Article:
Hayiou-Thomas, M.E., Kovas, Y., Harlaar, N. et al. (2006) Common aetiology for diverse language skills in 41/2-year-old twins. Journal of Child Language. pp. 339-368. ISSN: 0305-0009

https://doi.org/10.1017/S0305000906007331

J. Child Lang. **33** (2006), 339–368. © 2006 Cambridge University Press
doi:10.1017/S0305000906007331 Printed in the United Kingdom

Common aetiology for diverse language skills in $4\frac{1}{2}$-year-old twins*

MARIANNA E. HAYIOU-THOMAS
Department of Psychology, University of York

YULIA KOVAS, NICOLE HARLAAR
AND ROBERT PLOMIN
Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London

DOROTHY V. M. BISHOP
Department of Experimental Psychology, University of Oxford

AND

PHILIP S. DALE
Department of Communication Science and Disorders, University of Missouri-Columbia

(*Received* 29 *October* 2004. *Revised* 23 *September* 2005)

ABSTRACT

Multivariate genetic analysis was used to examine the genetic and environmental aetiology of the interrelationships of diverse linguistic skills. This study used data from a large sample of $4\frac{1}{2}$-year-old twins who were tested on measures assessing articulation, phonology, grammar, vocabulary, and verbal memory. Phenotypic analysis suggested two latent factors: articulation (2 measures) and general language (the remaining 7), and a genetic model incorporating these factors provided a good fit to the data. Almost all genetic and shared environmental influences on the 9 measures acted through the two latent factors. There was also substantial aetiological overlap between the two latent factors, with a genetic correlation of 0·64 and shared environment correlation of 1·00. We conclude that to a large extent, the same genetic

---

[*] We gratefully acknowledge the ongoing contribution of the parents and children in the Twin Early Development Study (TEDS). TEDS is supported by a programme grant (G9424799) from the UK Medical Research Council. Address for correspondence: Dr M. E. Hayiou-Thomas, Department of Psychology, University of York, Heslington, York YO10 5DD. e-mail: m.hayiou-thomas@psychology.york.ac.uk

and environmental factors underlie the development of individual differences in a wide range of linguistic skills.

INTRODUCTION

The relationship between diverse language abilities is a major issue in the study of child language development. The current paper addresses this topic from an aetiological perspective, using twin methodology to estimate the extent to which genetic and environmental factors underpin the covariation between diverse areas of linguistic skill in young children, from articulation and phonology to lexicon and grammar.

The issue of the relationships between different language components has usually been discussed at the species-universals level of analysis, but the same question can and should be addressed at the individual differences level. Substantial and stable individual variation occurs not only within components of early language, with some children acquiring vocabulary items or grammatical structures faster than other children, but it also occurs across language components. Developmental asynchrony between components such as the lexicon and grammar, for example, would suggest that they draw on different cognitive and neural mechanisms. Alternatively, evidence that they vary in tandem would suggest that individual differences in these components of language rely on similar mechanisms (Bates, Dale & Thal, 1995).

Much of the debate in the area of language acquisition has centred on the separation of the lexicon from a rule-based grammatical system. While dual-route accounts based in generative linguistics have traditionally focused on the species-universal level of analysis, some recent accounts have predicted similar patterns at the individual differences level. A recent account that takes this perspective is the procedural/declarative model of lexicon and grammar (Ullman, 2001), which draws a parallel between these well-established memory systems and the mental representation of semantic and syntactic information. In this model, the lexicon is stored in an associative network which utilizes the same mechanisms used to represent any other set of arbitrarily related information. Learning and using grammatical rules, on the other hand, depends on similar procedural mechanisms to those used in implicit learning of other cognitive and motor skills.

An alternative viewpoint argues for a much closer relationship between the lexicon and syntax. Those proposing this view note the striking correlations found between measures of early grammatical and lexical ability. One example is from the large cross-sectional norming study for the MacArthur Communicative Development Inventory (CDI: Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994), in which the linear correlation between the grammatical and vocabulary checklists was 0·84, and remained very strong when age was partialled out of the correlation. Similar results were obtained from longitudinal studies using language samples from free play and structured assessments in the laboratory which augmented the parent-report data (Goodman, 1995). Furthermore, individual growth curves for both late and early talkers showed that the level of grammatical complexity achieved reflects their lexical abilities and not their age (Thal, Bates, Goodman & Jahn-Samilo, 1997).

On the basis of these kinds of results, as well as connectionist simulations of the feasibility of such a mechanism (e.g. Plunkett & Marchman, 1993), it seems plausible that a certain ‘critical mass’ of vocabulary may be necessary for the emergence of combinatorial language and subsequent grammatical development. This version of events has been termed ‘lexical bootstrapping’, suggesting that vocabulary is the foundation for grammar. A complementary perspective that also highlights the connection between lexical and grammatical development is ‘syntactic bootstrapping.’ Gleitman (1990) suggests that young children can use syntactic knowledge – for example, the kinds of words that appear in certain parts of a sentence – to narrow down the possibilities for the meaning of a new word. These processes imply closely interlinked acquisition mechanisms for the early lexicon and syntax. In a similar vein, Tomasello’s (1992) VERB ISLAND HYPOTHESIS posits that early grammatical development is completely lexically specific. Thus, rather than using abstract categories like ‘subject’ and ‘object’, children use verb-specific concepts like ‘pusher’ and ‘pushee’; these then provide a platform from which processes of generalization and abstraction begin to create syntactic categories.

Despite the emphasis on semantics and syntax, these theoretical accounts also make some interesting predictions about the role of phonology in early language acquisition. For example, in addition to semantic and syntactic bootstrapping, there is also evidence for ‘phonological bootstrapping’, which utilises prosodic, phonetic and phonotactic information in the speech signal as a cue to both semantic and syntactic structures (Morgan & Demuth, 1996). From this perspective, phonology is considered to be closely related to both semantics and syntax, which in turn are closely related to each other. A different prediction is made by the procedural/declarative model discussed earlier, in which phonology is considered to be a rule-governed system like syntax, and is therefore part of the procedural system. Here, phonology is predicted to be related to syntax, but less so to semantics.

A third account focuses explicitly on the role of phonology, and in particular phonological working memory (Baddeley, Gathercole & Papagno, 1998). According to this theory, the phonological loop system of human working memory is a key cognitive resource for language learning. It is a short-term memory system that enables people to hold strings of speech sounds in a buffer, and it is argued that this facilitates both vocabulary acquisition and syntactic parsing. There is now a large body of empirical evidence (reviewed in Baddeley *et al.*, 1998) showing that individual variation in phonological loop functioning – usually measured by a nonword repetition task – is associated with novel word learning in both children and adults, and that deficits in nonword repetition are an excellent marker of specific language impairment (Botting & Conti-Ramsden, 2001). According to this account, phonological skills – or at least phonological memory – should be closely associated with semantic and syntactic skills.

Genetically informative designs provide a powerful technique for testing predictions from different theoretical perspectives, within an individual differences framework. Not only can we test to what extent genetic and environmental factors are likely to play a role in individual differences for any given area of language, but we can also evaluate the extent to which aetiological factors overlap for different areas of language. That is, are the same genes and/or the same environmental factors involved for different areas?

This approach is illustrated by a twin study that used multivariate behavioural genetic techniques to show a substantial (61%) genetic overlap between vocabulary and grammar in two-year-olds as assessed by their parents (Dale, Dionne, Eley & Plomin, 2000). Later longitudinal analyses that incorporated data on vocabulary and grammar when the twins were three years old, found that there were strong genetic correlations from two-year-old grammar to three-year-old vocabulary, and from two-year-old vocabulary to three-year-old grammar, providing evidence for both lexical and syntactic bootstrapping in this age group (Dionne, Dale, Boivin & Plomin, 2003). Both the concurrent and the longitudinal relationships between these two areas of language appear to be mediated to a large degree by shared genetic factors. Such shared genetic aetiology is consistent with a common mechanisms hypothesis for language development. However, it is incompatible with dual-route accounts of lexicon and grammar, at least where these make predictions about individual differences.

Recent work from our group, the Twins Early Development Study in the UK, examined the univariate heritabilities of nine measures, covering a wide range of linguistic skills, in a large sample of twins aged 4;6 (Kovas, Hayiou-Thomas, Oliver, Dale, Bishop & Plomin, 2005). The conclusion drawn from those analyses was that all nine measures demonstrated moderate heritability, for both the normal range and the low end of performance, and that none of the measures stood out as remarkably more or less heritable than any of the others (with a possible exception for disorder of receptive language, which showed low heritability). This complemented previous work on a subset of the same sample, showing moderate heritability for a composite of these language measures (Colledge, Bishop, Koeppen-Schomerus, Price, Happe, Eley, Dale & Plomin, 2002). However, univariate heritability analyses do not provide information on the relationships between measures: it is possible to find the same magnitude of genetic and environmental influence on two measures even if the specific genes and environments influencing them are completely different; the converse is also true, so that exactly the same aetiological factors can have a different magnitude of effect on different measures. The current paper therefore uses multivariate analyses to examine the degree to which the genetic and environmental factors influencing individual differences in linguistic abilities – from phonology to lexicon to grammar – are the same or different.

METHOD

Participants

The sampling frame for the present study was the Twins Early Development Study (TEDS), a longitudinal study of twins born in England and Wales in 1994, 1995 and 1996 (Trouton, Spinath & Plomin, 2002). After checking for infant mortality, all families identified by the UK Office for National Statistics (ONS) as having twins born in these years were invited to participate in TEDS when the twins were about 18 months old. The twins were assessed at 2, 3, and 4 years of age using parent questionnaires, which included measures of language, cognitive, and behavioural development.

A subset of TEDS twins was tested at home on an extensive battery of language and nonverbal measures, at age 4;6 (*S.D.* = 0;2). This subsample was selected on the basis of parent report at age 4 on measures of vocabulary & grammar (MacArthur-Bates Communicative Development Inventory UK Short Form – MCDI:UKSF; Dale, Price, Bishop & Plomin, 2003), and nonverbal ability (Parent Report of Children's Abilities – PARCA: Oliver, Dale, Saudino, Petrill, Pike & Plomin, 2002). Twin pairs were excluded where either member of the pair had any major medical or perinatal problems, documented hearing loss, or organic brain damage. Participants were selected to be ethnically white, in order to avoid ethnic stratification in molecular genetic studies using DNA from this group; however, over 94% of the population of England and Wales is also white. Maternal education levels were also comparable both to the overall TEDS sample, as well as UK ONS census data. In all selected families, English was the only language spoken at home.

A total of 1672 children (836 twin pairs) participated in the in-home testing; 98 children (49 pairs) were excluded from any analyses either on the grounds of serious medical conditions discovered at the time of the in-home visit, or because of missing data from one of the twins in a pair, or because of uncertainty regarding twin zygosity. Data were standardized (*M* = 0, *S.D.* = 1) on 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. The analyses presented in the current paper were based on same-sex pairs only, giving a total of 556 twin pairs (1112 individuals).

By design, this sample includes a large proportion of children with language difficulties, as a major aim of the overall study (although not the current set of analyses) is to elucidate the causes of language impairment. This may affect the generalizability of our findings, and should be borne in mind when interpreting our results. However, despite the oversampling at the low end, the distributions of scores for each of the measures are all unimodal, and in most cases near-normal, and the means are all within one standard deviation of the published age-norms (Kovas *et al.*, 2005).

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 1 hr 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). Each co-twin was assessed by a different tester.

Measures

The verbal battery was chosen on the basis of the following criteria: tests should be suitable for four-year-olds, should show variation across the range of ability at this age, and should have established psychometric properties. Furthermore, the tests were chosen to differ from each other with respect to the main 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 four-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 our 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, devised to measure children's knowledge of morphological rules. Children are presented with a nonsense label ('wug') for a single novel object, and asked what they would be called if there were two of these objects ('wugs'). Although good performance on this test provides evidence of knowledge of formation rules for noun 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 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 first part of 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, 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 is 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 s. 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, 1997*a*) 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, as suggested by Renfrew *et al.* 1997, which reflects the story content that the child includes in their 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 could potentially be biased in favour of finding commonalities between semantics and syntax if the same narrative was used to index both domains. Although we have categorised 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.

Expressive syntax

The Renfrew Action Picture Test (Renfrew, 1997*b*), Grammar score. This is an elicitation task designed to solicit utterances containing different types of grammatical construction. 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 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 highly correlated in our sample (0·77), However, because we did not want to bias our results in favour of finding associations between syntax and semantics, we used only the grammar measure from this task. Phonological accuracy of utterances is not taken into account when scoring AP Grammar, although it must be acknowledged that a child with an expressive phonological impairment could be handicapped by problems in producing inflected forms.

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*; determiner *a*, *the*.

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, and the child is awarded 1 point for each successfully repeated key word, and a maximum of 30 points was 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 both receptive phonological ability as well as expressive phonology, and is not explicitly adjusted for articulatory accuracy. A 20-item version of the test was used, with ten items at each of the 2 and 3 syllable lengths. 1 point is awarded for a correct response, and 0 for an incorrect response, with a maximum possible raw score of 20.

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 of the important role this aspect of language function plays in literacy development. 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 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. Four 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. For example: 'Which of these things would Lynn like?' 'Chair?' 'Bin?'. The child responds by picking up the chosen answer card and placing it in a special box. After 4 practice trials with feedback a further eight items are administered. 1 point is awarded for each correct response, with a maximum possible raw score of 8 points.

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 standardisation, reliability and validity of each test can be found in the published manuals.

Analyses

All measures were standardized for the entire sample of 1574 children to a mean of 0 and standard deviation of 1. Scores were corrected for the linear effects of age and sex, as these can inflate twin similarity (McGue & Bouchard, Jr., 1984). Same-sex twins only were included in the analyses, to simplify computational demands; although this precludes examination of differential aetiology for males and females, we note that no such sex effects have been found for any of the measures included here (Kovas *et al.*, 2005).

Phenotypic analyses included bivariate intraclass correlations between measures to gain a first indication of their interrelationships. Exploratory factor analysis with Oblimin oblique rotation was then carried out to examine the factor structure underlying these relationships. As twin data are not independent, all phenotypic analyses were based on half the sample, which included one randomly selected member of each twin pair.

Genetic analyses were based on the twin design, which capitalises on the fact that identical (MZ for monozygotic) twins share 100% of their varying DNA while fraternal twins (DZ for dizygotic) share on average 50%, just like any other sibling pair (Plomin, DeFries, McClearn & McGuffin, 2001). If the members of an MZ twin pair are more similar to each other on a given trait than the members of a DZ pair, this difference can be attributed to genetic influences. Comparing the members of MZ and DZ twin pairs on a single trait yields an estimate of univariate heritability, which is a measure of the extent to which variance on the trait can be attributed to genetic as opposed to environmental factors. It is possible to extend this model to examine the aetiological relationship between two (or more) variables, by comparing Trait 1 in Twin 1 to Trait 2 in Twin 2. If the cross-trait cross-twin correlation is higher in MZ than in DZ pairs, this is evidence for some shared genetic relationship (Dale *et al.*, 2000). This bivariate analysis yields an estimate of bivariate heritability, which is the extent to which the covariance between two traits can be attributed to genetic factors. Bivariate heritability is the genetic correlation between the two traits weighted by their heritabilities (Plomin *et al.*, 2001). The genetic correlation is the correlation between genetic factors on the two traits regardless of their heritability or their phenotypic correlation. That is, the heritability of the two traits could be low and the phenotypic correlation between them could be low but the genetic correlation between them could be high. The genetic correlation can be viewed roughly as the extent to which the same genes affect the two traits.

The actual derivation of the parameters was calculated using the models described below and the structural equation modelling package Mx (Neale, Boker, Xie & Maes, 2002). The basic genetic model employed uses the maximum likelihood method to obtain parameter estimates for the effects of additive genetic (A), shared environmental (C), and nonshared environmental (E) influences on a given trait. The additive genetic and shared environmental influences are what make the children within a twin pair similar to each other, while the nonshared – or unique – environmental influences contribute to differences within the pair. The E parameter also includes the effects of measurement error.

The multivariate genetic analyses were intended to be as closely parallel to the phenotypic analyses as possible. Two types of multivariate genetic analyses were conducted: a series of bivariate genetic analyses between each

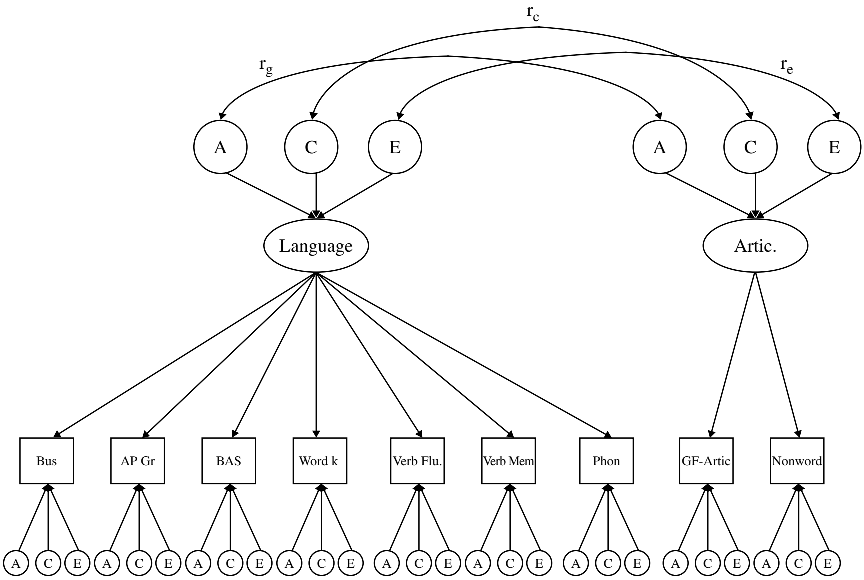

Fig. 1. Path diagram for multivariate common pathways model, with two correlated latent factors (general language and articulation). Square boxes represent the measured variables, while circles represent latent variables. These are the two phenotypic latent factors (above the boxes), with their A, C, E variance components, and the A, C, E variance components that are unique to the measured variables (below the boxes).

pair of tests and a multivariate analysis that examined the aetiological relationships between all tests simultaneously. The bivariate genetic analyses were conducted to estimate genetic correlations between each pair of test, analogous to bivariate phenotypic correlations. These bivariate genetic correlations were derived from a bivariate correlated factors model, in which the overlap in variance between two measures (or factors) is calculated in terms of their A, C, and E components. For the sake of simplicity, we present only the genetic bivariate correlations; information about the analogous environmental correlations reflecting the shared C and E influences on each of the pairs of traits can be obtained from the first author.

For the multivariate genetic analyses that examined the aetiological structure of all of the tests considered simultaneously, we used a common pathways genetic model (Figure 1).

In this model the measured variables are hypothesized to load onto latent factors, as suggested by the phenotypic factor analysis, and the model provides estimates for these loadings. The aetiology of the latent factors is then partitioned into the proportions of their variance explained by additive genetics (A), shared environment (C) and nonshared environment (E). The degree of overlap in the aetiology of the latent factors is also assessed. For example, the product of the three paths connecting the two latent variables is related to bivariate heritability, the extent to which genetic factors account for the correlation between the two latent variables. The genetic correlation, as explained earlier, provides an estimate of the extent to which the same or different genes affect the factors. The model also estimates A, C, and E parameters for influences that are specific to each of the measured variables; the specific E parameters also incorporate measurement error. Finally, there are estimates for the total effects of A, C, and E on each of the measures, which are the sum of the specific effects and the effects that are shared with the other variables loading onto the common latent factor, weighted for that variable's loading on the latent factor.

A note on extremes

The current analyses do not directly address the issue of language impairment, as fitting our multivariate model at the extremes would drastically reduce statistical power. Phenotypically, the results of a factor analysis carried out on the lower end of the distribution were highly similar to the results for the whole sample, although it does not necessarily follow that the genetic interrelationships will also be the same. On the basis of the univariate results presented in previous work (Kovas *et al.*, 2005), which showed similar results at the levels of the whole sample and the low extremes, we would predict that the genetic inter-relationships between measures would also be similar at the extremes. However, we have not as yet been able to test this prediction directly. This would be a worthwhile direction for future work.

RESULTS

Phenotypic analysis

Bivariate correlations. The phenotypic bivariate correlations between the 9 language measures are presented in Table 1. These are all significant at the $p<0{\cdot}001$ level, and they are all at least moderate. The strongest correlation, at 0·68, is that between Nonword Repetition and Goldman-Fristoe articulation, while the weakest, at 0·23, is between verbal memory and Goldman-Fristoe articulation. The highest average correlations with the other measures are 0·46 for Action Pictures Grammar and 0·45 for Nonword Repetition; the lowest average correlation is 0·31 for Phonological Awareness.

Factor structure. As expected from the moderate positive manifold of correlations shown in Table 1, the factor analysis revealed a strong first

TABLE 1. *Bivariate phenotypic correlations for individual language measures*

	Bus story	AP grammar	BAS comp.	Word knowledge	Verbal fluency	Verbal memory	Phon. awareness	GF-artic	Nonword rep.
Bus story	—								
AP gram.	0·59	—							
BAS comp.	0·44	0·44	—						
Word knowledge	0·51	0·49	0·49	—					
Verbal fluency	0·49	0·46	0·43	0·49	—				
Verbal memory	0·50	0·50	0·50	0·47	0·48	—			
Phono. aware.	0·33	0·35	0·34	0·29	0·29	0·34	—		
GF-artic	0·30	0·43	0·41	0·28	0·34	0·23	0·28	—	
Nonword rep.	0·37	0·45	0·44	0·26	0·41	0·35	0·27	0·68	—

TABLE 2. *Loadings for each measure on the unrotated first principal component, and Oblimin rotated factor loadings for each measure, from phenotypic factor analysis*

Test	First principal component loading	Rotated factor loadings	
		General language	Articulation
Bus story information	0·62	0·79	0·31
Action pictures grammar	0·60	0·76	0·47
BAS comprehension	0·47	0·63	0·51
Word knowledge: oral vocabulary	0·62	0·78	0·20
Verbal fluency	0·48	0·68	0·40
Verbal memory: words & sentences	0·58	0·75	0·20
Phonological awareness	0·30	0·54	0·30
Goldman-Fristoe articulation	0·83	0·37	0·91
Nonword repetition	0·80	0·43	0·89

principal component that accounts for nearly half (46%) of the total variance. All of the measures load on the first unrotated principal component (see Table 2). Bus Story and Verbal Memory show high loadings (0·62) and Phonological Awareness shows the lowest loading (0·30). Goldman-Fristoe Articulation and Nonword Repetition show the highest loadings (0·83 and 0·80, respectively), even though their average correlations with the other measures are 0·37 and 0·45, respectively. The high loadings on the first unrotated principal component for these two measures occur because they are so highly correlated with each other ($r = 0·68$) – as noted earlier, although the Nonword Repetition test is meant as a test of phonological short-term memory, it also makes substantial demands on articulation and phonemic segmentation.

A scree plot indicated that two rotated factors could be derived, and an Oblimin oblique rotation yielded two factors, with rotated factor loadings also shown in Table 2. One factor, which we will refer to as Articulation, showed high loadings (0·91 and 0·89, respectively) only for the Goldman-Fristoe Articulation and Nonword Repetition measures, although Action Pictures Grammar and BAS Comprehension also loaded moderately (0·47 and 0·51 respectively). The other seven language measures, including Action Pictures Grammar and BAS Comprehension, loaded more highly on the other factor, which we will refer to as General Language. The correlation between these two factors was 0·42.

Genetic analysis

Bivariate genetic correlations. Bivariate genetic correlation coefficients (r_g) are presented with 95% confidence intervals in Table 3. These were derived

TABLE 3. *Bivariate genetic correlations for individual language measures.* 95% *confidence intervals presented in parentheses*

	Bus story	AP grammar	BAS comp.	Word knowledge	Verbal fluency	Verbal memory	Phon. aware	GF-artic	Nonword rep
Bus story	—								
AP gram.	0·86 (0·52–1·00)	—							
BAS comp.	0·58 (0·27–1·00)	0·63 (0·26–1·00)	—						
Word knowledge	0·66 (0·42–1·00)	0·67 (0·22–1·00)	0·39 (0·00–1·00)	—					
Verbal fluency	0·61 (0·33–0·96)	0·77 (0·28–1·00)	0·52 (0·00–1·00)	0·64 (0·27–0·96)	—				
Verbal memory	0·96 (0·40–1·00)	0·45 (0·00–1·00)	0·54 (0·00–1·00)	1·00 (0·48–1·00)	0·96 (0·38–1·00)	—			
Phon. awareness	0·45 (0·13–0·75)	0·33 (0·00–0·69)	0·34 (0·00–0·88)	0·44 (0·00–0·84)	0·71 (0·41–1·00)	0·37 (0·00–1·00)	—		
GF-artic	0·29 (0·04–0·64)	0·79 (0·38–1·00)	0·63 (0·23–1·00)	0·51 (0·09–1·00)	0·48 (0·01–1·00)	0·34 (0·00–1·00)	0·34 (0·00–0·67)	—	
Nonword rep	0·59 (0·30–0·93)	0·76 (0·46–1·00)	0·52 (0·03–1·00)	0·37 (0·00–0·83)	0·76 (0·42–1·00)	0·35 (0·00–1·00)	0·48 (0·08–0·93)	0·89 (0·69–1·00)	—

from a series of 36 bivariate correlated factors models (each measure with each of the others). The full models provide a good fit to the data, with low χ^2 (range: 3·869–24·717, for 11 degrees of freedom), low (usually negative) Akaike's Information Criterion (−18·131–2·717) and in most cases high probability values (0·010–0·974).

The pattern of genetic correlations (r_g) between the 9 language measures is generally similar to the phenotypic correlations in suggesting a positive manifold among all tests, although the average correlation is higher for r_g (0·58) than for the phenotypic correlations (0·41). As noted earlier, there is no necessary relationship between genetic correlations and phenotypic correlations – the genetic correlation can be very high when the phenotypic correlation is low and vice versa. The strong inter-correlation between GF-Articulation and Nonword Repetition in the phenotypic analysis is also seen at the genetic level, with an r_g of 0·89. The average inter-correlation among the other seven tests included in the General Language factor is 0·62, whereas the average correlation between the two tests included in the Articulation factor and the seven tests included in the General Language factor is 0·52. In other words, the genetic structure among the tests appears to support the phenotypic structure. There are several other very high genetic correlations that do not correspond to particularly strong relationships in the phenotypic analysis. For example, the r_g between the Bus Story and Action Picture grammar measure ($r_g = 0{\cdot}86$) is noteworthy as are the correlations between verbal memory and Bus Story, Word Knowledge, and Verbal Fluency (at r_g of 0·96, 1·00 and 0·96 respectively). The lowest genetic correlation, at 0·29, is between GF-articulation and Bus Story. However, it should be borne in mind that the 95% confidence intervals for these estimates are very wide and one should be cautious in interpreting any differences between the genetic correlation coefficients.

Genetic factor structure. The aetiological links among the nine tests were assessed simultaneously using the common pathways model with two latent factors (Figure 1), reflecting those suggested by the phenotypic factor analysis: an Articulation factor, indexed by the Goldman-Fristoe articulation and Nonword Repetition measures, and a General Language factor, indexed by the remaining seven measures (Bus Story information, Action Pictures grammar, BAS Comprehension, Word Knowledge oral vocabulary, Verbal Fluency, Verbal Memory words and sentences, and Phonological Awareness). This structure provided a good fit to the data: $\chi^2 = 386{\cdot}21$ for d.f. = 298; $p = 0{\cdot}00$; Akaike's information criterion (AIC) = −209·79; and root mean square estimation of fit (RMSEA) = 0·036.

Figures 2, 3, and 4 show the path coefficients (squared) that describe the total genetic and environmental influences on each measured variable, as well as how much of this total is unique to each measure and how much is shared with the other measures (that is, acts through the latent factors). For

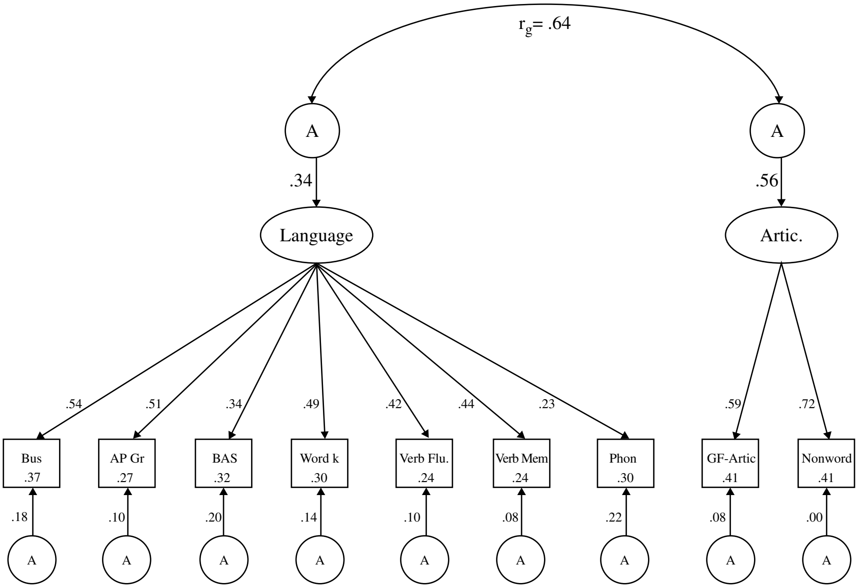

Fig. 2. Estimates for effects of additive genetics (A): Common pathways model with parameter estimates. Note that these are not path coefficients but rather the square of the path coefficients. The number in each square box represents the total heritability for that measure. The path between the box and the latent factor represents the measure's factor weighting, while the path between the latent factor and its variance component represents the additive genetic contribution to the latent factor. The path between the box and the unique variance components represents the unique genetic contribution to that measure.

clarity of presentation, the A, C, and E factors and pathways are presented in different figures, but they are all part of the same model (seen in Figure 1). Specifically, looking at the figures from the bottom up, they show (a) the unique A, C, and E parameters for each measured variable, (b) the total genetic and environmental influence on each measured variable, (c) the factor loadings from each measured variable onto its latent factor, (d) the genetic and environmental influences on the latent factors, and (e) the genetic, shared, and nonshared environment correlations between the two latent factors (r_g, r_c, and r_e).

Tables 4 and 5 present the same information as that illustrated in the model diagrams, with the addition of 95% confidence intervals. Table 4 shows the parameter estimates that are relevant for the individual language measures: total genetic, shared and nonshared environmental influences on each of the measured variables, as well as the unique genetic and environmental contributions to this total, and the loading of each measure onto its latent factor. Table 5 complements this by showing the parameter estimates that are relevant to the latent factors: genetic and environmental influences

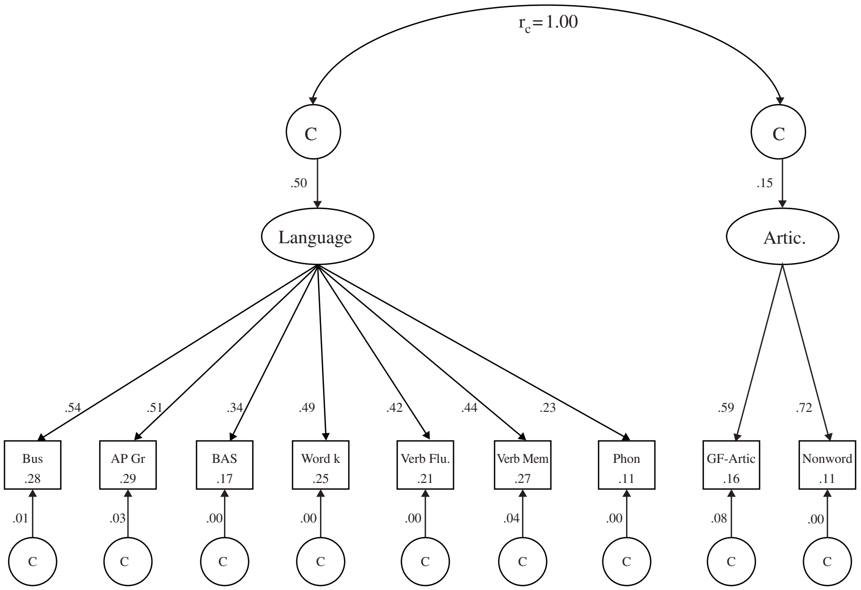

Fig. 3. Estimates for effects of shared environment (C): Common pathways model with parameter estimates.

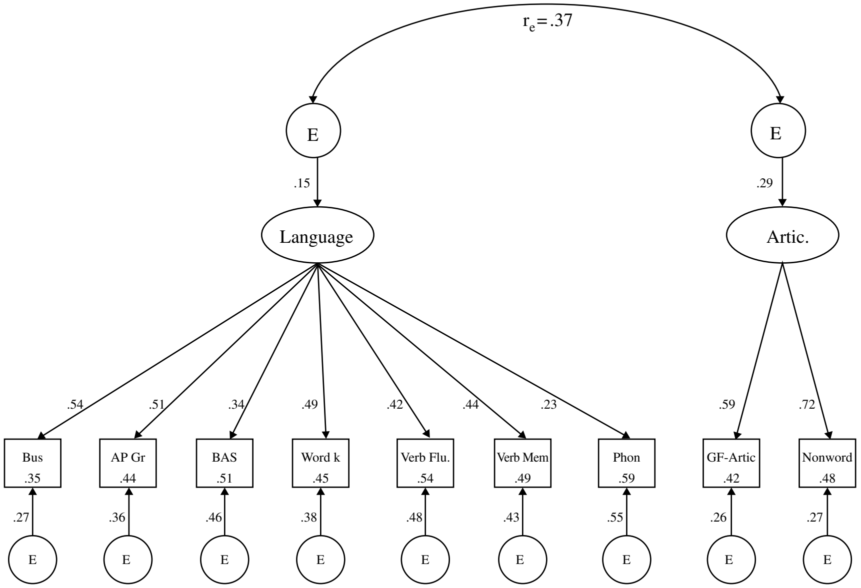

Fig. 4. Estimates for effects of nonshared environment (E): Common pathways model with parameter estimates.

TABLE 4. *Parameter estimates (with 95% confidence intervals) from the full common pathways model, for individual language measures: total influences of additive genetic (a), shared environmental (c), and non-shared environmental (e) influences; specific influences of a, c, and e; and factor loadings for each measured variable onto its latent factor*

Task	Total a^2	Total c^2	Total e^2	Specific A	Specific C	Specific E	Loading on Latent Factor
Bus info.	0·37 (0·17–0·51)	0·28 (0·16–0·44)	0·35 (0·29–0·42)	0·18 (0·01–0·26)	0·01 (0·00–0·15)	0·27 (0·21–0·34)	0·54 (0·48–0·60)
AP gram.	0·27 (0·09–0·43)	0·29 (0·15–0·45)	0·44 (0·36–0·52)	0·10 (0·00–0·20)	0·03 (0·00–0·15)	0·36 (0·29–0·44)	0·51 (0·45–0·57)
BAS gram.	0·32 (0·09–0·42)	0·17 (0·09–0·36)	0·51 (0·44–0·60)	0·20 (0·00–0·28)	0·00 (0·00–0·18)	0·46 (0·38–0·55)	0·34 (0·27–0·40)
Word know.	0·30 (0·12–0·43)	0·25 (0·14–0·39)	0·45 (0·38–0·53)	0·14 (0·00–0·21)	0·00 (0·00–0·12)	0·38 (0·31–0·46)	0·49 (0·43–0·55)
Verbal flu.	0·24 (0·07–0·36)	0·21 (0·12–0·37)	0·54 (0·46–0·64)	0·10 (0·00–0·17)	0·00 (0·00–0·13)	0·48 (0·40–0·58)	0·42 (0·36–0·48)
Verbal mem.	0·24 (0·07–0·40)	0·27 (0·13–0·42)	0·49 (0·42–0·58)	0·09 (0·00–0·20)	0·05 (0·00–0·17)	0·43 (0·35–0·51)	0·44 (0·38–0·50)
Phon. aware.	0·30 (0·11–0·40)	0·11 (0·06–0·25)	0·59 (0·50–0·69)	0·22 (0·04–0·31)	0·00 (0·00–0·13)	0·55 (0·46–0·65)	0·23 (0·17–0·29)
GF artic.	0·41 (0·21–0·59)	0·16 (0·02–0·34)	0·42 (0·35–0·51)	0·08 (0·00–0·22)	0·08 (0·00–0·19)	0·26 (0·19–0·33)	0·59 (0·51–0·66)
Nonword rep.	0·41 (0·23–0·53)	0·11 (0·02–0·26)	0·48 (0·41–0·56)	0·00 (0·00–0·03)	0·00 (0·00–0·02)	0·27 (0·21–0·34)	0·72 (0·66–0·79)

TABLE 5. *A, C, E estimates for latent factors (with 95% confidence intervals), and genetic, shared, and nonshared environment correlations between the latent factors*

Factor	General language	Articulation
	Genetic	
General language	A=0·34 (0·15–0·57)	
Articulation	r_g=0·64 (0·32–0·96)	A=0·56 (0·31–0·72)
	Shared environmental	
General language	C=0·50 (0·30–0·67)	
Articulation	r_c=1·00 (0·58–1·00)	C=0·15 (0·02–0·36)
	Nonshared environmental	
General language	E=0·15 (0·10–0·21)	
Articulation	r_e=0·37 (0·15–0·57)	E=0·29 (0·20–0·38)

on the two latent factors, and the genetic and environmental correlations between them.

The loadings of the measured variables on the latent factors – which are constant for the A, C and E path diagrams – are substantial, and significantly different from zero in all cases (see 95% confidence intervals presented in Table 5). The square roots of these loadings in Figures 2, 3 and 4 correspond roughly to the Oblimin factor loadings seen in Table 2. The additive genetic influence on both latent factors is moderate and significantly greater than zero. Although the point estimate for the Articulation factor is higher (0·56) than that for the General Language factor (0·34), this difference is not significant, as indicated by the overlapping confidence intervals (Table 5). Furthermore, the genetic correlation between the two latent factors is also substantial, at 0·64, indicating largely but not entirely overlapping sets of genes influencing these two areas of language function.

As shown in Table 4, the total amount of additive genetic influence on each of the measures is moderate and significantly greater than zero in that the 95% confidence intervals do not include zero.

The path diagram in Figure 2 divides the total additive genetic influence for each measure into two components – general genetic influence that is shared in common with the other measures loading on the latent factor and specific genetic influence that is unique to the measure. For example, for the Bus Story, the heritability of 0·37 (Table 4) is divided into general genetic influence that is shared with the other general language measures (0·54 × 0·34 = 0·18) and genetic influence that is specific to the Bus Story (0·18). For the general language latent factor, genetic influence on the seven constituent measures is nearly evenly divided between general and specific genetic influence. For the articulation factor, however, nearly all of the genetic influence on GF Articulation and on Nonword Repetition is general; the estimates of specific genetic influence are 0·08 and 0·00, respectively.

In summary, we conclude that genetic influences on these measures overlap substantially within each factor and that there is high genetic overlap between the latent factors of General Language and Articulation.

Turning to the effects of shared environment (Figure 3), the pattern of results is similar to that observed for genetic effects. The total influence of shared environment is moderate and significantly greater than zero for all measures (Table 4). There are significant effects of shared environment on both latent factors, although in this case the influence is greater on the General Language factor than on the Articulation factor – this difference is again non-significant, though there is only a small overlap between confidence intervals (Table 5). Strikingly, there is a perfect shared environment correlation between the two latent factors, implying that the same set of shared environmental influences is acting on both factors, though they have a slightly greater effect on General Language than Articulation. Finally, the unique influences of shared environment on each of the measured variables are negligible, and in no case significantly greater than zero.

The nonshared environmental influences have a rather different pattern of effect. The total effects are significant and substantial for all measures. However, the bulk of this effect is carried by the unique influences on each of the measures: in contrast to the case with A and C, the unique effects of E range from 0·26 to 0·55 and are all significantly greater than zero (Table 4). In contrast, the nonshared environmental influence on the latent factors is significant but modest, at 0·15 for General Language and 0·29 for Articulation. The nonshared environmental correlation between these factors is also lower, at 0·37, than that observed for the genetic and shared environmental correlations (Table 5). It should be noted that the effects of measurement error in this model are incorporated in the estimates of nonshared environment unique to the individual measures.

Nested models

The results reported above refer to the full model, in which all parameters are free for the model to find the set of estimates that yield the best fit between observed and expected covariance data. In order to clarify the results of the full model or to test specific hypotheses, it is possible to constrain certain parameters to a fixed value – or to drop them altogether – and to compare the fit of such nested models to that of the full model.

The parameter estimates produced by the full model suggested low values for almost all the unique A (additive genetic) parameters, and even more so for the unique C (shared environmental) parameters. A nested model which dropped the unique A parameters, however, resulted in a significant decrease in model-fit: χ^2 difference = 19·299, $p = 0·023$. This strengthens the conclusion one could draw from the results of the full model, that there is some, albeit modest, unique genetic variance (particularly for Bus Story information and phonological awareness). Dropping the general A parameters decreased model-fit substantially: χ^2 difference = 25·814, $p = 0·000$, highlighting the importance of these general genetic influences. By contrast, dropping the unique C parameters did not affect the model fit: χ^2 difference = 1·383, $p = 0·998$. As suggested by the full model, it appears that there are no shared environmental influences that are unique to any of the language measures; the marked worsening of model fit when the general C parameters are dropped also supports this conclusion (χ^2 difference = 17·834, $p = 0·000$).

Another question that can be addressed with nested multivariate genetic models concerns the aetiological correlations between the latent factors, and whether it is in fact necessary to have two factors to account for the data. Given the high (0·64) genetic correlation between the language and articulation factors, and perfect shared environment correlation, it is possible that a single set of genetic and environmental influences could account for covariation between the two factors. However, constraining these correlations to equal 1 results in severely worsened model-fit: χ^2 difference = 58·025, $p = 0·000$. Equally, setting the correlations to 0, implying that there are totally separate influences underlying the General Language and Articulation factors, also results in a striking reduction of model fit: χ^2 difference = 303·633, $p = 0·000$. Thus, we can conclude that despite the high genetic and shared environmental correlations between the General Language and Articulation factors, there are significant aetiological differences between the two factors.

DISCUSSION

In examining the aetiological interrelationships between the diverse language measures in our sample of four-year-old twins, both the analyses of bivariate genetic correlations and the multivariate common factor model suggest that many of the same genetic factors underlie the development of individual differences in a wide range of linguistic skills. More specifically, the main findings were as follows:

(1) The data are well-described by a genetic model mirroring the phenotypic factor analysis, in which the articulation and nonword repetition measures loaded on one latent factor while the remaining seven measures loaded on a second factor.

(2) The aetiology of the two latent factors was similar, with significant effects of additive genetics and shared and non-shared environment; there was a trend for stronger genetic and weaker shared environment

effects on the latent articulation factor, and stronger shared environment and weaker genetic effects on the latent general language factor.

(3) In addition, there was substantial though not complete overlap in the aetiology of the two latent factors, with a genetic correlation of 0·64 and a shared environment correlation of 1·00.

(4) Whatever genetic and shared environment effects there were on the language measures, they were substantially mediated by the latent factors: the influences of genetic and shared environment effects specific to individual language measures were non-significant, with only two exceptions (Bus Story and phonological awareness). By contrast, most of the nonshared environmental effects were specific to individual measures.

(5) The finding of substantially shared genetic influence was reinforced by the strong genetic correlations between measures. These complemented the pattern that emerged from the latent factor model, since r_g tended to be strongest for measures within the latent factors derived from the factor analysis. The correlations for the two articulation measures with the remaining language tasks were also high (especially with AP Grammar), reflecting the genetic correlation between the two latent factors in the common pathway model.

Our findings lend support to the view that similar processes – both genetic and shared environmental processes – underlie individual differences in a wide variety of skills within the domain of language in young children. The results of the present genetic analyses are consistent with reports from a large-scale phenotypic study of typical and poor language learners. Similar to our findings, measures of word and sentence level language proficiency, both receptive and expressive, grouped together into one factor; this language proficiency factor was differentiated from a speech production factor in both kindergarten and second grade (Tomblin & Zhang, 1999; Tomblin, Zhang, Weiss, Catts & Ellis Weismer, 2003).

General language factor

The general language factor included measures that assess both receptive and expressive language: however, there is no evidence that the measures are differentiated along this dimension at either the phenotypic or genetic level. Nor is there evidence of differentiation according to linguistic dimensions. For example, the two measures we might consider to be most lexically based, word knowledge and verbal fluency, are no more similar to each other – in terms of either genetic correlations or A, C, E estimates in the multivariate model – than they are to any of the other measures in the general language factor. Perhaps most interestingly in view of the debates surrounding the autonomy and innateness of syntax, there is no evidence that our measure of receptive syntax (BAS Comprehension) or of production of inflectional morphology (Action Pictures Grammar) either clustered together in any way or showed any significant specific genetic effects. It is intriguing, however, that production of inflectional morphology showed high genetic correlations with both GF Articulation and Nonword Repetition. One could speculate that of the expressive tasks, the one requiring accurate production of relatively brief, non-salient morphemes may draw most heavily on articulatory processes.

Within the general language factor, the moderate phenotypic correlations and the moderate factor loadings in the genetic model show that there is more to the individual measures than what is reflected in the latent factor. It is clearly not the case that they are all measuring the same thing. On the other hand, there is a substantial amount of phenotypic covariance, and genetic and shared environmental influences contribute substantially to this overlap. The quantitative genetic findings presented here make a prediction for molecular genetics work that seeks to identify specific genes responsible for the heritability of these language measures. Our results suggest that a gene associated with any of these language measures has at least even odds to be associated with all of them. Genes associated with all of the measures can be used in further research to understand the mechanisms responsible for such general effects. Genes associated with one measure but not the others will help to understand the mechanisms responsible for measure-specific effects.

Articulation factor

The two measures that load on the articulation factor – Goldman-Fristoe articulation and nonword repetition – might not have been expected to pair off together in such a determined way. Nonword repetition is generally regarded as an index of phonological short-term memory. Early findings that children with language impairment were poor at nonword repetition, particularly for longer nonwords, were interpreted directly as indicating a limitation in the phonological loop component of working memory (Gathercole & Baddeley, 1990). In the present study, given that the McCarthy verbal memory subtest – which required repetition of words and sentences – loaded on the general language factor, one might have predicted that nonword repetition would behave similarly if it was indeed measuring phonological working memory. However, it may be that the McCarthy verbal memory subtest groups with the general language factor because of its lexical/syntactic demands rather than its verbal memory demands (consistent with high genetic correlations between verbal memory and verbal fluency, word knowledge and Bus Story Information: 0·96, 1·00, and 0·96 respectively). Conversely, performance on the nonword repetition task depends not only on phonological working memory, but also on accurate phonological analysis and articulatory planning, and may also be influenced by prior lexical knowledge, depending on the word-likeness of the nonwords (Snowling, Chiat & Hulme, 1991). Its grouping in the current analysis with the GF-articulation measure, and away from the general language factor and the McCarthy verbal memory test, suggests that performance on the nonword repetition task in this group of four-year-olds reflects phonological output processes more than phonological working memory.

If the above analysis is correct, some of the previously reported high heritability of nonword repetition (Bishop, North & Donlan, 1996) may be due to articulatory factors. This conclusion suggests that if articulatory factors were controlled for, the heritability of residual phonological working memory would be found to be lower. However, it should be noted that this study did not include children with extremely poor articulation. Furthermore, a recent report found high heritability for impaired nonword repetition after correcting for articulatory skills by residualising scores on multisyllabic nonwords against scores on mono- and bisyllabic nonwords (Bishop, Adams & Norbury, in press).

In the current study, several methodological factors may have contributed to the close association between articulation and nonword repetition tasks. One is age: the children in our study were between 4 and 5 years of age, which is younger than most studies looking at nonword repetition. It is possible that the articulatory demands are particularly important for this age group. We also considered whether our study design confounded poor articulation with nonword repetition by failing to exclude children with articulation problems. However, a re-analysis that excluded all children whose GF-articulation score fell more than one standard deviation below the sample's mean did not alter the pattern of results. A third possibility relates to the shortened version of the nonword repetition task that we used, which included only two and three syllable nonwords: it may be that articulatory demands are paramount for short nonwords, and that phonological memory plays a more important role for longer nonwords.

Relationship between the general language and articulation factors

As the genetic correlation of 0·64 and the shared environmental correlation of 1·00 suggest, the two latent factors capture domains of ability that have a great deal of overlap in their aetiology, but also a substantial amount of independence (also reflected in the low non-shared environmental correlation). The genetic correlation of 0·64 can be interpreted to mean that if a gene were identified that is associated with the general language factor, there is about a 60 percent chance that the same gene would also be associated with the articulation factor. The shared environmental correlation of 1·00 means that any shared environmental factor found to be associated with general language would also be associated with articulation. At the phenotypic level of explanation, the division between the factors could be described as following the intuitive boundary between speech and language.

Speech encompasses purely articulatory processes as well as more general phonological processes, and these are probably both reflected in our 'Articulation' factor. Interestingly, the (receptive) phonological awareness task we used loaded on the general language and not the articulation factor. However, compared to the other 6 measures in the general language factor, it had the lowest loading on this factor and the highest loading on the articulation factor (Table 2). It also had the weakest average genetic correlation with the other 'general language' measures (Table 3).

The finding that the articulation and general language factors can be differentiated aetiologically is not predicted by either the procedural/declarative model (Ullman, 2001), or the phonological loop account (Baddeley *et al.*, 1998) discussed in the introduction. According to the procedural/declarative model, phonology is similar to syntax in being rule-governed and underpinned by a procedural implicit learning system: this model would presumably predict that syntax and phonology should also be linked at the aetiological level, and be separable from semantics. Instead, we found that semantics and syntax were very closely linked, while phonology could be at least partially separated from both of these.

Caveats and conclusions

A general caveat that in interpreting the results of any twin study, is the possibility of twin-specific effects that may limit the generalisability of the findings to singletons. Previous research has consistently found a language immaturity in twins at about 3–0;6 (e.g. Mittler, 1969; Rutter, Thorpe, Greenwood, Northstone & Golding, 2003), and a similar delay in early language (at 2 years) has also been found in the TEDS sample (Dale, Simonoff, Bishop, Eley, Oliver, Price, Purcell, Stevenson & Plomin, 1998). However, there appeared to be no evidence of delay in the four-year old twins in the present study (specifically, in a subgroup of 620 'controls') as compared to standardization data of the McCarthy Scales (Colledge *et al.*, 2002). 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 between identical and fraternal twins (Mittler, 1969). A strong test of the generalizability of twin findings to singletons is to include siblings of the twin pairs, using the same measures at the same age that the twins were tested. Unfortunately we do not have such data for the current sample.

Another point to note is that the results obtained here are specific to the measures that we used. In a study of this size, with children of this age, there are limits on how large a language battery can be given, and our decision was to sample briefly a range of language domains rather than assessing one or two domains in depth. However, it is possible that had we used different receptive language measures, or more measures of morphosyntax, we would have found evidence of specificity for these domains.

Our findings are necessarily limited not only by the measures, but also the sample that we used. For example, the mean age of our sample was 4;6, by which point much of 'early language acquisition' has already happened for typically developing children. Another point to note about our sample is that although it is large in relation to many studies on language development, the power is limited, and the standard errors around estimates of genetic and environmental effects, and especially genetic correlations, are substantial. When we report that most language tasks did not yield significant specific genetic influences, we are not implying that such influences are absent, but rather that we cannot say with any confidence that they were greater than zero.

Bearing in mind the foregoing provisos, our current results support a view of early language that treats the origin – and by implication the function – of different linguistic components as substantially integrated. We found that the moderate heritabilities of each of the diverse language measures (Kovas *et al.*, 2005) were, to a large extent, due to the genetic factors they have in common. The greatest evidence of aetiological dissociation mirrored the intuitive distinction between speech and language, and separated a factor of what we interpreted as articulation, from the rest of language.

REFERENCES

Baddeley, A., Gathercole, S. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review* **105**, 158–73.

Bates, E., Dale, P. S. & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (eds), *Handbook of Child Language*. Oxford: Basil Blackwell.

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**, 446–62.

Bishop, D. V. M., Adams, C. & Norbury, C. (in press). Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. *Genes, Brain and Behavior.*

Bishop, D. V. M., North, T. & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *Journal of Child Psychology & Psychiatry & Allied Disciplines* **37**, 391–403.

Botting, N. & Conti-Ramsden, G. (2001). Nonword repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders* **4**, 421–32.

Chiat, S. (2001). Mapping theories of developmental language impairment: premises, predictions and evidence. *Language & Cognitive Processes* **16**, 113–42.

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

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**, 324–28.

Dale, P. S., Dionne, G., Eley, T. C. & Plomin, R. (2000). Lexical and grammatical development: a behavioural genetic perspective. *Journal of Child Language* **27**, 619–42.

Dale, P. S., Price, T., Bishop, D. V. M. & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language delays at 3 and 4 years. *Journal of Speech, Language, and Hearing Research* **46**, 544–60.

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

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

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59**.

Gathercole, S. E. & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *Journal of Memory & Language* **29**, 336–60.

Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. (1994). The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory* **2**, 103–27.

Gleitman, L. R. (1990). The structural sources of verb meaning. *Language Acquisition* **1**, 3–55.

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

Goodman, J. C. (1995). The shape of change: longitudinal evidence about language development. *Society for Research in Child Development Abstracts* **10**, 111.

Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V. M. & Plomin, R. (2005). Genetic influences in different aspects of language development: the etiology of language skills in 4·5-year-old twins. *Child Development* **76**, 632–51.

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

McGue, M. & Bouchard, T. J., Jr. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics* **14**, 325–43.

Mittler, P. (1969). Genetic aspects of psycholinguistic abilities. *Journal of Child Psychology and Psychiatry* **10**, 165–76.

Morgan, J. L. & Demuth, K. (1996). *Signal to syntax: bootstrapping from speech to grammar in early acquisition.* Mahwah, NJ: Erlbaum.

Neale, M. C., Boker, S. M., Xie, G. & Maes, H. H. (2002). *Mx: Statistical Modeling* (6th ed.). VCU Box 900126, Richmond, VA 23298: Department of Psychiatry.

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

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

Plunkett, K. & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition* **48**, 21–69.

Renfrew, C. E. (1997*a*). *Bus Story Test* (4th ed.). Bicester, Oxon, England: Winslow Press.

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

Rice, M. L. (2000). Grammatical symptoms of specific language impairment. In D. V. M. Bishop & L. B. Leonard (eds), *Speech and language impairments in children: causes, characteristics, intervention and outcome*. Hove, UK: Psychology Press.

Rutter, M., Thorpe, K., Greenwood, R., Northstone, K. & Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay: I: Design; twin-singleton differences in language, and obstetric risks. *Journal of Child Psychology and Psychiatry* **44**, 326–41.

Snowling, M. J., Chiat, S. & Hulme, C. (1991). Words, nonwords, and phonological processes: some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics* **12**, 369–73.

Thal, D. J., Bates, E., Goodman, J. & Jahn-Samilo, J. (1997). Continuity of language abilities: an exploratory study of late- and early-talking toddlers. *Developmental Neuropsychology* **13**, 239–73.

Tomasello, M. (1992). *First verbs: a case study of early grammatical development*. New York: Cambridge University Press.

Tomblin, J. B. & Zhang, X. (1999). Language patterns and etiology in children with specific language impairment. In H. Tager-Flusberg (ed.), *Neurodevelopmental disorders*. Cambridge, MA: The MIT Press.

Tomblin, J. B., Zhang, X., Weiss, A., Catts, H. & Ellis Weismer, S. (2003). Dimensions of individual differences in communication skills among primary grade children. In M. L. Rice & S. F. Warren (eds), *Developmental language disorders: From phenotypes to etiologies*. Mahwah, NJ: Erlbaum.

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**, 444–8.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* **30**, 37–69.