Published in final edited form as:

J Speech Lang Hear Res. 2010 August ; 53(4): 982–992. doi:10.1044/1092-4388(2009/09-0108).

The etiology of diverse receptive language skills at 12 years

Philip S. Dale[1], **Nicole Harlaar**[2], **Marianna E. Hayiou-Thomas**[3], and **Robert Plomin**[4]

[1]University of New Mexico, Albuquerque

[2]Ohio State University, Columbus

[3]University of York, UK

[4]King's College London

Abstract

Purpose—In the second decade of life, language skills expand in both quantitative and qualitative ways. The etiology of these new skills and the relationships among them have been little explored.

Method—Taking advantage of widespread access to inexpensive and fast Internet connections in the United Kingdom, we administered four web-based measures of receptive language development – Vocabulary, Listening Grammar, Figurative Language, and Making Inferences – to a sample of 4892 12-year-old twin pairs participating in the Twins' Early Development Study (TEDS).

Results—The four measures showed moderate phenotypic intercorrelation. All four showed moderate genetic influence (a^2 between .25 and .36), and low shared environmental influence (c2 between .13 and .19). The median genetic correlation among the four measures was 0.87, indicating strong genetic overlap among them. A latent factor score for language, based on the common variance among the measures, showed substantial genetic influence ($a^2 = .59$) and moderate shared environmental influence ($c^2 = .28$). A small but significant sex difference favored females on the Listening Grammar and Making Inferences tests, but there was no evidence for sex differences in the etiology of any of the measures.

Conclusion—Despite the emergence of new skills at this developmental period, from the etiological perspective language skills remain relatively undifferentiated at an etiological level..

In the second decade of life, language skills expand in both quantitative and qualitative ways. As Nippold (2007) notes in her review of later language development, although many of these changes are more subtle than the dramatic changes of the first five or six years, they have equally far-ranging consequences for the child's cognitive, academic, and social life. Between the ages of 10 and 15, for example, Nippold identifies major changes in the lexicon (e.g., abstract nouns, double-function terms, prefixes and suffixes), verbal reasoning (e.g., more complex and abstract analogies), figurative language (e.g., slang, sarcasm, metaphor), syntax (e.g., more complex subordination, longer clause length), pragmatics (e.g., more complex and complete narratives, more coherent persuasive arguments), and literacy (e.g., comprehension of expository texts, distinguishing between fact and opinion).

Many of these changes appear to reflect some broad, general developments characteristic of adolescent thought. Nippold (2007) suggests four such developments that may underlie many of the specific advances in this period. The first is a growth in metacognitive competence; in the case of language, the ability to reflect upon and analyze language as an object itself. Although growth in this ability begins much earlier, e.g., phonological awareness in the preschool years, the skill is stronger and broader in the school years, and permits the analysis and understanding of new words, grammatical structures, and pragmatic acts, as well as serving as one foundation for understanding figurative language. A second change, also an aspect of cognitive development, is the ability to think abstractly. In language, this is manifested in the acquisition of abstract words and sentence meanings; and together with metalinguistic skills, makes possible the appreciation of ambiguity and humor. A third major change concerns social cognition, specifically, a growth in social perspective-taking. This is the foundation of nearly all the changes in pragmatics, which concern people's intentions when they use language, and their choices with respect to expressing those intentions. Finally, language development in the second decade is largely driven by print input, rather than oral input. Around the fourth grade, the oft-noted transition between "learning to read" and "reading to learn" occurs (Snow, Burns & Griffth, 1998). This transition applies not only to content areas such as history and science, but to language itself, as print introduces children to advanced vocabulary and syntax and figurative expressions.

The etiology of these new skills in adolescence and the relationships among them have been little explored, other than vocabulary which is often included in studies of verbal intelligence (Plomin, DeFries, McClearn, & McGuffin, 2008; Plomin & Kovas, 2005). In the present study, we use twin methodology to examine directly the diversity and interrelations of language advances in a large sample of 12-year-old twins. The age of 12 is well into the transition to adolescence, which is characterized by brain and hormonal changes that very likely play a role in the broad changes listed in the previous paragraph (Ernst & Mueller, 2008; Spear, 2000). Although genes themselves do not change with development, changes in the pattern of gene expression play an important role, and indeed the relative importance of genetic variants may change with development and with learning domain. These shifts have the potential to fundamentally affect the genetic architecture of abilities; that is, the way in which diverse abilities are related etiologically (Davis, Haworth & Plomin, in press).

Research on earlier stages of language development has utilized multivariate genetic analysis of twin data to demonstrate a surprising coherence across diverse abilities. For example, vocabulary and grammar are generally interpreted as distinct aspects of language on both linguistic and psychological grounds. However, the two are highly correlated in early development (Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007). Dionne, Dale, Boivin and Plomin (2003) sought the roots of this connection by examining measures of vocabulary and grammar at ages 2 and 3 years in a large sample of twins. They found evidence for very substantial overlap in the genetic influences on vocabulary and on grammar, and also for shared (within family) environmental influences on the two. They interpreted these results as strongly suggesting a high degree of overlap in the mechanisms underlying learning and using these two aspects of language.

Similarly, Hayiou-Thomas, Kovas, Harlaar, Bishop, Dale and Plomin (2006) examined the relationships among a diverse set of measures of articulation, phonology, grammar, vocabulary, and verbal memory at age 4½ for a subset of the twins in the Dionne et al. study. They found that 7 of the language measures formed a coherent latent factor on both phenotypic and genetic grounds, while 2 other measures constituted an articulation latent factor. The two factors themselves had substantial etiological overlap.

DeThorne, Petrill, Hart, Channell, Campbell, Deater-Deckard, Thompson and Vanderbergh (2008) included conversation-based measures such as mean length of utterance and number of different words along with test-based measures in their study of genetic influences on children's language at age 7. Although separate latent factors could be defined for the conversational and test-based measures, there was significant genetic correlation between the two.

These findings are consistent with an emerging body of evidence for the 'Generalist Genes' hypothesis (Plomin & Kovas, 2005; Kovas, Haworth, Dale & Plomin, 2007; Davis, Haworth & Plomin, 2009). Most genetic effects are general, in that their effects are found across a range of cognitive and academic domains. This is most clearly shown in the magnitude of genetic correlations across measures, which can be conceptualized as the probability that a gene which is found to influence one domain will be found to influence the other. Genetic correlations among a diverse range of abilities such as reading, language, mathematics, vocabulary, spatial and memory abilities are typically between 0.5 and 1.0 (Plomin & Kovas, 2005). However, except for the two studies described above, this issue has not been addressed within the domain of language. The conclusion of those two studies – high coherence across diverse language skills – cannot be assumed to hold in the second decade, given the substantial qualitative and diverse changes in language development that occur then.

Our broad goal was to examine the etiology of these new aspects of language at age 12. More specifically, we sought first to examine the etiology of four measures of language development individually, and second, to evaluate the overlap of genetic (and also environmental) influences on this set of measures. Because multivariate analyses require large sample sizes in order to have any precision in parameter estimates, we took advantage of widespread access to inexpensive and fast broadband connections in the United Kingdom to administer the measures in web-based form.

The large and representative sample made it possible to address the question of sex differences as well. Much less evidence is available concerning sex differences in adolescence than in early childhood or adulthood (the topic is not even mentioned in Nippold, 2007, and many other current reviews). Small, but significant sex differences, typically 1-2% of the variance, are frequently found on measures of early language development (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994; Halpern, Benbow, Geary, Gur, Hyde & Gernsbacher, 2007). However, the limited findings available to date are more mixed in adolescence (Hedges & Nowell, 1995). Higher means for females than males have been found for measures of language usage, reading comprehension, and especially writing. Higher means for males are sometimes found for tests of verbal analogies. Sex differences have also been found in measures of variability, rather than mean level, of cognitive and academic abilities, usually reflecting greater variability for boys, though this difference is more strongly supported for mathematics than for verbal abilities. In addition to asking if there are differences in mean or variance, we also sought to evaluate the possibility of sex differences in etiology, such as a difference in the relative influence of genetic and environmental factors on language abilities. For example, Spinath, Price, Dale and Plomin (2004) found higher heritability for language ability in boys than girls at the preschool stage. However, these are exploratory analyses, and no specific hypotheses were formulated.

Method

Participants

Twins in the present sample were 12-year-old participants in the Twins Early Development Study (TEDS), a longitudinal study of twins ascertained from population records of live twin births in England and Wales (Kovas et al., 2007). Since then, the sample has remained reasonably representative of the UK population, as determined by comparison with Office of National Statistics census data (Oliver & Plomin, 2007). Informed consent is obtained by post or online consent forms, and a test administrator is assigned who telephones the family and provides assistance and encouragement. We excluded from the present analyses data from twin pairs with any of the standard set of genetic, medical, and data quality exclusion criteria in TEDS (see Kovas et al. for details) or severe current medical problems. Zygosity was determined by a combination of parent questionnaires at 18 months, 3 years, and 4 years, and DNA tests for a subsample when the twins' zygosity was doubtful or the parents requested it (Kovas et al., 2007). We included only twins whose first language was English, to eliminate variability due to differences in exposure to English. Finally, we included only twins whose parents reported their ethnicity as “white,” which is 93% of this UK sample. This exclusion was selected to maximize genetic similarity in preparation for later molecular genetic analysis of the sample. The final sample included 4892 pairs: 1758 MZ, 1580 DZ same-sex, and 1554 DZ opposite-sex pairs.

Measures

Reliance on internet-based testing, necessary for assessment of a large sample, led to our focus on receptive measures. Standardized tests were selected that had demonstrated ability to discriminate children with language disability as well as being sensitive to individual differences across the full range of ability. Within those constraints, we attempted to select a diverse set of tests, sampling vocabulary and semantics, syntax, and pragmatics. An aspect of language that becomes increasingly important in adolescence, and shows substantial variability at this age, is metalinguistic ability, that is, knowledge about language itself (Nippold, 2007). For this reason, the four measures selected included two with low metalinguistic demands (Vocabulary, Listening Grammar) and two with higher demands (Figurative Language, Making Inferences).

Internet test development and administration—The language measures reported in this paper were part of a larger internet-based test battery covering multiple cognitive and academic domains (Davis et al., in press). The battery was developed in collaboration with

Planet Three Publishing (www.planet3.co.uk) and e-Business Systems (www.e-businesssystems.co.uk). The standardized tests were modified with a set of adaptive branching rules, so that all children started with the same items, but then were branched to easier or harder items depending on their performance. Adapting to children's competence increases their engagement, while limiting the number of items that need to be answered (Birnbaum, 2004). Points were awarded for correct responses, for unadministered items preceding the child's starting point, and for items skipped through branching to harder items. Streaming audio was used where appropriate for tests. The test battery was self-paced, and could be completed over a period of several weeks. Each child's performance was monitored online, and families were telephoned at the start of the testing and provided support and encouragement throughout the process. Games were interspersed with the activities to maintain engagement. For further information on the testing procedure, including steps to assure confidentiality, independence of data from the twins in each family, and internet access and connection speed, see Haworth et al. (2007). Cronbach's alpha, a measure of internal consistency for the web-based test, is reported for each measure (from Haworth et al., 2007).

Vocabulary—The WISC-III-PI Vocabulary Multiple Choice subtest (Wechsler, 1992) is a multiple-choice test, in which children select from three or four alternative definitions of a word. Note that the adaptation to multiple choice is part of the development of the UK adaptation of the WISC-III (Wechsler, 1992). (Alpha = .88)

Nonliteral Semantics—In additional to vocabulary, semantics was assessed using Level 2 of the Figurative Language subtest of the Test of Language Competence – Expanded Edition (Wiig et al., 1989). This subtest assesses the interpretation of idioms and metaphors; correct understanding of such nonliteral language requires rich semantic representation as well as an awareness of the ambiguity of many expressions between their literal and figurative meaning. The child hears a sentence orally and chooses one of four answers, presented in both written and oral forms. (Alpha = .66)

Syntax—Syntax was assessed using the Listening Grammar subtest of the Test of Adolescent and Adult Language (TOAL-3; Hammill et al., 1994). This test requires the child to select two sentences that have nearly the same meaning from a set of three options. The sentences are presented auditorily only. (Alpha = .94)

Pragmatics—Level 2 of the Making Inferences subtest of the Test of Language Competence – Expanded Edition (Wiig et al., 1989) requires participants to make permissible inferences on the basis of existing, but incomplete, causal relationships presented in short paragraphs presented orally. The child chooses two of four responses, presented in both written and oral form, that best explain what could have happened. (Alpha = .58)

Analysis—Data analysis proceeded through four phases. The first was a phenotypic description of performance on each of the four measures and their intercorrelations. The second was comprised of univariate genetic analyses of each measure, estimating the proportion of variance due to additive genetic (A), shared environment (C), and nonshared environmental (E) influences. In this phase, we also examined the possibility of sex differences in the etiology of each measure.. In the third phase, the etiology of the relationship between each pair of variables was estimated with genetic and environmental correlations. The genetic correlation is an index of the degree of overlap between the genetic influences on each variable in the pair; and similarly for the shared environmental and nonshared environmental correlations. In the fourth phase, based on the close connection among the measures demonstrated phenotypically and genetically, we fitted a common pathway model to the data. This model derives a latent factor for language using maximum likelihood factor analysis, and then partitions the variance in that factor – variance common to all four measures - into additive genetic, shared environment, and nonshared environmental components. The same latent factor model also estimates genetic and environmental components of the variance specific to each measure. The genetic analyses of the second, third and fourth phases were conducted in the Mx statistical program (Neale, Boker, Xie & Maes, 2002). We estimated 95% confidence intervals to indicate the significance of parameter estimates. For all later analyses, standardized residuals correcting for age and sex were used, because the age of twins is perfectly correlated across pairs, and sex is perfectly correlated for MZ pairs. Unless corrected, these factors would inflate the correlation between twins. Note that the sample sizes are slightly different because the phenotypic analyses were conducted in SPSS, which utilizes listwise deletion, whereas the genetic analyses were conducted in Mx, which estimates parameters using Full Information Maximum Likelihood techniques for missing data.

Results

Phenotypic analyses

Table 1 presents the means and standard deviations for each measure, subdivided by sex and zygosity. A small sex difference was observed for Listening Grammar and Making Inferences, but it accounted for only .2% of the variance in both cases. There was no significant sex differences in variance in performance. Overall, MZ and DZ twins performed at similar levels, although there was a trend for DZ pairs to score slightly higher on average; this difference was significant for Vocabulary and Listening Grammar, but again accounted for much less than 1% of the variance. Table 2 presents the correlations among the four measures, which are all at least moderate in magnitude. The Vocabulary and Listening Grammar scores showed bimodality; consequently these and later analyses were also conducted with those scores recoded in quintiles to produce a more uniform distribution. The results were virtually indistinguishable, and are not reported here (details available on request from the authors).

Univariate genetic analyses

Intraclass correlations are presented in Table 3 for the MZ and DZ twins at each age. The correlations are consistently higher for MZ than DZ twins, with a lack of overlap between the confidence intervals, suggesting a significant genetic contribution to each measure. An estimate of the genetic influence can be obtained by doubling the MZ-DZ correlation difference; this yields heritability estimates of 32%, 34%, 24%, and 34% for the four measures (Vocabulary, Listening Grammar, Making Inferences, and Figurative Language, respectively). More accurate estimates of heritability and the other etiological parameters, shared environmental influence and nonshared environmental influence are provided by standard univariate model-fitting, utilizing the Mx statistical program (Neal, Boker, Xie & Maes, 2006). Parameter estimates derived from this model-fitting are also included in Table 3, as well as 95% confidence intervals. The overlapping confidence intervals seen in each of the last three columns of the table suggest that the four measures do not differ significantly in the magnitude of genetic and environmental influences.

Although Table 1 showed only very small, albeit significant, sex differences on two of the measures and none on the other two with respect to mean levels of performance, it is possible that there are etiological differences between the sexes on one or more of the measures. To evaluate this possibility, we computed a set of univariate sex-limitation models. The basic approach is to constrain various parameters (such as A, C, and E) to be the same for the two sexes for each measure, and determine if the fit of the model is significantly worsened. As the results of this analysis in the Appendix show, in no case was the fit worsened significantly, even when all parameters were constrained to be equal. This result demonstrates that there are no quantitative differences in etiology for the sexes, that is, differences in the balance of genetic or environmental factors. It also implies that there are no qualitative sex differences, that is, evidence that different genetic factors or different environmental factors influence scores for the two sexes. Consequently, all further analyses combined the data from both sexes.

Bivariate (pairwise) analyses

We next performed six bivariate genetic analyses, one for each pair of variables. Each of these analyses estimates three etiological correlations, all included in Table 4. The genetic correlations are reported above the diagonal. This statistic is a measure of the degree of overlap of genetic influences on both variables, that is, the extent to which it is the same genes which influence the pair. (It should be noted that genetic correlations are independent of the heritability, the actual influence of genes on each measure, so that genetic correlations can be high even when heritabilities are modest.) The figures below the diagonal are the shared environmental and nonshared environmental correlations, defined analogously. The genetic correlations (r_A) are all high, exceeding .71 and with a median value of .87. The shared environmental correlations (r_C) are almost equally high, exceeding .61, with a median value of .78. The nonshared environmental correlations (r_E) are much lower, with a range of .08 to .13. Thus it appears that the same genetic and shared environmental factors appear to influence all four measures, but that nonshared environmental factors differ among the measures.

Common pathway model

The moderate phenotypic correlations (Table 2) and high genetic correlations (Table 4) among the four measures suggest that an analysis of shared variance among the measures would be appropriate. In the common pathways model, a latent factor for language is estimated which eliminates measure-specific variance, including errors of measurement, to the extent that errors are not correlated across all measures. Measure-specific reliable variance is also eliminated at this step. The shared variance in that latent factor is then partitioned into additive genetic, shared environment, and nonshared environmental components. The model also estimates the influences on the remaining, measure-specific variances.

The results of this model-fitting are displayed in Figure 1. All four measures load strongly on the latent factor. The heritability of the latent language factor ($a^2 = .59$) is higher than for the individual factors (in Table 3), accounting for more than half of the variance, and shared environment influence ($c^2 = .28$) is also increased. Nonshared environment influence is modest ($e^2 = .13$).

Figure 1 also includes the etiology of measure-specific influence. Only for Listening Grammar is there significant evidence for measure-specific genetic influence ($a^2 = .32$), and only for Making Inferences is there significant evidence for measure-specific shared environmental influence ($c^2 = .26$). The majority of the measure-specific variance for each of the measures is non-shared environment, which includes both twin-specific experience and error variance.

Table 5 summarizes information from this model-fitting in a different way. The path diagram in the figure divides the total additive genetic influence for each measure into two components: general genetic influence shared in common with the other measures, and specific genetic influence that is unique to the measure. For example, the genetic influence on Figurative language (.36, in Table 3) is the sum of genetic influence shared with other measures ($.59 \times .71^2 = .29$) and the measure-specific genetic influence ($.26^2 = .07$) The shared environmental influence and nonshared environmental influence can also be divided in this way. In Table 5, the variance for each of the four measures is thus divided among six sources: A, C, and E influences on the latent factor which indirectly influence the measure, and A, C, and E influences which are measure-specific. (The figures in Table 5 do not sum exactly to the univariate parameters listed in Table 3 due to the use of multivariate model-fitting, which is more accurate, but the differences are very small.)

Discussion

The four aspects of language tested at age 12 are qualitatively different from measures of language development in the preschool years, as well as incorporating considerable diversity in themselves. Furthermore, they appear to have strong and complex relationships with other aspects of development, including cognitive, social, and academic aspects. Our first goal was to examine the etiology of the measures individually. All four were moderately heritable (.25 - .36), comparable to the results for the individual measures assessed at 4½ (Kovas, Hayiou-Thomas, Oliver, Dale, Bishop and Plomin, 2005), although the influence of shared environment was lower at the older age. A second goal was to examine the relationship among the measures. The four were moderately related phenotypically, as shown by the similar factor loadings for the four measures in Figure 1 as well as the correlations in Table 2. They are even more strongly related etiologically; as the correlations in Table 4 illustrate, it is largely the same genes and the same shared environmental factors that influence all four measures. Again, this outcome is analogous to the pattern observed at 4 ½ years (Hayiou-Thomas et al., 2006). But it is even more notable at age 12 because of increased diversity among the measures. In particular, there is no evidence for dissociation between measures with low metalinguistic requirements (Vocabulary, Listening Grammar) and those with high requirements (Making Inferences, Figurative Language).

The finding of a high degree of overlap in the genetic factors that influence these measures as well as overlap among the shared environmental factors that play a role suggests common mechanisms for learning aspects of language as diverse as vocabulary, grammar, inference, and figurative language. Much research has confirmed that even for the learning of vocabulary, a rich repertoire of social, statistical, and cognitive skills are required, a repertoire which would be equally useful for other aspects of language (Golinkoff, Hirsh-Pasek, Bloom, Smith, Woodward, Akhtar, Tomasello & Hollich, 2000; Bavin, 2009) . In this view, genetic and environmental influences are seen as affecting development in all domains simultaneously. Other interpretations of these empirical findings on etiology are possible, however. Language development may be strongly cumulative in nature; if so, anything which influences an earlier stage will necessarily affect later stages. Vocabulary forms an important early foundation, such that effective mastery of vocabulary may be essential (though not necessarily sufficient) for grammar learning (Marchman & Thal, 2005). Similarly, achieving a certain level of competence in vocabulary and grammar may be essential before inference and figurative comprehension are possible. This situation may be analogous to a well-established example from the literacy domain: the relationship of word recognition, especially fluency of recognition, to reading comprehension. Because comprehension is so strongly dependent on fluency of word recognition, factors which are related to the former will be related to the latter (Olson & Byrne, 2005; Catts, Hogan & Adlof, 2005). Several aspects of Figure 1, particularly the measure-specific effects, are consistent with this cumulative model. There is neither genetic nor shared environment measure-specific influence on Vocabulary. For Listening Grammar, there is a measure-specific genetic influence beyond that of the latent factor, and for Making Inferences, there is a measure-specific shared environment influence. These two measures are, on a priori analysis, the most distinct from Vocabulary learning. Deciding between these and other models will require other kinds of evidence, including longitudinal and intervention research.

The close relationship among the measures led to the development of a common pathways model, which examined the etiology of the shared variance among the measures. This analysis led to the major difference between the present results and those of Hayiou-Thomas et al. (2006) for 4 ½ year olds. In the present study, the heritability estimated for the latent factor, .59, is substantially higher than the .34 estimated for the language factor at the earlier age. Increasing heritability of traits with development is a frequent finding (Haworth et al., in press). As Haworth et al. (in press) note, longitudinal genetic research (e.g., Kovas et al., 2007) most often finds evidence that genes contribute to continuity rather than change. That finding runs counter to a prediction based on more genes becoming functional in the brain during the transition to adolescence. A more plausible mechanism that might underlie this increase in heritability for language is genotype-environment correlation (Plomin, DeFries, McClearn & McGuffin, 2008). If children influence, or select, their environments on the basis of their genetic predispositions – whether, for example, they prefer talking, bookreading, and dramatic play to puzzles, blocks, and construction play – and the practice that occurs in those environments facilitates further development, genetically-driven variation among children will increase with development. In fact, Oliver, Dale & Plomin (2005) found a modest, but significant heritable component to the amount of early literacy experience, such as bookreading, which children experienced at age 4.

Our third goal was to evaluate the possibility of sex differences in language skills at age 12, including potential differences in mean level, variability, and etiology. Only for two of the measures was there a significant sex difference with respect to mean level, in both cases favoring girls, but they were of very small magnitude (less than 1%). No differences were found for variability or etiology. These results again are similar to those for 4 ½ year olds. Utilizing the same dataset as Hayiou-Thomas et al. (2006), Kovas et al. (2005) found small but significant sex differences favoring girls for five of the seven language measures, ranging from .2% to .7% of the variance. Thus, despite some small sex differences in mean performance, there is no evidence at either 4½ or at 12 years for differences in the underlying etiology for boys and girls.

Limitations and Conclusions

The conclusions of this study must be qualified by several limitations in the design. First, reliance on internet-based testing required limiting our assessment to receptive language measures. Whether expressive language measures fit into the same pattern can only be decided on the basis of further research. More generally, inclusion of a wider battery of measures, such as speech (Hayiou-Thomas et al., 2006), derivational morphology, verbal analogies, irony, or literate lexicon, might produce different results. The second is evidence for limited reliability, and hence validity, of some of the individual measures. Correlations between MZ twins are often used as ‘lower-bound’ estimates of reliability, and they are only moderate here (.43 - .53). Consistent with these figures, the internal consistency of Figurative Language (.66) and Making Inferences (.58) was quite modest. (The figures for this web-administered, adaptive testing format were only slightly below the internal consistency figures for the test as originally developed, however: .71 and .66, respectively.) Moreover, Haworth et al. (2007) did not directly test the concurrent validity of these web-based measures for language. However, the effect of limited reliability is to decrease estimates of heritability and shared environmental influence for those measures, and therefore the figures in Table 3 may be taken as conservative. Limited reliability was a major motivation for use of a latent factor analysis, which abstracts away from error variance, assigning it to measure-specific nonshared environment. The latent factor which remains is inherently more reliable.

Despite these qualifications, however, the present results do support a characterization of language skills at age 12 as substantially integrated, consistent with the Generalist Genes Hypothesis. In this paper, we have provided evidence for general effects within the domain of language. Similar findings emerge from the analyses of Davis et al. (in press), who examined a wider range of measures of cognitive and academic abilities.

Following Haworth et al (in press), we suggest that finding substantially greater heritability for language at age 12 than earlier has an intriguing implication: it should be easier to identify genes with effects on language development by studying adolescents, or even adults, than young children. It will be possible to evaluate the accuracy of this prediction, as genome-wide association studies of language development are conducted in the near future.

Acknowledgments

We are grateful for the ongoing contribution of the children and their parents in the Twins Early Development Study (TEDS; R. Plomin, P.I.). TEDS is supported by a program grant (G0500079) from the U.K. Medical Research Council; research on academic achievement is also supported by grants from the U.S. National Institutes of Health (HD44454 and HD46167).

Appendix: Model-fitting evaluation of qualitative genetic differences and quantitative differences in ACE between boys and girls

Appendix

Model-fitting evaluation of qualitative genetic differences and quantitative differences in ACE between boys and girls

	−2LL	df	AIC	BIC	DIC
Vocabulary					
1. Full	25053.815	9011	7031.815	−11280.377	−17316.639
2. Fix r_A	25054.111	9012	7030.111	−11282.871	−17319.803
3. Equate A	25055.511	9013	7029.511	−11284.813	−17322.415
4. Equate C	25056.316	9013	7030.316	−11284.411	−17322.012
5. Equate E	25056.684	9013	7030.684	−11284.227	−17321.828
6. Equate all	**25062.501**	**9015**	**7032.501**	**−11286.602**	**−17325.543**
Listening Grammar					
1. Full	23996.278	8656	6684.278	−10684.324	−16482.699
2. Fix r_A	23996.684	8657	6682.684	−10686.741	−16485.786
3. Equate A	23997.487	8658	6681.487	−10688.960	−16488.675
4. Equate C	23997.089	8658	6681.089	−10689.159	−16488.873
5. Equate E	23998.270	8658	6682.270	−10688.569	−16488.283
6. Equate all	**23998.500**	**8660**	**6678.500**	**−10693.695**	**−16494.749**
Making Inferences					
1. Full	24778.535	8939	6900.535	−11188.897	−17176.911
2. Fix r_A	24778.535	8940	6898.535	−11191.535	−17180.219
3. Equate A	24778.554	8941	6896.554	−11194.163	−17183.517
4. Equate C	24778.595	8941	6896.595	−11194.143	−17183.496
5. Equate E	24778.911	8941	6896.911	−11193.984	−17183.338
6. Equate all	**24782.715**	**8943**	**6896.715**	**−11197.358**	**−17188.051**
Figurative Language					
1. Full	25424.334	9274	6876.334	−11946.061	−18158.565
2. Fix r_A	25424.334	9275	6874.334	−11948.720	−18161.894
3. Equate A	25424.337	9276	6872.337	−11951.377	−18165.221
4. Equate C	25424.341	9276	6872.341	−11951.375	−18165.219

	−2LL	df	AIC	BIC	DIC
5. Equate E	25424.996	9276	6872.996	−11951.048	−18164.892
6. Equate all	**25425.676**	**9278**	**6869.676**	**−11956.025**	**−18171.209**

Note: For all measures, 6 models were fit. Model 1: Full model, in which r_A within DZO twins was allowed to fall below .5 and A, C, E parameters were allowed to differ by sex (i.e., allowing for both qualitative genetic differences and quantitative sex differences). Model 2: r_A within DZO twins was set to .5; A, C, E parameters were allowed to differ by sex. Model 3: r_A within DZO twins was set to .5; A parameters were equated for boys and girls but C and E parameters were allowed to differ by zygosity. Model 4: Similar to Model 3, but C parameters were equated for boys and girls. Model 5: Similar to Model 3, but E parameters were equated for boys and girls. Model 6: r_A within DZO twins was set to .5; A, C and E parameters were equated for boys and girls (i.e., no qualitative or quantitative sex differences). No significant differences were obtained among the models for each measure.

References

Bavin, EL., editor. The Cambridge Handbook of Child Language. Cambridge University Press; New York: 2009.

Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nature Reviews Genetics. 2002; 3:872–882.

Catts, HW.; Hogan, TP.; Adlof, SM. Developmental changes in reading and reading disabilities. In: Catts, HW.; Kamhi, AG., editors. The Connections Between Language and Reading Disabilities. Erlbaum; Mahwah NJ: 2005. p. 25-40.

Davis OSP, Haworth CMS, Plomin R. Learning abilities and disabilities: Generalist genes in early adolescence. Cognitive Neurospsychiatry. 2009; 14:312–331.

Davis OSP, Kovas Y, Harlaar N, Busfield P, McMillan A, Frances J, et al. Generalist genes and the Internet generation: Etiology of learning abilities by web testing at age 10. Genes, Brain and Behavior. 2008; 7:455–462.

Dale PS, Simonoff E, Bishop DVM, Eley TC, Oliver B, Price TS, Purcell S, Stevenson J, Plomin R. Genetic influence on language delay in two-year-old children. Nature Neuroscience. 1998; 1:324–328.

DeThorne LS, Petrill SA, Hart SA, Channell RW, Campbell RJ, Deater-Deckard K, Thompson LA, Vandenbergh DJ. Genetic effects on children’s conversational language use. Journal of Speech-Language-Hearing Research. 2008; 51:423–435.

Dionne G, Dale PS, Boivin M, Plomin R. Genetic evidence for bidirectional effects of early lexical and grammatical development. Child Development. 2003; 74:394–412. [PubMed: 12705562]

Ernst M, Mueller SC. The adolescent brain: Insights from functional neuroimaging research. Developmental Neurobiology. 2008; 68:729–743. [PubMed: 18383544]

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ. Variability in early communicative development. Monographs of the Society for Research in Child Development. 1994; 59(5, Serial No. 242)

Golinkoff, RM.; Hirsh-Pasek, K.; Bloom, L.; Smith, LB.; Woodward, AL.; Akhtar, N.; Tomasello, M.; Hollich, G. Becoming a Word Learning: A Debate on Lexical Acquisition. Oxford University Press; Oxford:

GOAL plc. GOAL formative assessment: Key stage 3. Hodder & Stoughton; London: 2002.

Halpern DF, Benbow CP, Geary DC, Gur RC, Hyde JS, Gernsbacher MA. The science of sex differences in science and mathematics. Psychological Science in the Public Interest. 2007; 8(1)

Hammill, DD.; Brown, VL.; Larsen, SC.; Wiederholt, JL. Test of Adolescent and Adult Language (TOAL-3). Pro-Ed; Austin, TX: 1994.

Haworth CMA, Harlaar N, Kovas Y, Davis OSP, Oliver B, Hayiou-Thomas, et al. Internet testing of large samples needed in genetic research. Twin Research and Human Genetics. 2007; 10:554–563. [PubMed: 17708696]

Haworth CMA, Wright MJ, Luciano M, Martin NG, de Geus EJC, van Beijsterveldt CEM, et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. Molecular Biology. (in press).

Hayiou-Thomas ME, Kovas Y, Harlaar N, Plomin R, Bishop DVM, Dale PS. Common aetiology for diverse language skills in 4 ½ -year-old twins. Journal of Child Language. 2006; 33:339–368. [PubMed: 16826830]

Hedges LV, Nowell A. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science. 1995; 269:41–45. [PubMed: 7604277]

Kovas Y, Haworth CM, Dale PS, Plomin R. The genetic and environmental origins of learning abilities and disabilities in the early school years. Monographs of the Society for Research in Child Development. 2007; 72(3, Serial No. 288)

Kovas Y, Hayiou-Thomas ME, Oliver B, Dale PS, Bishop DVM, Plomin R. Genetic influence in different aspects of language development: The etiology of language skills in 4.5-year-old twins. Child Development. 2005; 76:632–651. [PubMed: 15892783]

Kovas Y, Plomin R. Generalist genes: Implications for the cognitive sciences. Trends in Cognitive Sciences. 2006; 10:198–203. [PubMed: 16580870]

Marchman, VA.; Thal, DJ. Words and grammar. In: Tomasello, M.; Slobin, DI., editors. Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates. Erlbaum; Mahway, NJ: 2005. p. 141-164.

Neale, MC.; Boker, SM.; Xie, G.; Maes, HH. Mx: Statistical modeling. 7th ed. Department of Psychiatry, VCU Box 900126; Richmond, VA 23298, USA: 2006.

Nippold, MA. Later language development: The school-age and adolescent years. 3rd ed. Pro-Ed; Austin, TX: 2007.

Oliver BR, Dale PS, Plomin R. Predicting literacy at age 7 from preliteracy at age 4. Psychological Science. 2005; 16:861–865. [PubMed: 16262770]

Oliver BR, Plomin R. Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. Twin Research and Human Genetics. 2007; 10:96–105. [PubMed: 17539369]

Olson, R.; Byrne, B. Genetic and environmental influences on reading and language ability and disability. In: Catts, HW.; Kamhi, AG., editors. The Connections Between Language and Reading Disabilities. Erlbaum; Mahwah NJ: 2005. p. 173-200.

Plomin, R.; Davis, OSP. Gene-environment interactions and correlations in the development of cognitive abilities and disabilities. In: O'Daly, J.; O'Daly, O.; Murray, RM.; McGuffin, P.; Wright, P., editors. Beyond nature and nurture in psychiatry: Genes, environment and their interplay. Informa Healthcare Medical Books; Oxford, UK: 2006.

Plomin R, Davis OSP. The future of genetics in psychology and psychiatry: Microarrays, genome-wide association, and non-coding RNA. Journal of Child Psychology and Psychiatry. 2009; 50:63–71. [PubMed: 19220590]

Plomin, R.; DeFries, JC.; McClearn, GE.; McGuffin, P. Behavioral genetics. 5th ed. Worth; New York: 2008.

Plomin R, Kovas Y. Generalist genes and learning disabilities. Psychological Bulletin. 2005; 131:592–617. [PubMed: 16060804]

Spear LP. The adolescent brain and age-related behavioral manifestations. Neuroscience and Biobehavioral Reviews. 2000; 24:417–463. [PubMed: 10817843]

Spinath FM, Price TS, Dale PS, Plomin R. The genetic and environmental origins of language disability and ability: A study of language at 2, 3, and 4 years of age in a large community sample of twins. Child Development. 2004; 75:445–454. [PubMed: 15056198]

Wechsler, D. Wechsler Intelligence Scale for Children manual. 3rd ed. Psychological Corporation; London: 1992. UK; WISC-IIIUK

Wiig, EH.; Secord, W.; Sabers, D. Test of Language Competence. Psychological Corporation; San Antonio, TX: 1989. (Expanded ed.)

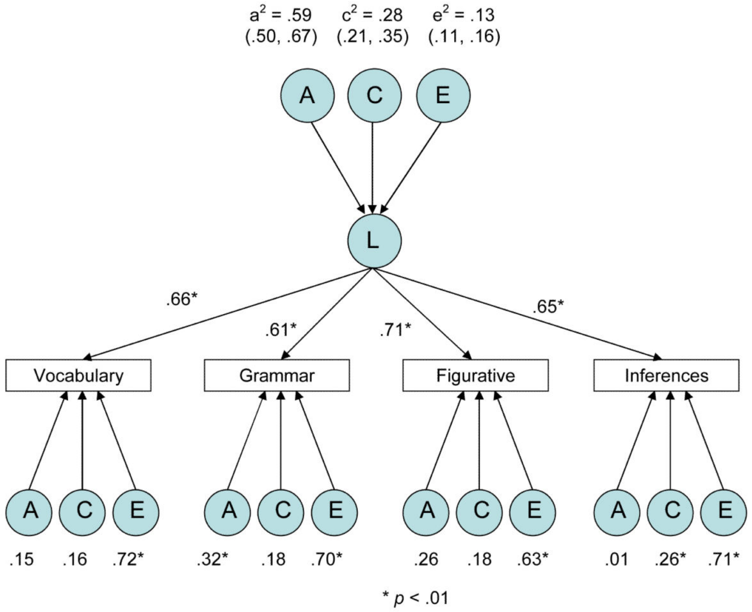

Figure 1.

Common pathway analysis of latent language factor; confidence intervals in parentheses (results for analysis utilizing Vocabulary and Listening Grammar quintiles were highly similar)

Table 1

Means and standard deviations by sex and zygosity

	Sex					Zygosity					Sex * Zygosity		
	Girls* *M (SD)	***Boys* *M (SD)***	***F***	***P***	***μ*** ^{2}	***MZ* *M (SD)***	***DZ* *M (SD)***	***F***	***P***	***μ*** ^{2}	***F***	**p**	***μ*** ^{2}
Vocabulary	.01	.01	.04	.84	.000	−.04	.03	4.59	.03	.001	1.30	.25	.000
	(1.01)	(.99)				(1.02)	(.99)						
Listening	.03	−.06	8.29	.00	.002	−.05	.01	4.54	.03	.001	.24	.62	.000
Grammar	(1.00)	(.99)				(.99)	(1.00)						
Making	.06	−.04	8.46	.00	.002	.00	.02	1.58	.21	.000	.61	.44	.000
Inferences	(.99)	(1.02)				(1.00)	(1.01)						
Figurative	.00	.03	.81	.84	.000	−.01	.03	1.58	.21	.000	.64	.42	.000
Language	(.99)	(.99)				(.99)	(.99)						

Note: M = mean; SD = standard deviation; MZ = monozygotic twins; DZ = dizygotic twins; F, p, and μ^2 derived from MANOVA performed using one member of each twin pair (total N = 4319; 925 MZ girls, 1476 DZ girls, 659 MZ boys, 1259 DZ boys).

Manova

Intercept: F (4, 4312) = 1.39, $p = .24$

Sex: F = (4, 4312) = 5.86, $p < .00$

Zygosity: F = (4, 4312) = 1.75, $p = .137$

Sex * zygosity: F = (4, 4312) = .362, $p = .00$

Table 2

Phenotypic correlations (confidence intervals) among the four language measures

	Listening Grammar	**Making Inferences**	**Figurative Language**
Vocabulary	.40 (.38, .41)	.43 (.41, .44)	.48 (.40, .43)
Listening Grammar		.42 (.40, .43)	.43 (.41, .44)
Making Inferences			.45 (.43, .46)

All correlations significant at $p < .01$

Table 3

Intraclass correlations (95% confidence intervals and n pairs) and model-fitting parameters (95% confidence intervals) for 12-year language measures

Measure	MZ corr	DZ corr	a^2	c^2	e^2
Vocabulary	.43 (.38, .47) 1702	.27 (.23, .30) 3028	.30 (.22, .38)	.13 (.06, .19)	.58 (.57, .61)
Listening Grammar	.46 (.42, .49) 1633	.29 (.26, .33) 2897	.30 (.22, .39)	.15 (.09, .22)	.54 (.51, .58)
Making Inferences	.44 (.40, .48) 1692	.32 (.28, .35) 2997	.25 (.22, .33)	.19 (.17, .25)	.56 (.53, .59)
Figurative Language	.53 (.50, .56) 1758	.36 (.33, .39) 3134	.36 (.31, .44)	.18 (.12, .24)	.46 (.44, .49)

Table 4

Pairwise genetic (above diagonal), shared environment, and nonshared environmental (below diagonal) correlations

	Vocabulary	**Listening Grammar**	**Making Inferences**	**Figurative Language**
Vocabulary	---	.71	.89	.97
Listening Grammar	.86 / .12	---	.94	.72
Making Inferences	.75 / .12	.61 / .10	---	.85
Figurative Language	.66 / .13	.81 / .11	.83 / .08	---

All correlations significant at $p < .01$

Table 5

Decomposition of variance in measured variables into common (across measures) ACE effects and unique (measure-specific) ACE effects.

	Vocabulary	**Listening Grammar**	**Making Inferences**	**Figurative Language**
Factor loading	.66	.61	.65	.71
	.64, .68	.60, .63	.63, .67	.69, .72
Total measured	.44	.38	.42	.50
variance	.43, .45	.36, .40	.40, .45	.48, .52
Proportion of total measured variance:				
Common a^2	.25	.22	.27	.29
	.21, .25	.19, .26	.20, .36	.25,.34
Common c^2	.12	.11	.11	.14
	.09, .16	.09, .13	.10, .14	.10, .18
Common e^2	.06	.05	.06	.07
	.05, .07	.04, .06	.04, .07	.05, .08
Unique a^2	.02	.10	.00	.07
	.00, .02	.02,.17	.00, .00	.00, .13
Unique c^2	.03	.03	.06	.03
	.00, .06	.00, .09	.01, .09	.00, .09
Unique e^2	.52	.49	.50	.40
	.48, .55	.46,.53	.47,.53	.37, .43

Note: Common a^2, c^2, e^2, reflect genetic, shared environmental, and nonshared environmental variance that covaries among the four measures (captured by the latent Language factor); unique a^2, c^2, e^2 refers to genetic, shared environmental, and nonshared environmental variance that is measure-specific.