Genetic and environmental influences on early speech, language and literacy development

Marianna E. Hayiou-Thomas

Department of Psychology, University of York, York YO10 5DD, UK

Received 4 March 2008; accepted 12 March 2008

Abstract

The genetic and environmental etiology of speech and broader language skills was examined in terms of their concurrent relationships in young children; their longitudinal association with reading; and the role they play in defining the ‘heritable phenotype’ for specific language impairment (SLI). The work was based on a large sample of 4 1/2-year-old twins, who were assessed at home on a broad range of speech and language measures as part of the Twins Early Development Study. We found that genetic factors strongly influence variation in young children’s speech in typical development as well as in SLI, and that these genetic factors also account for much of the relationship between early speech and later reading. In contrast, shared environmental factors play a more dominant role for broader language skills, and in relating these skills to later reading; isolated impairments in language as opposed to speech appear to have largely environmental origins.

Learning outcomes: 1) Readers will be able to discuss some of the ways in which behavioural genetic methods can make a useful contribution to the field of communication disorders. 2) Readers will be able to compare the genetic and environmental contributions to general language versus speech skills in young children. 3) Readers will be able to describe the likely relationship between early speech and language and later reading development, in terms of shared genetic and environmental resources. 4) Readers will be able to discuss how different ascertainment methods for clinical samples may lead to very different understandings of the nature of a disorder such as specific language impairment.

© 2008 Elsevier Inc. All rights reserved.

The current paper explores the relationship between speech and language skills in young children, using a behavioural genetics approach. I illustrate this theme using three recently available online at www.sciencedirect.com

Journal of Communication Disorders 41 (2008) 397–408
completed studies: the first examines the nature of the relationship between speech and language in terms of the shared genetic and environmental factors that influence them; the second looks at the extent to which genes and environments that are important for speech and language skills in preschool also have long term impact on children’s reading; and the third study considers how a differentiation between speech and broader language skills may be useful for clarifying the ‘heritable phenotype’ in specific language impairment.

1. Background on TEDS and the twin method

This work is based on data from the Twins Early Development Study (TEDS), which is a large population-based sample of twins in the UK (PI: Robert Plomin). The children have been followed longitudinally from the ages of 2–12, and assessed regularly on measures of language and reading, as well as other cognitive and behavioural variables (Oliver & Plomin, 2007). I am going to focus specifically on a subsample of approximately 1600 children from TEDS who were visited at home when they were 4 1/2 years old, and assessed on a broad range of language and nonverbal measures. Part of the purpose of this project was to examine the causes of language impairment, and this subsample was selected to have a relatively high proportion of children at risk of language difficulties, based on a parental report of vocabulary at 4 years. This is an important point to bear in mind when interpreting the results of the first two studies presented here: the focus of the studies is on individual differences among children across the full range of ability, but it is possible that we would see a different pattern if our sample were not over-represented for children with poor language skills.

Twins provide a natural experiment which is ideally suited to examining the relative contributions of nature and nurture to variations in human traits and behaviours. Identical twins (also known as monozygotic, or MZ), share 100% of their DNA, while fraternal twins (dizygotic, or DZ) share approximately 50% of the DNA that varies from person to person. The twin method is based on comparing the similarity of the members of a twin pair on a given trait, such as vocabulary size, for MZ and DZ twin pairs. If MZ twins have much more similar vocabulary scores than DZ twins, we can infer that the extra similarity in MZ vocabulary comes from these twins’ extra genetic similarity. On the other hand, if MZ and DZ twins are equally similar, then we can infer that something in the twins’ environment is driving that similarity, since we assume that MZ and DZ twins share their environments to the same extent (the ‘equal environments assumption’). Using this comparison we can decompose the total variance on any given trait into its genetic and environmental components. The environmental sources of variance can be further broken down into ‘shared’ and ‘nonshared’ environments: shared environments are defined as factors which make children within a family more similar to each other (such as attending the same school), while nonshared environments are factors which make children different from each other (for example, an illness that only one child experiences; measurement error is also included in the nonshared environmental estimate). More extensive discussion of the twin method, as well as of its limitations, can be found in Plomin, DeFries, McClearn, & McGuffin (2001). For the purposes of this paper, the key points to underline are that the MZ–DZ comparison allows us to estimate the contribution of genetic, shared
environmental and nonshared environmental factors on a given trait or disorder. This method can also be extended to examine the genetic and environmental contributions to the relationship between two traits. Finally, it is important to emphasise that behavioural genetics focuses on individual differences between people, rather than on species-universals. The studies I discuss are concerned with the factors that underlie the differences in children’s language skills, rather than on the common patterns that characterise all children’s language development and which may be due to completely different factors.

2. Speech and language: etiology of the relationship

There is a large literature examining the relationship between different components of the developing language system, both at the species-universal and the individual differences levels of analysis. There is also a growing literature examining the genetic and environmental etiology of different language skills and difficulties (e.g., Stromswold, 2001). However, there is relatively little research that focuses explicitly on the etiology of the relationships between different components of language. One recent example of such a study examined the relationship between lexical and syntactic growth in 2 and 3 year olds, using data from a parental checklist of words and syntactic structures produced by the child (an adaptation of the McArthur–Bates Communicative Developmental Inventory (Dale, Reznick, & Thal, 1998; Fenson et al., 1994; Reznick and Thal, 1998). There were strong genetic correlations between vocabulary and grammar at 2 and 3 years of age, indicating that many of the genes influencing individual differences in these two skills are shared. Similarly, strong shared environmental correlations indicated that many of the environments influencing vocabulary and grammar are shared. This etiological overlap was also evident longitudinally, in that the genes and environments important for 2-year vocabulary were also important for 3-year grammar, and vice versa (Dionne, Dale, Boivin, & Plomin, 2003). Thus, there seem to be strong genetic and environmental links between vocabulary and grammatical skills in toddlers, when looking at individual differences in these skills across the whole range of ability. Conversely, there do not seem to be strong genetic links between grammatical and phonological skills. In a study focusing on poor language skills in 6-year olds, Bishop, Adams, and Norbury (2006) found that deficits in phonological short-term memory and in grammatical inflections were both highly heritable, but that there was no significant overlap in the genetic influences on these two tasks. The two sets of results just described are not directly comparable, in that they focus on different ages as well as different levels of ability, but put together they nonetheless suggest a pattern in which vocabulary and grammar are genetically related skills, but phonology and grammar are not.

Study 1 (Hayiou-Thomas et al., 2006) aimed to examine the ‘cognitive architecture’ of young children’s language, using behavioural genetic methodology. Specifically, we were interested in the etiology of the inter-relationships of different language components, including vocabulary, grammar and phonology, within a single study. A subsample of 4 1/2-year-old twins from the Twins Early Development Study was assessed at home on nine language measures which were selected to tap a broad range of language skills. Phenotypic factor analysis yielded two latent factors, which suggested a broad distinction between
general language and speech. The ‘general language’ factor included 7 of the 9 measures, tapping expressive semantics (MSCA Word Knowledge, MSCA Verbal Fluency (McCarthy, 1972), Bus Story Test (Renfrew, 1997a)); expressive syntax (Action Picture Test grammar score (Renfrew, 1997b)); receptive syntax (BAS verbal comprehension subtest (Elliot, Smith, & McCulloch, 1996)); verbal memory (MSCA Verbal Memory Words and Sentences (McCarthy, 1972)); and receptive phonology (Phonological Awareness task (Viding et al., 2003)). The ‘speech’ factor included the remaining two measures in the battery: a test of articulation and a test of phonological memory (Sounds-in-Words Subtest (Goldman & Fristoe, 1986), and the Children’s Test of Nonword Repetition (Gathercole & Baddeley, 1996)). This broad grouping, whereby there is a clear differentiation between measures of language versus speech but relatively little differentiation within these domains, is consistent with other reports in the literature (Tomblin & Zhang, 1999; Tomblin, Zhang, Weiss, Catts, & Ellis Weismer, 2003).

Our primary interest was to see whether and how this phenotypic pattern would be reflected in the genetic and environmental etiology, in terms of the following questions: (a) What is the relative magnitude of genetic and environmental influences on the language and speech latent factors? (b) To what extent do the genetic and environmental factors affecting speech and language overlap? (c) Are there unique genetic or environmental influences acting on individual language measures that do not act through the latent factors?

Using a latent factors model (Fig. 1), we found that the main source of variance in the language factor was the children’s shared environment, accounting for half of the total
variance in this factor \((c^2 = .50)\); in addition, there was a significant though more moderate contribution from additive genetic influences \((h^2 = .34)\). Nonshared environmental influences accounted for a modest – though still significant – proportion of the total variance on the language factor \((e^2 = .15)\). A somewhat different pattern was apparent for the speech factor: more than half the total variance was attributable to genetic influences \((h^2 = .56)\) with much smaller (though significant) contributions from shared environment \((c^2 = .15)\) and nonshared environment \((e^2 = .29)\). The language and speech factors were strongly related phenotypically \((r_p = .60)\), and this was reflected in their etiological relationship as well: the genetic correlation between them was \(r_g = .64\), and the shared environmental correlation was \(r_c = 1.00\). That is, two thirds of any genetic influences acting on language also play a role in individual differences in speech, and any aspect of the shared environment that is important for language also plays a role in speech. It is important to differentiate here the overall magnitude of the genetic and environmental effects, from the overlap of these effects on speech and language. For example, although the shared environmental correlation of 1.00 means that there are no distinct environmental factors on language that will not also have some impact on speech; it is also the case that those environments will be much more important for language than they will be for speech. Finally, although the latent factors captured much of the variance in the individual language measures, we were interested to see whether there were any additional and unique genetic or environmental influences acting on individual measures. The short answer is No: the unique genetic and shared environmental influences were negligible and in nearly all cases were not significantly different from zero.

In summary, the results of this study of 4 1/2-year language suggest that diverse linguistic skills in young children are very closely related etiologically, and that genetic and shared environmental influences are largely shared among measures. Nonetheless, there is a significant etiological distinction between ‘general language’ and ‘speech’, in that the genetic overlap between them, though strong, is not complete. There is also an etiological distinction between these two factors in that, though both genetic and environmental influences play a role for both speech and language, the dominant influences on language stem from children’s shared environment, while the dominant influences on speech are genetic.

3. Speech, language and reading: etiology of the longitudinal association

Following on from these findings, an interesting question that arises is how early speech and language skills relate to subsequent literacy development. Again, there is an extensive phenotypic literature on the relationship between oral language and literacy, with a consensus that oral language – and in particular phonological abilities – forms the foundation on which reading and related skills are based. The close relationship between

---

\(^1\) Heritability is defined as the proportion of variance on a given trait that can be accounted for genetic factors, and is commonly denoted as \(h^2\). Shared and nonshared environmentality are denoted as \(c^2\) and \(e^2\), respectively. In multivariate analysis, the genetic, shared and nonshared environmental correlations are denoted as \(r_g\), \(r_c\) and \(r_e\), respectively; \(r_p\) refers to the total phenotypic correlation.
phonology and reading is supported by behavioural genetic evidence showing that these two abilities are both highly heritable, and that to a large extent they are influenced by the same genetic factors. The concurrent genetic correlation between phonological awareness and reading has been estimated in the region of .70–.80, in children with typical literacy development as well as those with reading difficulties (Gayán & Olson, 2001, 2003). In addition, reading difficulties have been found to be more heritable in groups of children who also have deficits in phonological short-term memory (Bishop, Adams, & Norbury, 2004), suggesting that the close genetic relationship between phonology and reading extends beyond phonological awareness to the phonological system more generally. This pattern also holds longitudinally, in that the genetic factors influencing phonological awareness in preschool go on to affect children’s reading a year later (Byrne et al., 2005).

Less well established is the nature of the relationship between nonphonological language skills and reading. From the behavioural genetics literature, there is now some evidence that shared environmental factors, in addition to genetics, may be important in mediating the relationship between lexical and syntactic abilities in toddlers and preschool children, and literacy skills in the early school years (Harlaar, Hayiou-Thomas, Dale, & Plomin, in press).

In Study 2 (Hayiou-Thomas, Harlaar, Dale, & Polmin, submitted for publication), we used behavioural genetic methodology to examine the longitudinal issue of how preschool language and speech predict literacy skills in the primary school years, in terms of shared etiology. Specifically, we focused on the genetic and environmental relationship between the language and speech latent factors at 4 1/2 years and reading ability assessed at 7, 9 and 10 years of age. Reading was assessed by teacher-ratings, following the UK National Curriculum Criteria, at ages 7, 9 and 10 (Department for Education and Employment, 1998). In addition, word and nonword recognition was directly measured using the TOWRE at 7 years, and sentence reading comprehension was measured using the PIAT at 10 years. We used a set of latent factors models similar to those described in Study 1, in which the Language and Speech latent factors were each related to three reading measures: a latent factor comprising the three teacher-rated reading scores, the TOWRE, and the PIAT. The key statistics of interest in these analyses are the bivariate heritability and environmentality, which indicate the proportion of the overall phenotypic relationship between two variables which can be accounted for by shared genetic or environmental factors.

We found a strong phenotypic relationship ($r_p = .47–.68$) between the language factor at 4 1/2 years and each of the three reading variables (Fig. 2). This relationship was mediated by both genetic and shared environmental factors. In the case of teacher-rated reading and the PIAT, the association with the language factor was attributable roughly equally to genetic and shared environmental factors that affect both language and reading (bivariate $h^2 = .49$ and .56, respectively, bivariate $c^2 = .50$ and .44, respectively). In the case of the TOWRE, most of the phenotypic association was attributable to shared environmental factors affecting both preschool language and word recognition (bivariate $h^2 = .19$, bivariate $c^2 = .80$). By contrast, the relationship between the Speech factor and the reading variables was predominantly due to genetic factors (Fig. 3). In this case, the overall phenotypic association was moderate ($r_p = .39–.51$), and the Speech–TOWRE and Speech–PIAT associations were largely accounted for by shared genetic factors (bivariate
\( h^2 = .68 \) and \( .72 \), respectively); the bivariate environmentality was nonsignificant. There was, however, a significant contribution from shared environment to the association between speech at 4 1/2 years and teacher-rated reading (bivariate \( h^2 = .60 \), bivariate \( c^2 = .40 \)).

In conclusion, we found a stable and moderate-to-strong phenotypic prediction from language and speech skills at 4 1/2 years of age, to a range of reading measures at 7, 9 and 10 years of age. The etiology of this prediction differed for the language and speech factors: while the relationship between early language and later reading appears to be mediated by both genetic and shared environmental factors, the association between early speech and later reading is dominated by genetic factors that these two skills have in common.
4. Speech or language: the heritable phenotype in SLI?

The previous two studies described were concerned with individual differences across the whole range of ability. The final study (Bishop & Hayiou-Thomas, 2008) focuses specifically on a group of children who were identified as having specific language impairment. This work was motivated by a striking anomaly in the literature looking at genetic influences on SLI: while four out of five twin studies reported very strong genetic effects (Bishop, North, & Donlan, 1995; Lewis & Thompson, 1992; Tomblin & Buckwalter, 1998), one study – based on the 4 1/2-year TEDS sample – found negligible heritability for SLI (Hayiou-Thomas, Oliver, & Plomin, 2005). On re-examining the data, we found that the reason for this large discrepancy had to do with ascertainment. While the previous studies initially recruited their SLI samples on the basis of clinical or parental concern, the TEDS children with SLI were defined purely on the basis of psychometric criteria: more than 1 S.D. below the mean on the 4 1/2-year language composite described earlier, and better than 1 S.D. below the mean on a nonverbal composite (comprising the MSCA Block Building, Puzzle Solving, Draw-a-Design and Tapping Sequence (McCarthy, 1972)). When the SLI sample was redefined in terms of children who had had contact with speech and language therapy, however, we found that the heritability estimate was .97; that is, entirely in line with all the previous studies. The etiology of ‘clinical SLI’ therefore appears to be very different to the etiology of ‘psychometric SLI’ (Fig. 4).

The critical question following on from this finding is: what differentiates these two groups of children? One possibility is that the clinically referred children presented with more severe language impairment, as our psychometric cut-off criterion of −1 S.D. would allow children with quite mild impairments to be classified as SLI. However, the level of initial impairment was not significantly different for the ‘clinical’ and ‘psychometric’ SLI groups. A second possibility is that the clinically referred children had more persistent language impairment: however, a comparison of vocabulary scores at age 7 showed that these also did not differ significantly between the two groups. The final possibility we examined is that the two groups had different phenotypic profiles: that there was something distinctive about the type of language impairment that the ‘clinical SLI’ group exhibited, and that made it more likely that a child would be referred for therapy. We found that this
did differentiate the two groups, and that the ‘clinical SLI’ group had significantly lower scores on the speech composite at 4 1/2 years than the ‘psychometric SLI’ group. This finding is consistent with previous work showing that poor speech skills are more likely to result in a referral than poor general language skills (Zhang & Tomblin, 2000). Finally, in directly testing the heritability of poor performance on the language composite as compared to the speech composite, we found that speech deficits were indeed more heritable than broader language deficits, and that the more extreme the speech deficit, the stronger the genetic influence.

This study suggests that SLI is highly heritable only when the sample is selected on the basis of parental or professional concern. The children who are most likely to provoke such concern are children with poorer speech skills, and it appears that it is these speech deficits that are the locus of the genetic effects on language impairment. Although it is likely that many children with speech problems will also have broader language difficulties and vice versa, it seems that ‘pure’ language impairments are largely environmental in origin.

5. Conclusions

The set of studies described in this paper suggest that there is a useful distinction to be made, in terms of etiology, between general language and speech skills. Differences in young children’s language skills, such as vocabulary and grammar, appear to be largely due to environmental influences, though genetic effects also play a significant role. Differences in speech skills, on the other hand, appear to be mostly due to genetic effects, though environmental factors also play a significant role. This pattern is reflected in the longitudinal relationship between early speech and language skills, and subsequent literacy development. While the prediction from general language skills at 4 1/2 years, to reading at 7, 9 and 10 years was mediated by both genetic and environmental factors, the prediction from speech to later reading was mediated predominantly by genetic factors. Finally, the distinction between speech and language appears to be particularly important in defining the ‘heritable phenotype’ for specific language impairment: ‘pure’ language impairments seem to have environmental origins, while speech impairments are largely genetic.

More generally, I hope that this work demonstrates that behavioural genetic methods go beyond estimating simple heritability, and can be a useful tool for addressing issues that are of theoretical interest, and perhaps even practical use, to the field of language development and childhood language disorders.

Acknowledgements

Many thanks to the families of participating twins, and to Robert Plomin, Andy McMillan, and staff at the Twins Early Development Study. Thanks also to Philip Dale, who kindly presented this work on my behalf at the 2007 ASHA Research Symposium.
Appendix A. Continuing education

1. Two identical twins had the same score on a reading test (50th centile). Two fraternal twins had very different scores on the reading test (30th and 70th centiles). What does this suggest about the etiology of individual differences in reading? (Bearing in mind that one would need a very large sample of twins, rather than two pairs!).
   a. Reading skill is entirely a function of experience.
   b. There is at least some genetic influence on reading.
   c. Fraternal twins are better readers than identical twins.
   d. Both genes and environments always contribute to any given trait.

2. Shared environmental factors include:
   a. Anything in the family environment.
   b. Measurement error.
   c. Any environmental factor that makes two children in the same family more similar to each other.
   d. Environmental factors that are shared by identical but not fraternal twins.

3. The first study reported in this paper found that speech and language skills:
   a. Were both influenced by genetic as well as environmental factors.
   b. Had different patterns of genetic and environmental influence.
   c. Could be distinguished genetically, even though there was a significant genetic correlation between them.
   d. All of the above.

4. According to the findings of the second study reported in this paper, the relationships between early language and speech and subsequent reading, are:
   a. Moderate.
   b. Stable.
   c. Mediated by both genetic and environmental factors in the case of language, and mostly genetic factors in the case of speech.
   d. All of the above.

5. The third study reported in this paper suggests that different methods of ascertainment for SLI result in:
   a. Groups of children with different profiles of difficulty, and different etiologies.
   b. Erroneous identification of the disorder.
   c. Similar estimates of the heritability of SLI.
   d. Essentially the same groups of children being identified with SLI.

References


