Understanding the genetic architecture of language- and literacy-related abilities during mid-childhood and adolescence: Evidence for genetically shared factors with early vocabulary

Ellen Verhoeft1,2, Chin Yang Shapland1, Philip S. Dale3, Simon E. Fisher1,4, Beate St Pourcain1,4,5

1. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
2. International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
3. Speech & Hearing Sciences, University of New Mexico, Albuquerque, USA
4. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
5. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

Background

Mastering language skills is an important development milestone and influences children's cognitive development during later life. The heritability of language abilities increases during child and adolescent development, as shown by longitudinal twin studies in 2 to 12-year old twins. Moreover, language- and literacy-related abilities (LRAs) in mid-childhood are highly genetically correlated with each other. However, knowledge about early genetic predictors of language and literacy skills is scarce. Here, we apply genetic-relationship-matrix structural equation modelling (GSEM), a multivariate analysis framework in unrelated individuals, to model genetically shared factors between early vocabulary and multiple LRAs during mid-childhood and adolescence.

Methods

Expressive and receptive vocabulary at 38 months, as well as thirteen LRAs related to reading, spelling, phonemic awareness, listening comprehension, non-word repetition and verbal intelligence (7-13 years) were studied in children from a UK birth cohort (ALSPAC; N≤ 6,092). GSEM based on Cholesky decomposition was applied to investigate developmental genetic architectures using rank-transformed measures. Due to computational constraints, thirteen GSEM models each consisting of the two early vocabulary measures and one mid-childhood LRA were defined. Path coefficients across multiple GSEM models were combined using random-effects meta-regression. In addition, we estimated the factorial co-heritability as the proportion of genetic variance explained by a genetic factor with respect to the total genetic variance of a trait.

Results

Across all Cholesky models, we observed one shared genetic factor between both vocabulary measures, with up to 0.75(SE=0.25) of the genetic variance in early receptive vocabulary being accounted for by genetic variance in expressive vocabulary. This common factor explained ~8.5% phenotypic variation in receptive vocabulary (standardised path coefficient: -0.29(SE=0.08)). A second genetic factor that was unique to early receptive vocabulary accounted for ~3.8% of phenotypic variation (standardised path coefficient: 0.20(SE=0.04)). The early common genetic factor also predicted ~8.6% phenotypic variation in verbal intelligence scores (standardised path coefficient: -0.29(SE=0.10)). However, there was little evidence that this factor contributed to phenotypic variation in other LRAs. The second genetic factor, unique to early receptive vocabulary, explained genetic variance for many literacy, but not language-related traits. Specifically, it accounted for ~43% of
phenotypic variance in pooled reading-related abilities, with little evidence for heterogeneity (pooled standardised path coefficient: 0.66(SE=0.03), $\rho_{het}=0.93$). This captured the majority of the total genetic variance for reading-related measures during mid-childhood and adolescence (factorial co-heritability for e.g. reading accuracy and comprehension age 7: 0.94(SE=0.08)). There was no evidence for genetic factors that are specific for mid-childhood LRAs, across any of the models studied.

**Discussion**

Genetic variation in literacy-related abilities during mid-childhood can be fully accounted for by genetic factors influencing receptive, but not expressive vocabulary, at the age of 38 months. In contrast, genetic variation in verbal intelligence is predicted by genetic factors influencing both early expressive and receptive vocabulary. Our findings suggest that genetically predictable biological mechanisms affecting LRAs during mid-childhood and adolescence emerge already at an early age.