntax if the same narrative was used to index both domains. Although we have categorized the Bus Story as an expressive semantic test, task demands are considerably more complex than for the Word Knowledge test, insofar as the child has to both understand and re-tell the story. Thus this test assesses both expressive and receptive abilities, and makes demands on both semantic and working memory. The Bus Story information score has been shown to be a sensitive index of SLI, and a good predictor of outcome in language-impaired children (Bishop & Edmundson, 1987).

Expressive syntax

The Renfrew Action Picture Test (Renfrew, 1997b), grammar score. This is an elicitation task designed to solicit utterances containing different types of grammatical construction. It has been shown to be sensitive to variations in grammatical development in 4-year-olds, and differentiates children with SLI from typically-developing children (Bishop & Edmundson, 1987). In this test, the child is presented with 10 picture cards, depicting scenes of increasing complexity, and asked to describe each one; the examiner can use a limited number of indirect prompts to encourage a full description. As with the Bus Story, separate Information and Grammar scores can be derived from the child's response: the Information score is based on the content of the child's response (similar to the Bus Story); the Grammar score reflects the use of inflectional morphology and function words.

For example, the first card shows a girl cuddling her teddy bear. The maximum Information points a child could get for this card is 2, for mentioning 'cuddle' and 'teddy'. The maximum Grammar score is 1 point, for using the progressive – ing on 'cuddling'. The Information and Grammar scores were very highly correlated in our sample (.77). However, because we did not want to bias our results in favor of finding associations between syntax and semantics, we used only the grammar measure from this task. The constructions elicited in the Action Picture test are as follows: present participle *-ing*; future tense; regular past tense *-ed*; irregular past tense; regular plural noun -*s*; irregular plural nouns; possessive -*s*, nominative pronouns *she*, *he*, *it*; relative pronouns *that*, *which*, *who*; auxiliary *is*, *has*, *was*; passive *got*, *been*; coordinating conjunction *and*; subordinating conjunction *because*; determiners *a*, *the*. Phonological accuracy of utterances is not taken into account when scoring the Action Pictures test (the grammar score), although it must be acknowledged that a child with an expressive phonological impairment could be handicapped by problems in producing inflected forms.

Receptive syntax

The BAS Verbal Comprehension subtest (BAS; Elliot, Smith, & McCulloch, 1996) is a test of receptive language. The child is presented with a set of toys, and asked to arrange them according to the examiner's instruction. For example, 'Put the house on each side of the car'; the child receives one point for a correct response (no verbal response required), and zero points for an incorrect response. We used a subscale consisting of the last 11 items of the BAS I Verbal Comprehension subscale, which required comprehension of grammatical morphology and syntax (a maximum raw score of 11 is therefore possible). The scores from the first section of this subtest, which consisted of items requiring only lexical comprehension, showed a clear ceiling effect, and were excluded from further analyses.

Verbal memory

a) Memory for meaningful materials

MSCA Verbal Memory Words and Sentences (McCarthy, 1972). The Words and Sentences subtest requires the child to repeat words presented in three or four word sequences or sentences. The child is awarded 1 point for each successfully repeated key word, and a maximum of 30 points is possible on this subtest. Note that performance with the sentence stimuli in this subtest will be influenced by receptive and expressive syntactic ability, in addition to the memory requirement. MCSA also includes a Story subtest that requires the repetition of a short story; however, this subtest showed a floor effect and was excluded from further analyses.

b) Phonological short-term memory

The Children's test of Nonword Repetition (Gathercole & Baddeley, 1996) is a test of phonological working memory in which the child is asked to repeat nonsense words (e.g. skiticult, rubid). This task also makes substantial demands on receptive phonological ability as well as expressive phonology. A 20-item version of the test was used, with ten items at each of the 2 and 3 syllable lengths based on the pilot work suggesting that longer words produced floor effect at this age. 1 point is awarded for a correct response, and 0 for an incorrect response, with a maximum possible raw score of 20.

Because children's articulation at this age is often immature, it was not feasible to attempt to adjust scoring to allow for mis-repetitions that were consistent with the child's expressive phonological repertoire. Thus results from this measure will be sensitive to articulatory accuracy as well as phonological short-term memory.

Receptive phonology

We considered using a test of speech sound discrimination to assess basic receptive phonology skills, but decided against this on the basis of pilot work that showed that 4-year-olds lacked the necessary attentional skills to complete the kind of multiple-choice test that is typically used in this area.

A test of phonological awareness was included in our battery because there is ample evidence that this language skill plays a unique role in literacy development (Bradley & Bryant, 1983), independent of vocabulary and verbal memory. At the time this study was conceived, there were no good standardized tests of phonological awareness suitable for 4-year-olds, and we therefore devised our own materials.

The Phonological Awareness task (based on Bird, Bishop, & Freeman, 1995) is a purely receptive task that does not involve any expressive language from the child, but requires the child to judge whether phonemes presented in different word contexts are the same. The test has substantial metalinguistic and memory demands, but every effort is made to reduce the memory load. The child is introduced to puppets and told that the puppets like things that sound like their names. Pictured choice items are named by the experimenter and left in front of the child. The child is required to choose one item from the set of four (two in the practice trials) on the basis of rhyme. After 4 practice trials with feedback a further eight items are administered. For example: 'Which of these things would Lynn like?' 'Chair?' 'Bin?'. 1 point is awarded for each correct response, with a maximum possible raw score of 8 points.

Expressive phonology

The Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986) Sounds-in-Words Subtest is designed to assess production of specific speech sounds. The child is asked to name pictures depicting objects and actions that are familiar to young children. The examiner listens for specific target phonemes – most of which are tested for in initial, medial and final positions - and codes these as correct (1 point) or incorrect (0 points). 23 simple consonants and 12 blends are tested, with a maximum possible raw score of 74.

With the exception of the phonological awareness task, which is based on materials used by Bird et al. (1995), all tasks used in this study are published measures, well established and widely used. Full information on standardization, reliability and validity of each test can be found in the published manuals: The McCarthy Scales of

Children's Abilities (McCarthy, 1972); The Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986); Action Picture Test and the Bus Story Test from the Renfrew Language Scales (Renfrew, 1997a, 1997b); the Verbal Comprehension subtest from the British Ability Scales (Elliot, Smith, & McCulloch, 1997); The Children's test of Nonword Repetition (Gathercole & Baddeley, 1996).

Analyses

The twin method is based on estimating the relative genetic and shared and nonshared environmental components of variance by comparing intraclass correlations for monozygotic twins (MZ) who are genetically identical and dizygotic (DZ) twins whose genetic relatedness is on average .50. The relatedness for shared (common) environmental influences is assumed to be 1.0 for both MZ and DZ twin pairs who grow up in the same family because they experience similar prenatal and postnatal environments. Thus, if linguistic abilities are more similar within MZ twin pairs than those within DZ pairs, genetic influences are suggested.

Heritability can be estimated as twice the difference in MZ and DZ twin correlations. Shared environment can be estimated by subtracting the heritability estimate from the MZ correlation, which reflects similarity between twins beyond that accounted for by genetic similarity. The remaining proportion of variance, not accounted for by the genetic relation of a twin pair or by the fact that the twins are growing up in the same family, is attributed to nonshared environment but also includes measurement error. Model-fitting analysis is a more comprehensive way of estimating variance components based on the same principles. Further details of the twin method and its assumptions are described elsewhere (Plomin et al., 2001).

Model fitting

Individual Differences Analyses

The analyses of individual differences reported in this study employed variations on the full ACE model that apportions the phenotypic variance into genetic (A), shared environmental (C), and nonshared environmental (E) components, assuming no effects of non-additive genetics or non-random mating.

Sex differences in the genetic and environmental parameter estimates were assessed by testing a full sex-limited model and a series of three nested models, corresponding to the three alternatives outlined earlier: qualitative differences, quantitative differences, and no differences. Each of these possibilities is associated with a set of parameters in the models. Qualitative differences are reflected in the genetic correlation (r_g) between DZ opposite-sex twins. In DZ same-sex pairs, the assumption is that on average the twins share 50% of their varying DNA, and the coefficient of genetic relatedness (the genetic correlation between the two children) is therefore 0.5. If there are qualitative differences in etiology between boys and girls (different genetic and environmental factors), the genetic correlation in DZ opposite-sex twins will be less than 0.5. If there are quantitative differences (the same factors, but exerting different magnitudes of effect) rather than qualitative differences, the genetic correlation for DZ opposite-sex pairs will be 0.5, but the parameter estimates for the A, C, and E components will be significantly different for male-male pairs and female-female pairs. If there are no differences between boys and girls, the DZ opposite-sex (DZos) pairs will have a genetic correlation of 0.5 and the A, C, and E estimates for male-male and female-female pairs will be the same, although the phenotypic variance might nonetheless differ for the two sexes because mean differences are often associated with variance differences (i.e., higher means have higher variances). The full model allows all parameters to vary: r_g in the DZ opposite-sex pairs, A, C, and E estimates, and variance estimates. The first nested model is a common-effects sex-limitation model, which fixes r_g to 0.5 in the DZos, but allows different A, C, E and variance estimates. The second nested model is a scalar effects sex-limitation model, which constrains the r_g in the DZos, as well as the A, C, and E parameters, but allows differences in phenotypic variance between males and females by modeling the variance in one sex to be a scalar multiple of the variance in the other sex. The third and final nested model tests the null hypothesis, and constrains all the parameters to be equal for males and females.

The ACE parameters and their confidence intervals were estimated by fitting the full and the nested models to variance/covariance matrices using the model-fitting program Mx (Neale, 1997). The overall fit of each model was evaluated using three indices. The χ^2 statistic, where degrees of freedom equal the number of observed correlations minus the number of estimated parameters, indicates the fit of the full model and also tests the fit of nested models, with a lower value indicating better fit (with degrees of freedom equal to the difference in degrees of freedom between the full and nested models). However, the χ^2 statistic is inflated with large sample sizes. The other two indices – Akaike's information criterion (AIC = χ^2-2df; Akaike, 1987) and the root mean square error of approximation (RMSEA) – give more interpretable estimates of fit for larger samples, with lower values representing better fitting models.

Extremes Analyses

For each of the measures, we defined probands as those children who scored more than one standard deviation below the sample's mean (15.87% of the whole sample). This cut-off does not correspond to any standard diagnosis of language disability, but rather identifies statistically low performance on that measure. The correspondence between the cut-offs applied in the current study (based on the distributional properties of our sample) and the published test norms is reported in the Results section.

Probandwise concordances (the ratio of the number of probands in concordant pairs to the total number of probands) were calculated separately for each measure and each of the 5 sex-by-zygosity groups. Probandwise concordances represent the risk that a co-twin of a proband is affected (Plomin et al., 2001). Greater MZ than DZ concordances suggest genetic influence but unlike twin correlations, twin concordances cannot be used to estimate genetic and environmental parameters because they do not in themselves include information about the population incidence.

The liability-threshold model, which is a natural extension of quantitative genetic models for quantitative traits, is widely used in genetics to analyze concordance data (Sham, 1998). The model assumes an underlying continuous liability that has a normal distribution with a mean of 0 and a variance of 1 in the general population. If the liability to a disorder is quantitative rather than categorical, the disorder is assumed to be present in all individuals whose liability is above a certain threshold value and to be absent in all other individuals. The value of the threshold can be estimated from the population frequency of the disorder. The liability is not measured directly, but is estimated from the observed categorical data. For the purposes of this study, the data from the entire twin sample were organized into 2 x 2 contingency tables, where cells represent pairs in which both twins are unaffected, both twins are probands, and two discordant cells where twin one or twin two are probands. These data can be used to quantify genetic and environmental sources of variation in liability in the population. In this study a structural equation model was fit to the contingency tables by maximum likelihood, using the Mx program to estimate ACE parameters (Neale, 1997).

A full sex-limited liability-threshold model and a series of nested models were tested; the full model, common effects model, and scalar effects models used were the same as those used in the individual differences modeling. In addition, the null model in this case equated thresholds for males and females; the threshold corresponds to the proportion of affected individuals for the two sexes. Whether the same threshold could be fit for males and for females tested whether rates of disorder differed for males and females.

Differences in the results for the extremes analyses and the individual differences analyses were treated as statistically significant when the point parameter estimates for the individual differences analyses fell outside the 95% confidence intervals of the estimate in the extreme analysis.

Results

In order to investigate the extent to which the intentional over-representation of low-language children in our sample affected the distribution of scores for each individual measure, descriptive statistics from the total sample of 1574 children were compared with those from the standardization samples reported in the test manuals (see Table 1). Although means were somewhat lower and standard deviations (SDs) somewhat greater for most of the measures, the means and SDs were comparable with the reported norms. These results suggest that although the distribution of most measures was shifted towards the low end, this shift was within 1 standard deviation of the reported norms for 4 ½ year olds.

The enrichment of the sample for cases who were at risk for low-language was advantageous in terms of increasing the statistical power needed to study language disability in the context of language ability. Table 1 shows the raw scores from each measure corresponding to the –1SD cut-off for proband selection in the present sample. It can be seen that the cut-off for each measure in our sample corresponds to between –1 and –2.2 SD in the published norms for the tests.

Insert Table 1 here

To allow for comparisons among the different measures, standardizations were carried out separately for all 9 measures using the means and standard deviations of the entire sample (after exclusions described in the method section), so that each test had zero mean and unit variance for the total sample of 1574. With the exception of the Goldman-Fristoe Test of Articulation (which was subjected to log transformation) all measures reported in this study have unimodal (and in most cases near-normal) score distributions (see Figure 1). Although all children were tested at around 4½ years of age the results could be affected even by small differences in age at this important stage of language development. Therefore the linear effects of age were regressed from these standardized data.

Insert Figure 1

Descriptive statistics for the age-regressed scores for the 9 measures are summarized in Table 2. It can be seen that MZ twins consistently have lower means than DZ twins, which may be due to greater perinatal complications of MZ twins (Lenneberg, 1967). Analysis of variance revealed a significant main effect of zygosity for all nine measures favoring DZ twins. However, the effect size (η^2) of zygosity is small, accounting for between .5% and 1.6% of the variance. Similarly, girls generally performed significantly better than boys for seven of the nine measures. The largest mean sex difference was found for nonword repetition, which accounted for 3.2% of the variance. However, the other significant sex effects were negligible, accounting for .2% to .7% of the variance. Sex by zygosity interaction was not significant for any of the measures.

Insert Table 2 here

Pairs in which one or both twins scored 3 or more SDs below or above the mean were excluded from further analysis of individual differences (but not extremes) for each individual measure: 1 pair for The Renfrew Bus Story Information Test, 14 pairs for The BAS Verbal Comprehension subtest, 14 pairs for The MSCA Word Knowledge subtest, and 3 pairs for the MSCA Verbal Fluency subtest. This exclusion was necessary because extreme scores can lead to distortion of results in correlational analyses. Pairs in which one or both twins had missing data were also excluded from further analysis for each individual measure. Final numbers of participants included in the genetic analyses for each measure are presented in Tables 2 and 4.

Phenotypic relationships between the measures have been investigated and described in another TEDS paper (Hayiou-Thomas et al., submitted), which focuses on multivariate genetic analysis. Correlations were varied, but mostly moderate (.29-.68).

Genetic Analyses

Individual Differences Analyses

Genetic and environmental influences on individual differences were first assessed by comparing MZ and DZ twin correlations. It can be seen from Table 3 that MZ twins performed more similarly than DZ twins on all 9 measures, with average MZ and DZ correlations of .55 and .35, respectively. Doubling the difference between the MZ and DZ twin correlations suggests an average heritability estimate of

.40. Shared environmental influence can be estimated as .15, the extent to which the heritability estimate of .40 does not account for the MZ correlation of .55. The rest of the variance, .45, can be attributed to nonshared environment and measurement error. Generally similar results are found for boys and girls separately, especially when taking into account the smaller sample sizes, suggesting that ACE parameter estimates are similar for boys and girls. Another interesting finding is that DZ opposite-sex twins are not less correlated than the same-sex DZ twins – the average correlation for opposite-sex twins is .32 and the average correlation for same-sex twins is .34. This is consistent with the hypothesis of similar etiologies of individual differences in these measures of language development for boys and girls.

Insert Table 3 here

The full sex-limitation model and the reduced models were tested and compared as described in the analyses section. The results of the model fitting are summarized in Table 4. All model comparisons favored the null model. In other words, fixing ACE parameters and variances to be the same for boys and girls, and fixing the genetic correlation for DZ opposite-twin pairs to .5 resulted in non-significant changes in the fit of the model. This suggests that the quantity and quality of genetic and environmental effects are the same in boys and girls. For this reason only the parameter estimates from the null model are presented in Table 4.

Insert Table 4 here

As suggested by the twin correlations in Table 3, the overall pattern of results suggests that all measured aspects of language are moderately heritable (.29-.53), whereas shared environment has a much smaller influence on these abilities (.06-.26). Non-shared environmental influence is moderate for all measures (.34-.56), which is partly due to the inclusion of measurement error in this component. As can be seen from Table 4, the confidence intervals of the genetic and shared environmental estimates are wide and overlapping across the nine language measures. For example, the confidence intervals overlap between the most heritable measure (Bus Story Information) and the least heritable measure (Action Picture Grammar), which indicates that these heritability estimates do not differ significantly. Therefore only very broad conclusions can be drawn for these nine language measures: consistent and moderate genetic and nonshared environmental influences, and modest shared environmental influence.

Extremes Analyses

Genetic and environmental influences on performance at the low end of the distribution were first examined by comparing MZ and DZ probandwise concordances. Table 5 shows that concordances for MZ twins are generally higher than for DZ twins, with the exception of Phonological Awareness and BAS Comprehension. The average MZ and DZ concordances are .51 and .33, respectively, suggesting that genetic influence at the low extreme is of a similar magnitude to the results found for individual differences throughout the distribution. Again, results were similar for boys and girls.

Insert Table 5 here

The full sex-limitation liability-threshold model and the reduced models were tested and compared as described in the analysis section. All model comparisons favored the scalar effects sex-limitation model that constrains genetic and environmental parameters to be equal between the sexes. In other words, as was the case for the individual differences analyses, fixing ACE parameters to be the same for boys and girls, and fixing the genetic correlation for DZ opposite-sex twin pairs to .5 resulted in non-significant changes in the fit suggesting that the quantity and quality of genetic and environmental effects are the same in boys and girls. However, the thresholds could not be equated for boys and girls as this resulted in a significantly worsened model fit. These results suggest that although boys are more likely to have a language disability, these mean differences must be explained by factors other than those that drive individual differences (see Discussion).

Table 6 summarizes the results of the best-fitting scalar models for the low extremes of the nine measures. As expected from the twin concordances in Table 5, Phonological Awareness and BAS Comprehension show zero heritability estimates. Bus Story Information and Verbal Fluency show substantial heritability, zero shared environmental influence, and moderate non-shared environmental influence. The other five measures show modest to moderate heritability, modest shared environmental influence and modest to moderate non-shared environmental influence. However, the wide confidence intervals for these measures do not allow for direct comparisons between them and indicate that heritability might not be significantly less for Phonological Awareness and BAS Comprehension than for the other 7 measures.

Insert Table 6 here

As described in the analysis section, ACE estimates were compared for individual differences and the extremes (see Tables 3 and 5). Despite the apparent differences in the estimates for the two analyses, heritability estimates from the individual differences analyses never fell outside the confidence intervals of the estimates from the analyses of extremes. Indeed, the point estimates from the two types of analyses were strikingly similar for most measures, with the exception of the two receptive measures. Estimates for shared environment fell just outside the confidence intervals for only one measure (BAS Comprehension). Overall, it can be concluded that the ACE estimates for the individual differences and the extremes are similar, at least within the limits of power provided by this largest-yet twin study. In other words, liability to different language disabilities and individual differences in language abilities are influenced by moderate genetic influence and modest to moderate shared environmental influence. Figures 2a and 2b summarize the main findings of the study.

Insert Figures 2a and 2b here

It should be noted that, although the present study is by far the largest twin study of low language performance across several measures within the same sample tested at the same age, the smaller sample size and categorical nature of the data of the liability-threshold analysis led to the reduction in power to detect significant effects (Neale, Eaves, & Kendler, 1994). Thus, no direct comparisons across the measures or the two types of analyses are possible. However, the overall pattern of results suggests moderate genetic influence and modest to moderate shared environmental influence for most measures with the possible exception of the two receptive measures (Phonological Awareness and BAS Comprehension) that may be less influenced by genetic factors, particularly at the low extreme of the continuum. It should also be noted that although the broad pattern of A, C, and E estimates is similar for the low extremes and the full range of performance, the reduced power for the extremes analyses resulted in more variable estimates.

Discussion

The main objectives of this study were: (1) to use the twin method to estimate the extent to which genes and environment influence different aspects of language; (2) to investigate genetic and environmental influences on the low-language extremes of language development as well as on the normal range of variation; and (3) to explore sex differences in genetic and environmental influences on diverse aspects of language using opposite-sex twins.

Overall, the results of this study suggest that such diverse aspects of language as expressive and receptive grammar, phonology, articulation, lexical knowledge and verbal memory show moderate heritability and moderate influence of nonshared environment. Shared environmental influence is modest for most measures. The results are similar when only the low end of the continuum is studied, with the possible exception of the two receptive measures (BAS Comprehension and Phonological Awareness). This similarity is consistent with the hypothesis that the same genetic and environmental influences are involved in shaping individual differences and differences in liability to a disorder.

One important caution in interpreting our findings concerns possible heterogeneity in the etiology of low language performance. Our method cannot differentiate between cases whose language problems stem from different etiologies, if these causal factors have the same magnitude of effect (i.e. identical heritabilities and environmentalities could emerge from different sets of genes and environments). It is possible therefore, that our sample includes cases who represent the low end of the normal continuum, as well as cases with single-gene disorders (although not FOXP2, since genotyping in our sample did not find a single case of the FOXP2 mutation, Meaburn et al., 2002). Although this is logically possible, monogenic disorders are typically extreme and are associated with very high heritabilities; if such cases were present in our sample we would probably have found higher heritabilities in our extremes analyses.

The results of the present study can be compared with previous findings from the TEDS sample at 2 and 3 years of age, when the twins' expressive vocabulary and grammar were assessed by their parents. For example, consistent with our results, moderate influences of genetic factors on expressive grammatical ability were found in 2-year-old twins (Dale et al., 2000; Dionne et al., 2003), and in 3-year-old twins (Dionne et al., 2003). For expressive vocabulary, smaller genetic influences were found at ages 2 and 3 (Dionne et al., 2003). Overall, the influences of shared environmental factors for both vocabulary and expressive grammar at 2 and 3 were larger than found in the present study. These differences may reflect a developmental trend by which genetic influences increase and shared environmental differences decrease with age. Such a trend has been previously reported for general cognitive ability, with increasing heritability for 'g' continuing into adolescence and adulthood (McGue, Bouchard, Iacono, & Lykken, 1993). In the future we plan to test directly whether this longitudinal pattern also holds for language ability, using the longitudinal nature of the data in TEDS.

The statistical power granted by the relatively large sample size of this study allows us to conclude with confidence that most aspects of language are moderately influenced by additive genetic effects. However, an even larger sample size would be necessary to yield small enough confidence intervals to compare rigorously the estimates for the 9 measures. Nevertheless, it is intriguing that the only notable exceptions to the rule of moderate heritability were the zero heritability estimates for the phonological awareness and BAS Comprehension measures, at the low extreme of ability. We can only speculate about the reasons for this result: the two measures assess different aspects of language and rely on memory and on metalinguistic awareness to a different extent. However, relative to the other seven measures, they pose more selective attention demands in that the child has to select items from an array; they can also be better described as receptive than expressive measures. If it is the receptive nature of these measures that is associated with negligible genetic influence at the low end, our finding should replicate with other receptive measures in other samples. This would be consistent with the results of a study reported by the SLI Consortium (2002) in which a systematic genomewide QTL linkage analysis found linkages to specific (and different) loci for two expressive language measures, but no linkage for a receptive measure. A small genetic effect implies a large environmental effect, and an important avenue of exploration would be to identify the environmental mechanisms that may influence specific language abilities. In the case of phonological awareness, for example, a lack of exposure to reading experiences could be a critical factor pushing children into the low end of phonological ability (Castles & Coltheart, 2004). Indeed, some candidate environmental factors have been identified in a recent family study (Rutter, Thorpe, Greenwood, Northstone, & Golding, 2003; Thorpe, Rutter, & Greenwood, 2003). Encouraging the child to speak, providing elaborating comments, and engaging in reading to the child and talking about the story and illustrations were all found to be significantly related to the child's linguistic ability. Further research is needed to determine whether these factors are of particular importance for receptive language abilities, and whether genes and environments interact in such a way that genetic effects can be completely overwhelmed if the quality of the relevant environmental factors is very poor.

The third aim of this study was to explore sex differences in genetic and environmental influences on language ability and disability. Phenotypically, the present study found some evidence of mean differences in language level between the two sexes, but these accounted for only a small proportion of variance in all investigated aspects of language. The genetic analyses complemented this finding across our nine language measures, by showing neither qualitative nor quantitative sex differences in the etiology of individual differences. This was true both for the whole sample and the extreme end of the distribution, and suggests that the same genes and environments affect individual differences in various linguistic abilities in males and females to the same extent. These findings replicate the absence of sex differences in the composite measure (Viding et al., 2004) at the level of each of the individual measures comprising the composite measure. The small mean differences observed for males and females, which also lead to overrepresentation of males at the low end of the distribution (as shown in our extremes analyses), may stem from a small mean difference in sex-specific genetic make-up, or from differences in exposure to a particular environmental influence. For example, if talking to the child was found to influence lexical competence in children, we would expect girls to be more lexically competent on average if parents on average talked more to girls than boys.

Alternatively, a biological factor - such as hormones - could lead to mean sex differences; for instance, it has been hypothesized that gonadal hormones may be implicated in developmental language disabilities (Tallal, 1991). It is also possible that the observed sex differences are residuals of those found in very early development and are driven by factors affecting the timing of development. This hypothesis can be investigated by looking at the stability of children's performance longitudinally. Some indication of sex differences in transient as compared to persistent language delay has been found in 2-year-old twins who were assigned to groups on the basis of outcome at 3 and 4 (Dale et al., 2003). Although boys were nearly twice as likely as girls to be in the language delayed group at 2 years of age, sex was not found to be a significant predictor of a persistent delay. In other words, although boys are more likely to show early language delay, similar proportions of boys and girls continue to be language delayed at 3 and 4 years of age. Finally, it is possible that there are sex differences in the etiology of individual differences in language ability, but that these are so small that they can only be detected with samples of thousands of twins (Galsworthy, Dionne, Dale, & Plomin, 2000).

A general caveat in interpreting our results overall, is the possibility that there are twin-specific effects which may limit the generalisability of our findings to singletons. Previous research with toddlers and preschool children has consistently found mean differences between twins and singletons in their linguistic abilities in that twins show a language immaturity of about 3-6 months (e.g. Mittler, 1969; Rutter et al., 2003; Thorpe et al., 2003). A small delay in the early language development of twins in comparison to singletons was indeed found in TEDS (e.g. Dale, Simonoff, Bishop, Eley, Oliver, Price, Purcell, Stevenson, & Plomin, 1998). However, no delay was found in the four-year old twins in the present study (specifically, in the subgroup of 620 ‘controls’) as compared to standardization data of the McCarthy measure (Colledge et al., 2002). The twins’ scores on the Verbal Scale and General Cognitive Scale of the McCarthy Scales of Children’s Abilities (MSCA) were comparable with the norms given in the McCarthy manual (Colledge et al., 2002), although caution is warranted because the McCarthy norms were obtained more than 30 years ago. Furthermore, twins do not seem to show any distinctive pattern of linguistic organization, and the twin-specific delay is similar across different aspects of language with no differences in this respect between identical and fraternal twins (Mittler, 1969).

Determining the extent to which different linguistic abilities are influenced by genetic and environmental factors gives us a starting point for understanding their nature and origin. A logical next step is to determine how diverse aspects of language are related. This question cannot be addressed by univariate genetic analyses – for example, heritabilities can be the same for two traits but completely different sets of genes can affect the two traits. Multivariate genetic analyses are needed in order to advance our understanding of the extent to which different aspects of language are affected by the same genetic and environmental factors. Further work exploring commonalities and differences in the etiologies of the diverse language abilities discussed in the present study, are reported elsewhere (Hayiou-Thomas et al., submitted). Another direction for future research is multivariate genetic analysis of the links between these different aspects of language and other cognitive domains such as nonverbal ability and disability as well as learning disabilities such as reading and mathematics. For example, the finding of a genetic correlation of .63 between language and nonverbal abilities in 4 year-old twins using the present dataset (Colledge et al., 2002) suggests both substantial genetic overlap and language-specific genetic effects. This finding is supported by other TEDS research using parental assessment instruments (Dale et al., 2000). In fact, several studies addressing the issue of genetic overlap within and across cognitive domains (reviewed in Plomin & Kovas, in press) consistently find evidence for substantial overlap.

Establishing the role of genetic influences in diverse aspects of language is only a first step that provides a foundation and a motivation for molecular genetic studies to find the multiple specific genes involved. Similarly, establishing the relative importance of shared and nonshared environmental influences is just a first step toward future research to identify specific environmental sources of these components of variance. As specific genes and environments are identified, we can begin to understand the complex mechanisms through which genotypes interact with the environment to develop into phenotypes.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317-332.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150-177.

Bates, E., Dale, P.S., & Thal, D. (1995) Individual differences and their implications for theories of language development. In Fletcher and MacWhinney (Ed.) *Handbook of Child Language* (pp.???). Academic Press.

Bird, J., Bishop, D. V. M., & Freeman, N.H. (1995). Phonological Awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research*, *38*, 446-462.

Bishop, D.V.M. (1997). *Uncommon understanding: Development and disorders of language in children.* Hove: Psychology Press.

Bishop, D.V.M., Bishop, S.J., Bright, P, James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *Journal of Speech, Language & Hearing Research*, *42*, 155-68.

Bishop, D.V.M., Bright, P., James, C., Bishop, S.J. & Van der Lely, H. K. J. (2000). Grammatical SLI: A distinct subtype of developmental language impairment? *Applied Psycholinguistics, 21*, 159- 181.

Bishop, D.V.M. & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156-173.

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

Bradley, L. & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*, 419-421.

Byrne, B., Delaland, C., Fielding-Barnsley, R., Quain, P., Samuelsson, S., Høien, T., Corley, R., DeFries, J.C., Wadsworth, S., Willcutt, E., & Olson, R.K. (2002). Longitudinal twin study of early reading development in three countries: preliminary results. *Annals of dyslexia*, *52*, 49-73.

Castles, A. & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91(1)*, 77-111.

Cazden, C.B. (1966). On individual differences in language competence and performance. *Journal of Special Education, 1*, 135-150.

Cole, K.N., Dale, P.S. & Mills, P.E. (1990). Defining language delay in young children by cognitive referencing: Are we saying more than we know? *Applied Psycholinguistics, 11*, 291-302.

Cole, K.N., Dale, P.S., & Mills, P.E. (1992). Stability of the intelligence quotient-language quotient relation: Is discrepancy modeling based on myth? *American Journal of Mental Retardation, 97 (2)*, 131-145.

Colledge, E., Bishop, D. V. M., Dale, P. S., Koeppen-Schomerus, G., Price, T.S., Happpé, F., Eley, T. C., & Plomin, R. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology*, *38(5)*, 749-757.

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

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

Dale, P.S., Simonoff, E., Bishop, D.V.M., Eley, T.C., Oliver, B., Price, T.S., Purcell, S., Stevenson, J., & Plomin, R. (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 bi-directional effects of early lexical and grammatical development. *Child Development*, *74 (2)*, 394-412.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised.* Circle Pines, Minn.: American Guidance Service.

Elliot, C.D., Smith, P., & McCulloch, K. (1996). Verbal Comprehension Scale. In *British Ability Scales (2nd ed.).* Windsor, England: nferNelson.

Elliot, C.D., Smith, P., & McCulloch, K. (1997). *British Ability Scales (2nd ed.). Technical Manual.* Windsor, England: nferNelson.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking Innateness: A connectionist perspective on development.* The MIT Press, Cambridge, Massachussetts.

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.

Galsworthy, M.J., Dionne, G., Dale, P.S., and Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. *Developmental Science*, *3(2)*, 206-215.

Gathercole, S.E. & Baddeley, A.D. (1996). *The Children's Test of Nonword Repetition.* Psychological Corporation, UK.

Gleitman, L.R. (1988). Biological dispositions to learn language. In M.B. Franklin & S.S. Barten (Eds.), *Child language: A reader* (pp.158-175). New York: Oxford University Press.

Goad, H. & Ingram, D. (1987). Individual variation and its relevance to a theory of phonological acquisition. *Journal of Child Language, 14*, 419-432.

Goldman, R. & Fristoe, M. (1986). *Goldman Fristoe Test of Articulation.* Circle Pines, MN: American Guidance Service.

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

Hayiou-Thomas, M.E., Oliver, B., & Plomin, R. (in press). Genetic influences on specific versus non-specific language impairment in 4-year-old twins. *Journal of Learning Disabilities.*

Hayiou-Thomas, M., Kovas, Y., Harlaar, N., Bishop, D. V. M., Dale, P.S. & Plomin, R. (submitted). Common genetic aetiology for diverse language skills in 4½-year-old twins.

Hohnen, B, and Stevenson, J. (1999). The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology*, *35*, 590-603.

Huttenlocher, J., Haight, W, Bryk, A, Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27 (2)*, 236-248.

Karmiloff, K., & Karmiloff-Smith, A. (2001). *Pathways to Language.* Harvard University Press, Cambridge, Massachusetts, and London, England.

Lenneberg, E.H. (1967). *Biological foundations of language.* New York: John Wiley & Sons.

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

Matheny, A.P. & Bruggemann, C.E. (1973). Children's speech: Hereditary components and sex differences. *Folia Phoniatrica*, *25*, 442-49.

Mather, P.L. & Black, K.N. (1984). Hereditary and environmental influences on preschool twins' language skills. *Developmental Psychology*, *20*, 303-8.

Mc Carthy, D. (1972). *McCarthy Scales of Children's Abilities.* New York: The Psychological Corporation.

Meaburn, E., Dale, P.S., Craig, I.W., & Plomin, R. (2002). Language-impaired children: No sign of the FOXP2 mutation. *NeuroReport, 13(8)*, 1075-1077.

McGue, M., Bouchard, T.J., Iacono, W.G., & Lykken, D.T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin & G.E. Mc Clearn (Eds.), *Nature, nurture, and psychology* (pp.59-76). Washington, DC: American Psychological Association.

Mittler, P. (1969). Genetic aspects of psycholinguistic abilities. *Journal of Child Psychology and Psychiatry*, *10*, 165-176.

Neale, M.C. (1997). *Mx: Statistical Modeling.* Box 126 MCV, Richmond, VA 23298, Department of Psychiatry.

Neale, M.C., Eaves, L.J., & Kendler, K.S. (1994). The power of the classical twin study to resolve variation in threshold traits. Behavior Genetics, 24(3), 239-258.

Neale, M.C. & Maes, H.H.M. (2003). *Methodology for genetic studies of twins and families.* Kluwer Academic Publishers B.V. Dordrecht, The Netherlands.

Newbury, D. F., Bonora, E., Lamb, J. A., Fisher, S. E., Lai, C. S. L., Baird, G., Jannoun, L., Slonims, V., Stott, C. M., Merricks, M. J., Bolton, P. F., Bailey, A. J., Monaco, A. P., & International Molecular Genetic Study of Autism

Consortium. (2002). FOXP2 is not a major susceptibility gene for autism or specific language impairment. *American Journal of Human Genetics, 70,* 1318-1327.

Oliver, B., Dale, P.S., Saudino, K.J., Pike, A., & Plomin, R. (2002). The validity of a parent-based assessment of the nonverbal cognitive abilities in three-year-olds. *Early Child Development & Care, 172(4)*, 337-348.

Plomin, R., DeFries, J.C., McClearn, G.E., & McGuffin, P. (2001). *Behavioural Genetics (4th ed).* New York, Worth.

Plomin, R. & Kovas, Y. (in press). Generalist genes and learning disabilities. *Psychological Bulletin*,

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science*, *264*, 1733-1739.

Renfrew, C.E. (1997a). *Bus Story Test – A test of narrative speech (4th ed.).* Bicester, Oxon, England: Winslow Press.

Renfrew, C.E. (1997b). *Action Picture Test (4th ed.).* Bicester, Oxon, England: Winslow Press.

Rice, M.L. (2003). A unified model of specific and general language delay: Grammatical tense as a clinical marker of unexpected variation. In Y. Levy & J. Schaeffer (Eds.), Language competence across populations: Toward a definition of specific language impairment. Mahwah, NJ: Lawrence Erlbaum Associates (pp63-95).

Robinson, R. (1987). The causes of language disorder: introduction and overview. *Proceedings of the First International Symposium on Specific Speech and Language Disorders in children,* (pp.1-19). London: Association for All Speech Impaired Children.

Rutter, M., Thorpe, K., Greenwood, R., Northstone, K., & Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay: I: Design; twin-singleton differences in language, and obstetric risks. *Journal of Child Psychology and Psychiatry, 44(3)*, 326-341.

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 cognitive abilities of 2-year-olds. *British Journal of Developmental Psychology, 16*, 349-363.

Sham, P. (1998). *Statistics in Human Genetics.* London: Edward Arnold.

Spinath, F.M., Price, T.S., Dale, P.S, & Plomin, R. (2004). The genetic and environmental origins of language disability and ability. *Child Development, 75(2),* 445-454.

Stromswold, K. (2001). The heritability of Language: a review and metaanalysis of twin, adoption, and linkage studies. *Language, 77(4)*, 647-723.

Tallal, P. (1991). Hormone influences in developmental learning disabilities. *Psychoneuroendocrinology, 16*, 203-211.

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

Thompson, L.A., Detterman, D.K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychological Science, 2*, 158-65.

Thorpe, K., Rutter, M., & Greenwood, R. (2003). Twins as a natural experiment to study the causes of mild language delay: II: Family interaction risk factors. *Journal of Child Psychology and Psychiatry, 44(3)*, 342-355.

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*, 444-448.

Van der Lely, H. K. J., Rosen, S., & McClelland, A. (1998). Evidence for a grammar-specific deficit in children. *Current Biology, 8(23)*, 1253-1258.

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*, 45(2), 315-325.

Table 1. Means (Standard deviations) for the nine measures for the whole sample and for the tests' norms; and scores corresponding to the –1 SD cut-off.

	Mean (SD) from the Test Norms	Mean (SD) for the whole sample	Cut-off score at –1 SD for the probands in the study
Bus Information	18.3 (7.0)	15.1 (9.9) N=1530	5
AP-Grammar	20 (5.8)	16.4 (7.0) N=1515	9
BAS Comprehension[1]	~ 8.5 (n/a)	7.39 N=1562	5
Word Knowledge[2]	14.5 (2.9)	12.1 (3.4) N=1574	8
Verbal Fluency	11.1 (5.2)	10.9 (6.0) N=1574	5
Verbal Memory	20.1 (6.5)	14.7 (8.2) N=1574	6
Phon. Awareness	(n/a)	4.3 (2.2) N=1517	2
GF-Articulation[3]	~9.5 (n/a)	9 N=1554	16
NW-Repetition	12 (n/a)	11.1 (5.4) N=1478	5

Note: n/a = not available. Bus Information = Bus Story Test total information score, AP-Grammar = Action Picture Test grammar score, BAS Comprehension = BAS Verbal Comprehension, Phon. Awareness = Phonological awareness task, GF-

Articulation = Goldman-Fristoe Test of Articulation, NW-Repetition = Nonword repetition task. [1]For the BAS Comprehension Test, the mean score in this study represents a point at approximately the 25th percentile on the normative distribution; the cut-off of 5 corresponds to between 6th and 7th percentile. [2]For the MSCA Word Knowledge subtest: although only oral vocabulary subtest was analyzed in this study, picture vocabulary scores had to be included in this analysis, because the normative information was available for the composite of the two subtests. [3]For GF-Articulation (Sounds-in-Words Subtest) the statistics refer to the number of errors; the mean score represents approximately the 46th percentile on the normative distribution; the cut-off of 16 corresponds to approximately the 20th percentile.

Table 2. Means (Standard Deviations) and ANOVA results by sex and zygosity for the nine measures.

	MZ	DZ	Males	Females	MZ m	DZ m	MZf	DZ f	DZ opp	ANOVA		
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	Sex	Zyg	Sex*Zyg
Bus Info	-.10	.06	.05	-.05	.00	.08	-.23	.10	.02	p = .540	p =.003	p = .112
	(.99)	(1.00)	(.97)	(1.03)	(.97)	(1.00)	(1.00)	(1.00)	(1.00)	η^2 =.002	η^2 =.008	η^2 =.003
AP-	-.13	.07	-.08	.09	-.15	-.02	-.10	.19	.06	p =.001	p =.001	p = .240
Grammar	(1.02)	(.98)	(.99)	(1.00)	(.99)	(1.02)	(1.05)	(.95)	(.97)	η^2=.007	η^2 =.009	η^2 =.002
BAS	-.09	.05	-.06	.07	-.11	-.04	-.06	.09	.08	p =.012	p =.023	p = .416
Comp.	(1.02)	(.99)	(1.01)	(1.00)	(.98)	(1.02)	(1.06)	(.96)	(.97)	η^2=.004	η^2 =.005	η^2 =.001
Word	-.17	.09	-.02	.02	-.17	-.07	-.16	.07	.12	p =.579	p =.000	p = .890
Know.	(.92)	(1.03)	(.99)	(1.02)	(.90)	(1.01)	(.95)	(1.02)	(1.05)	η^2=.000	η^2 =.016	η^2 =.000
Verbal	-.11	.06	-.07	.08	-.12	-.02	-.09	.12	.08	p =.002	p =.005	p = .072
Fluency	(.97)	(1.01)	(1.01)	(.98)	(.98)	(1.03)	(.96)	(.98)	(1.01)	η^2=.006	η^2 =.007	η^2 =.003

Verbal	-.15	.08	-.07	.08	-.12	-.06	-.18	.24	.08	p =.003	p =.000	p = .007
Memory	(.99)	(1.00)	(.98)	(1.02)	(.96)	(.98)	(1.03)	(0.92)	(1.03)	η^2=.006	η^2 =.013	η^2 =.006
Phon.	-.10	.06	-.06	.06	-.16	-.05	-.03	.15	.07	p =.030	p =.017	p = .234
	(.96)	(1.02)	(.97)	(1.03)	(.91)	(1.02)	(1.01)	(1.03)	(1.00)	η^2=.003	η^2 =.005	η^2 =.002
GF-	-.10	.06	-.17	.19	-.29	-.12	.12	.14	.12	p =.000	p =.004	p = .367
Artic.	(1.03)	(.98)	(.99)	(.98)	(.97)	(.99)	(1.06)	(.92)	(.99)	η^2=.032	η^2 =.007	η^2 =.001
NW-Rep.	-.09	.05	-.14	.15	-.20	-.07	.04	.11	.09	p =.000	p =.027	p =.067
	(1.04)	(.97)	(1.00)	(.97)	(1.02)	(.99)	(1.05)	(.96)	(.96)	η^2 =.002	η^2 =.005	η^2 =.004

Note: All calculations are based on age-corrected scores. MZ = monozygotic twins (N: 534-562), DZ = dizygotic twins (same and opposite sex) (N: 943-1012), Males = all male twins (N: 776-835), Females = all female twins (N: 702-739), MZm = monozygotic males (N: 291-310), DZm = dizygotic males (N: 260-294), MZf = monozygotic females (N: 239-252), DZf = dizygotic females (N: 247-256), DZopp = dizygotic opposite sex twins (N: 436-462); Zyg = zygosity; Bus Info = Bus Story Test total information score, AP-Grammar = Action Picture Test grammar score, BAS Comp. = BAS Verbal Comprehension, Word Know. = Word Knowledge, Phon. = Phonological awareness task, GF- Artic. = Goldman-Fristoe Test of Articulation, NW-Rep = Nonword repetition task. The N represents the numbers of individuals.

Table 3. Intraclass correlations by sex and zygosity for the nine measures.

	MZ	DZ	MZ m	DZ m	MZ f	DZ f	DZ opp
	ICC (CI)	*ICC (CI)*	*ICC (CI)*	*ICC (CI)*	*ICC (CI)*	*ICC (CI)*	*ICC (CI)*
Bus Info	.65	.40	.68	.37	.62	.57	.35
	(.58 - .72)	*(.33 - .48)*	*(.58 - .75)*	*(.21 - .50)*	*(.50 - .72)*	*(.44 - .68)*	*(.23 - .46)*
AP-Grammar	.57	.40	.63	.48	.50	.28	.39
	(.48 - .65)	*(.32 - .47)*	*(.52 - .72)*	*(.33 - .60)*	*(.35 - .62)*	*(.10 - .43)*	*(.27 - .50)*
BAS Comp.	.51	.35	.45	.31	.57	.37	.35
	(.42 - .60)	*(.27 - .42)*	*(.32 - .57)*	*(.15 - .45)*	*(.43 - .68)*	*(.21 - .52)*	*(.23 - .46)*
Word Know.	.59	.36	.56	.43	.59	.46	.25
	(.51 - .66)	*(.28 - .44)*	*(.44 - .66)*	*(.29 - .56)*	*(.47 - .70)*	*(.32 - .59)*	*(.13 - .37)*
Verbal Fluency	.50	.30	.53	.26	.46	.41	.25
	(.41 - .58)	*(.22 - .38)*	*(.40 - .63)*	*(.11 - .41)*	*(.32 - .60)*	*(.25 - .54)*	*(.13 - .37)*
Verbal Memory	.54	.35	.52	.51	.56	.36	.25

	(.45 - .62)	*(.27 - .43)*	*(.39 - .62)*	*(.38 - .62)*	*(.43 - .67)*	*(.20 - .50)*	*(.12 - .36)*
Phon.	.42	.26	.34	.16	.50	.25	.34
	(.32 - .52)	*(.17 - .34)*	*(.19 - .48)*	*(-.01 - .32)*	*(.35 - .62)*	*(.08 - .41)*	*(.22 - .45)*
GF-Artic.	.63	.41	.59	.35	.67	.40	.40
	(.55 - .69)	*(.34 - .48)*	*(.47 - .68)*	*(.20 - .49)*	*(.56 - .75)*	*(.24 - .54)*	*(.28 - .50)*
NW-Rep.	.52	.28	.49	.22	.57	.27	.26
	(.43 - .61)	*(.20 - .37)*	*(.35 - .60)*	*(.04 - .38)*	*(.43 - .68)*	*(.10 - .43)*	*(.13 - .38)*

Note: All correlations are based on age-corrected scores. Confidence intervals are presented in brackets. MZ (N: 257-281), DZ (N: 446-506), MZm (N: 140-155), DZm (N: 121-147), MZf (N: 116-126), DZf (N: 120-128), DZopp (N: 220-250). The N represents the number of twin pairs.

Table 4. Individual differences analyses. Parameter estimates (and confidence intervals) from the model of best fit for the nine measures.

	χ^2	Df	Probability	AIC	RMSEA	a^2	c^2	e^2
Bus Story Information	18.94	13	.125	-7.061	.018	.53 *(.35-.70)*	.13 *(.00-.28)*	0.34 *(.28-.40)*
AP-Grammar	13.03	13	.446	-12.971	.000	.29 *(.09-.50)*	.26 *(.09-.42)*	.45 *(.38-.53)*
BAS Comp.	11.53	13	.567	-14.474	.000	.30 *(.07-.51)*	.21 *(.03-.37)*	.50 *(.42-.59)*
Word Knowledge	11.15	13	.598	-14.852	.000	.52 *(.32-.67)*	.09 *(.00-.25)*	.39 *(.33-.46)*
Verbal Fluency	7.09	13	.898	-18.913	.000	.40 *(.17-.58)*	.11 *(.00-.28)*	.49 *(.42-.58)*
Verbal Memory	22.13	13	.053	-3.866	.026	.36 *(.15-.57)*	.17 *(.00-.33)*	.47 *(.40-.55)*

Phonological Awareness	9.95	13	.698	-16.053	.003	.38 *(.13-.53)*	.06 *(.00-.24)*	.56 *(.47-.66)*
GF-Articulation	7.67	13	.864	-18.326	.000	.37 *(.19-.56)*	.24 *(.08-.38)*	.39 *(.33-.46)*
Nonword Repetition	8.27	13	.825	-17.729	.000	.41 *(.18-.57)*	.09 *(.00-.27)*	.50 *(.43-.59)*

Note: The best-fitting model does not allow for sex differences, i.e. male and female variance component estimates were constrained to be equal (quantitative differences removed) and both genetic correlations and shared environment correlations are fixed to .5 and 1 respectively (qualitative differences removed).

Table 5. Probandwise twin concordances and number of affected individuals (probands) for verbal disability at 15 % cut-off organized by sex and zygosity for the nine measures.

	MZ	DZ	MZ m	DZ m	MZ f	DZ f	DZ opp
Bus Info	58%	24%	57%	14%	59%	34%	25%
	n = 100	*n* = 150	*n* = 42	*n* = 43	*n* = 58	*n* = 35	*n* = 72
AP-Grammar	57%	42%	62%	58%	52%	38%	31%
	n = 105	*n* = 130	*n* = 55	*n* = 45	*n* = 50	*n* = 26	*n* = 59
BAS Comp.	42%	41%	36%	37%	49%	44%	42%
	n = 90	*n* = 137	*n* = 45	*n* = 43	*n* = 45	*n* = 32	*n* = 62
Word Know.	62%	36%	60%	45%	63%	27%	35%
	n = 110	*n* = 139	*n* = 53	*n* = 40	*n* = 57	*n* = 30	*n* = 69
Verbal	55%	26%	54%	29%	57%	33%	21%
Fluency	*n* = 105	*n* = 113	*n* = 59	*n* = 56	*n* = 46	*n* = 30	*n* = 67

Verbal Memory	47%	34%	41%	47%	54%	34%	25%
	n = 131	*n* = 164	*n* = 68	*n* = 55	*n* = 63	*n* = 29	*n* = 80
Phon.	24%	28%	24%	24%	23%	30%	31%
	n = 93	*n* = 163	*n* = 50	*n* = 51	*n* = 43	*n* = 40	*n* = 72
GF-Artic.	64%	43%	58%	51%	73%	22%	44%
	n = 113	*n* = 155	*n* = 72	*n* = 55	*n* = 41	*n* = 27	*n* = 73
NW-Rep.	46%	27%	41%	33%	53%	28%	24%
	n = 114	*n* = 153	*n* = 69	*n* = 49	*n* = 45	*n* = 36	*n* = 68

Table 6. Extremes analyses. Parameter estimates (and confidence intervals) from the liability-threshold model of best fit for the nine measures.

	χ^2	df	Probability	AIC	RMSEA	a^2	c^2	e^2
Bus Story Information	19.64	12	.074	-4.361	.033	.69 *(.39-.81)*	.00 *(.00-.23)*	.31 *(.19-.47)*
AP-Grammar	24.28	12	.019	0.283	.029	.21 *(.00-.63)*	.48 *(.13-.72)*	.30 *(.18-.46)*
BAS Comp.	9.61	12	.651	-14.395	.000	.00 *(.00-.41)*	.56 *(.23-.67)*	.44 *(.29-.57)*
Word Knowledge	13.17	12	.357	-10.832	.012	.51 *(.10-.85)*	.25 *(.00-.58)*	.24 *(.14-.38)*
Verbal Fluency	20.87	12	.052	-3.129	.031	.67 *(.33-.80)*	.00 *(.00-.27)*	.33 *(.20-.48)*
Verbal Memory	26.91	12	.008	2.910	.039	.19 *(.00-.64)*	.32 *(.00-.56)*	.49 *(.34-.65)*

Phonological Awareness	7.39	12	.831	-16.614	.000	.00 *(.00-.39)*	.22 *(.00-.37)*	.78 *(.60-.93)*
GF-Articulation	13.24	12	.352	-10.760	.011	.36 *(.00-.74)*	.41 *(.08-.70)*	.22 *(.13-.36)*
Nonword Repetition	12.00	12	.446	-12.000	.015	.41 *(.00-.66)*	.09 *(.00-.48)*	.50 *(.33-.69)*

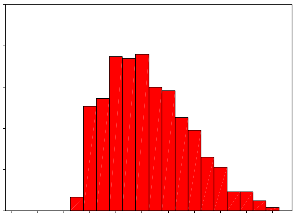
Note: In the model male and female parameter estimates are constrained to be equal (quantitative differences removed) and both genetic correlation and shared environment correlations are fixed to .5 and 1 respectively (qualitative differences removed). The thresholds are not equated for males and females.

Figure Legends

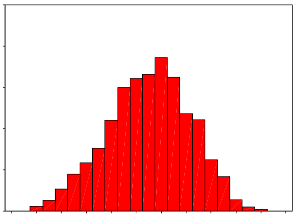
Figure 1. The distribution of age-corrected scores for the nine measures: a-The Renfrew Bus Story Information Test; b-The Renfrew Action Picture Grammar Test; c- The BAS Verbal Comprehension subtest; d- The MSCA Word Knowledge subtest; e- The MSCA Verbal Fluency subtest; f- The MSCA Verbal Memory Words and Sentences subtest; g- The Phonological Awareness task; h- The Goldman-Fristoe Test of Articulation (log-transformed); i- The Nonword Repetition task. N = 1478 - 1574.

Figure 2a**.** Sex-limitation model-fitting results for individual differences: proportions of variance explained by additive genetic (a^2), shared environmental (c^2) and nonshared environmental factors (e^2) for the nine measures.

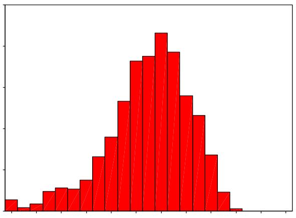
Figure 2b. Liability-threshold sex-limitation model-fitting results for dichotomous analysis using a 16% liability threshold: proportions of variance explained by additive genetic (a^2), shared environmental (c^2) and nonshared environmental factors (e^2) for the nine measures.



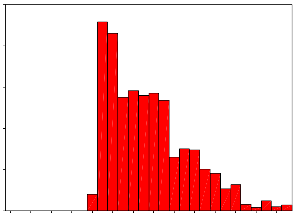
a



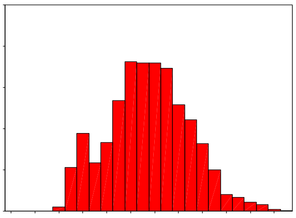
b



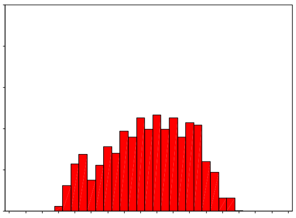
c



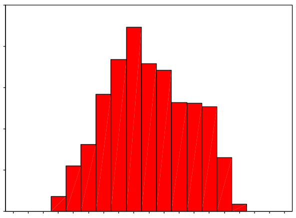
d



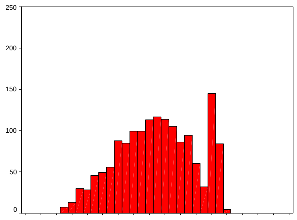
e



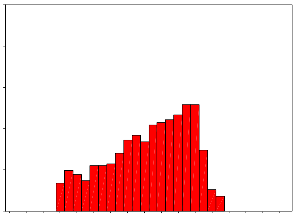
f



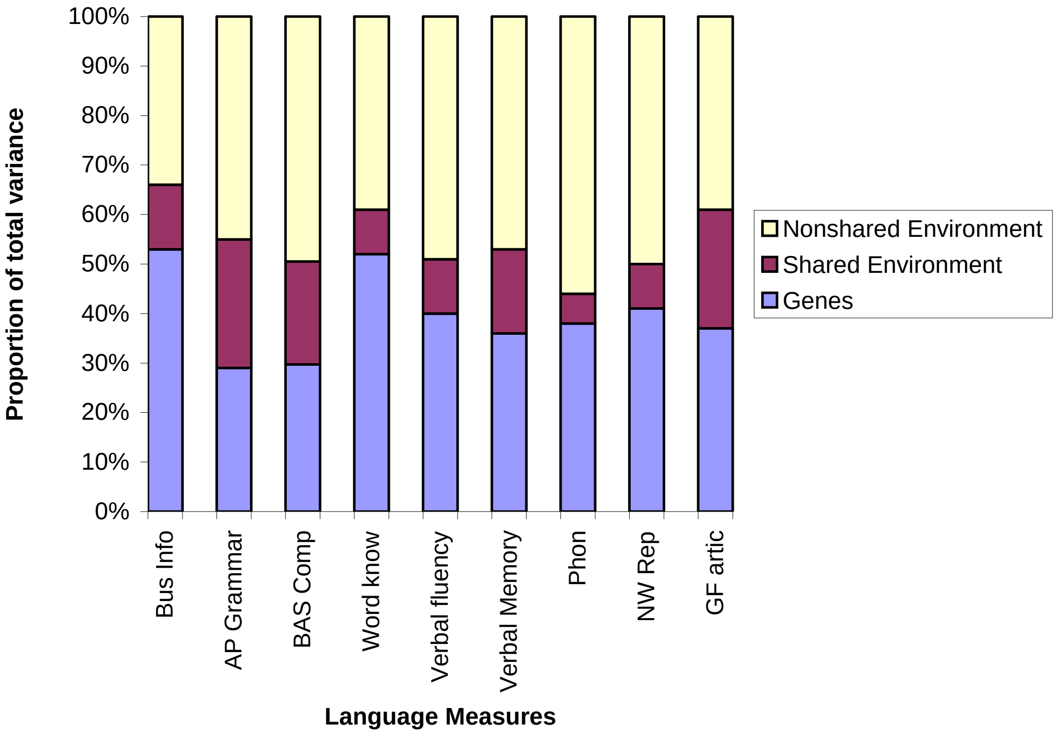
g

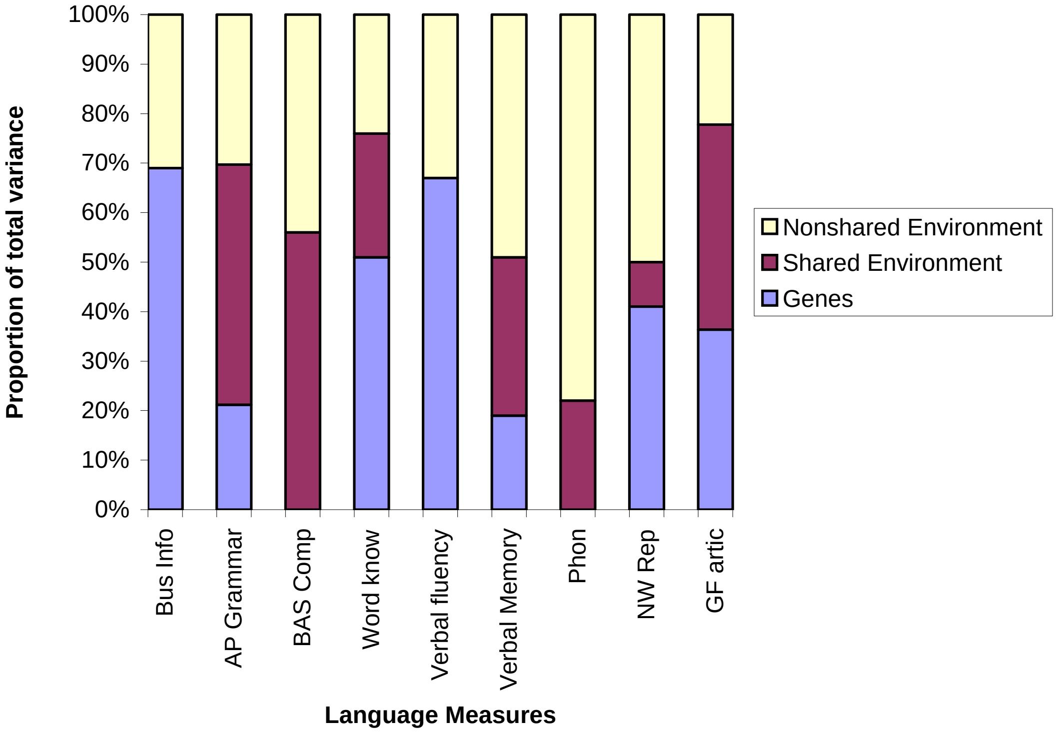


h



i


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Proportion of total variance
Bus Info
AP Grammar
BAS Comp
Word know
Verbal fluency
Verbal Memory
Phon
NW Rep
GF artic
Language Measures
Nonshared Environment
Shared Environment
Genes


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Proportion of total variance
Bus Info
AP Grammar
BAS Comp
Word know
Verbal fluency
Verbal Memory
Phon
NW Rep
GF artic
Language Measures
Nonshared Environment
Shared Environment
Genes