Genetic and Environmental Influences on Literacy and Numeracy in Australian School Children

Katrina Lyn Grasby
BPsych(Hons), University of New England, 2012

A thesis submitted for the degree of Doctor of Philosophy
of the University of New England

January 2016

Certification

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree or qualification.

I certify that any help received in preparing this thesis and all sources used have been acknowledged in this thesis.

Signature

Acknowledgements

I thank William Coventry, who suggested undertaking this PhD. Unsurprisingly, this work as been challenging and rewarding in turns, as each new hurdle is crossed and each new script conquered. Will, I am so deeply appreciative of your enthusiasm, patience, and kindness. I thank Brian Byrne, who, with a ready sense of humour, would calmly and astutely remind me of the bigger picture. I am incredibly privileged to have had the opportunity to work with both of you. I thank Sarah Medland, who has rescued me from scripting errors that were intractable to me. I also thank Sally Larsen, who tirelessly chased up states and families for data and whose efficiency and administrative organisation made it possible for me to have data for this thesis.

For my parents, who with boundless love and kindness have supported me emotionally, financially, and physically helped me move house twice though these PhD years. I thank my brother for his attention to detail on reading this thesis through. Particularly, for all my friends and family, who have indirectly helped me accomplish this work by having sympathetic ears and sound advice when the challenges came and sharing in my pleasure when the challenges were negotiated.

Preface

This thesis is by publication. The thesis is composed of an introductory chapter, which outlines and provides context on the topic, methodology, and purpose of the subsequent papers. The body of the thesis is composed of manuscripts, either accepted, in submission, or in preparation to submit to a journal. For each of these chapters the author's contribution and statement of originality are provided. The concluding chapter links together the manuscripts and highlights the contribution of the body of work. Cited works are referenced at the end of the chapter in which they are cited.

Table of Contents

List of Tables

List of Figures

List of Appendix Tables

List of Abbreviations

A	Additive genetic
ACARA	Australian Curriculum Assessment and Reporting Authority
AIC	Akaike's Information Criterion
BST	Basic Skills Test
C	Common or shared environment
CLDRC	Colorado Learning and Disabilities Research Centre
DZ	Dizygotic
E	Unique environment
FTPR	Florida Twin Project on Reading
ICSEA	Index of Community Socio-Educational Advantage
IA	Individually administered
LBOTE	Language Background Other Than English
ILTS	International Longitudinal Twin Study
LRT	Likelihood ratio test
MCEETYA	Ministerial Council on Education, Employment, Training, and Youth Affairs,
MZ	Monozygotic
NAPLAN	National Assessment Program: Literacy and Numeracy
NSW	New South Wales
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
r_A	Additive genetic correlation
r_C	Shared environment correlation
r_E	Unique environment correlation
SES	Socioeconomic status
TAFE	Technical And Further Education
TEDS	Twins Early Development Study
UK	United Kingdom
USA	United States of America
WRRMP	Western Reserve Reading and Math Project

Abstract

Each year, Australian students in Grades 3, 5, 7, and 9 sit nationwide tests in literacy and numeracy. These tests inform government, principals, and parents about student, school, and state performance in five domains: reading, spelling, grammar and punctuation, writing, and numeracy. As such, the results of these tests are of wide interest for diverse reasons depending on the stakeholder in question. In this thesis I examine the influence of genes and the environment on individual differences in performance on these tests. Using longitudinal data collected from a large sample of Australian twins and their siblings.

Initially, as a test of validity, I compared the performance of large-scale reading tests against three literacy tests in comprehension, word reading and vocabulary individually administered to twins in Grade 3. The individually administered tests accounted for a substantial amount of the variance in the large-scale reading tests. Additionally, they were preferentially related, both genetically and environmentally, to large-scale reading tests compared to large-scale numeracy tests, confirming that large-scale school reading tests measure, at least in part, the literacy skills tapped by individual tests considered "gold-standard" in testing.

In the second paper, I examined the extent to which genes and the environment contributed to variation in and covariation among the five domains in each grade. Averaged across domains and grade, genetic factors explained 60%, shared environment 10%, and unique environment 30% of the variation. Independent pathway models showed similar genetic and environmental structures at each grade with approximately one third to one half of the variation in each domain due to genes that influenced all domains.

In the third paper, I explored the genetic and environmental influences on stability and growth in each of the domains. Stability in performance was primary due to genes. For growth, reading followed a compensatory growth pattern, and variation in growth was due to the genes that also influenced differences in performance at initial testing. By contrast, growth in numeracy was principally influenced by unique environmental factors. These results suggest individual differences in growth of reading are primarily due to a genetically influenced developmental delay in the acquisition of necessary skills, while for numeracy, differences are due to environmental influences, such as different teachers or interests.

In the fourth paper, I tested if family or school SES moderated heritability of performance. Genetic influence was substantial and stable across all levels of family and school SES, with some evidence of a stronger influence of the shared environment when SES was lower, particularly for Grade 3 literacy. A final chapter presents a discussion summarising the principal findings, their implications, and their limitations.

Chapter 1

Introduction

Chapter 1

Literacy and numeracy

Literacy and numeracy are the founding skills of educational achievement. Literacy is a somewhat elusive term despite its extensive use. In English, historically, a literate person was "familiar with literature" or was well-educated, but it came to mean a person was simply able to read and write the information of daily life (UNESCO Institute for Statistics, 2008). By the turn of the 20th century the concept of multiliteracy had developed, which extended the traditional view of literacy to recognise that literacy might have a different function and meaning depending on the cultural and social group (Moss, 2009). Multiliteracy also explored the multiple modalities required to effectively communicate information in the modern world, such as visual, computer, and multimedia literacies. This has lead to terms such as emotional literacy and political literacy where "literacy" becomes a synonym for "competence" (Street, 2009). For the purpose of this thesis, literacy will refer to traditional print literacy, the ability to read and write.

Reading and writing are the culmination of an array of dependent skills. In learning to read a child must recognise that oral language is composed of individual units of sound (phonemic awareness; (Griffith & Olson, 1992). They must be able to manipulate phonemes, both segment words into syllables, onset, and rime, and blend phonemes together to form syllables and words (phonological awareness; Adams, 1990). They must acquire the *alphabetic principle*: That words are composed of letters (graphemes) that systematically represent phonemes (Byrne, 1998). Developing orthographic knowledge (memory for spelling) and chunking words into syllables and morphemes facilitates the process of encoding, into memory, both regular and irregular words (Ehri, 2005). For reading comprehension, words and text must be assigned a meaning via vocabulary and syntax (Hoover & Gough, 1990). Reading fluency is accomplished with the maturity of these skills. Most of these component skills are constrained, such that all competent readers acquire them (although they might differ in age or rate of acquisition); however, extent of vocabulary and degree of comprehension continue to develop (Paris, 2005).

Reading is integral to writing, but writing requires additional skills. Writing has three broad components: planning, formulation, and revision. Planning requires ideas to be generated, the organizing and sequencing of relevant ideas, and how to achieve the purpose of the text (Alamargot & Fayol, 2009). Content is dependent on knowledge and directed by the nature, or genre, of the text and the relationship of the author to the reader (Myhill, 2008). Formulation first involves the translating of thought to linguistic representations (text generation) through the selecting and structuring of words. Secondly, generated text must be represented with written symbols (transcription), which is dependent on spelling, grammar, punctuation, and fine motor skills (Berninger et al., 1992). A writer must employ appropriate sentence structure for writing, as

distinguished from speech structure, and to communicate clearly must account for the comprehension needs of an absent reader (Kress, 2003). Revision requires: text to be read and comprehended; the written text is compared with the intended goal; discrepancies and errors detected and diagnosed, and correction strategies generated, decided upon, and executed (Chanquoy, 2009). Errors and corrections might apply at the level of word, phrase, sentence, or idea. While there is developmental dissociation among the component skills of reading and writing (Berninger & Hart, 1992), there is also a dynamic association at play in developing the component skills between these two literacy domains (Adams, 1990; Graham & Hebert, 2011).

Numeracy, like literacy, is defined by context; it is the ability to use numbers and apply mathematical skills to the problems of daily life (Collins & O'Brien, 2003). At the most basic level is the understanding of numbers. Despite an innate sense of small numerosities (number of objects in a set), several number principals that derive from numerosity must be acquired, such as stable order, one-to-one correspondence, cardinality, and abstractness (Butterworth, 2005). Moreover, in order to solve mathematical problems both the concept of numerosity and skills at manipulating numerosities need to develop (Butterworth, 2005). Beyond understanding numbers, there are several broad strands of mathematical skills including measurement, geometry, algebra, statistics, and probability. In addition to mathematical knowledge, reasoning and problem solving skills have to be acquired. Applying mathematical skills to a problem makes various cognitive demands: the problem must be defined, the necessary information encoded, a strategy to employ selected, the necessary computations made, and the response executed (Anderson & Fincham, 2014; Anderson, Lee, & Fincham, 2014). Thus, to be numerate, numerosity and specific mathematical knowledge for the nature of the problem must be employed in conjunction with appropriate solving problems skills.

Improving our understanding of the cognitive and developmental processes engaged in acquiring literacy and numeracy are of particular use in refining methods to teach these fundamental skills, and the benefits of literacy and numeracy are wide-reaching. Literacy and numeracy are related to employment, and this relationship is largely independent of years of education (Chiswick, Lee, & Miller, 2003). Higher literacy and numeracy are correlated with higher income, better health, and greater civic engagement (Organisation for Economic Co-operation and Development [OECD], 2013). On a national level, higher average literacy is related to greater gross domestic product (GDP; Coulombe, Tremblay, & Marchand, 2004). As such, there are ample reasons—political, social, and individual motivation—to be literate and numerate. With social and political incentive to have a literate and numerate society, standardised national testing has become a feature of the education systems in most economically developed countries in recent

decades (Masters, Rowley, Ainley, & Khoo, 2008; Parveva, De Coster, & Noorani, 2009; Porter, McMaken, Hwang, & Yang, 2011). The Program for International Student Assessment (PISA), which started in 2000, expanded standardised assessment to an international stage; consequently performance could be compared between as well as within countries (www.oecd.org). Standardised measures allow for quantitatively assessment of national performance through an objective lens.

Education and national testing in Australia

Historically, education in Australian is under state jurisdiction; although, the 2008 Melbourne Declaration on Educational Goals for Young Australians set objectives that led to greater federal involvement (Ministerial Council on Education, Employment, Training, and Youth Affairs [MCEETYA], 2008). The intent of reforms and agreements with the federal government, culminating in the National Education Reform Agreement and the Australian Education Act of 2013, was to increase consistency in education across the nation (Council of Australian Governments, 2013). Consequently, the Australian Curriculum, Assessment and Reporting Authority (ACARA) was established and mandated to develop a national curriculum and manage national assessment (MCEETYA, 2009). The National Assessment Program: Literacy and Numeracy (NAPLAN) developed out of state-level assessments, to provide nationally comparable data to inform governments, communities, schools, and parents (Masters et al., 2008). Since 2008, students in Grades 3, 5, 7, and 9 have sat tests in reading, spelling, grammar and punctuation, writing, and numeracy (Senate Standing Committee on Education and Employment, 2014). Since 2010, summaries of the results of all schools, at the group level, have been available to the public on the *My School* website (www.myschool.edu.au).

Since the *My School* website began, public interest in NAPLAN has been extensive, and opinion pieces in newspapers abound (e.g. Athanasou, 2013; Bantick, 2015; Coulson, 2011; Job, 2013). Certainly the tests are limited. They are administered biennially, are a single appraisal of student proficiency, and the results are not available until four months after the tests are administered. As such the tests are not formative assessments (to monitor learning), rather they are a type of summative assessment (to evaluate learning). The tests assess the domains of reading, spelling, grammar and punctuation, writing, and numeracy at a broad level. (These tests are described in detail in the method sections of the research chapters of this thesis). As such, the component skills are not specifically assessed, and diagnosis of the source of illiteracy or innumeracy in an individual is not possible. Regardless of the imperfections inherent to these tests, there are several strengths. They are administered nationally to students living in diverse regions of the country, from the extremely isolated to the metropolis. All types of schools administer the tests,

including government, selective, Catholic, independent, and alternative delivery of education such as home schools and distance education.

Behaviour genetics and the classic twin design

When going beyond the question of *if* a person is literate and numerate to the question of *how well* a person can read, write, or enumerate, it is possible to explore both the average ability and variation in ability. The field of behaviour genetics explores genetic and environmental influences on the latter.

Introduction to genetics

Using twins to systematically explore the influence of nature and nurture stretches back to the mid-1800s. In 1875, Galton [35]reported a qualitative review of twin similarity and suggested there were strong inherent influences on differences among individuals. Quantitative assessment of covariation developed from an interest in measuring the relative similarity on characteristics among related individuals, primarily parent-offspring and fraternal dyads (Galton, 1888; Pearson & Lee, 1903). From these methodological advances it became apparent that (a) many human traits that varied followed a normal distribution, (b) measures on a trait were linearly related between relatives, and (c) related individuals correlated on traits according to degree of relatedness. However, it would be a different field of enquiry that would understand the degree of relatedness on a genetic level.

Also in the mid-1800s, Mendel (1866) reported on a series of experiments in pea plants and described what became known as the basic laws of heredity. The law of segregation: At a genetic locus, each offspring carried two alleles (versions of a gene), one from each parent, and these alleles assort into sex cells, which in turn become parent cells to a future generation. The law of independent assortment: Alleles for different traits assort into sex cells independently. Mendel's laws were fused with the findings of biological cytologists and the chromosomal theory of heredity was proposed early in the 20th century (Baxter & Farley, 1979), thus explaining how each individual received half of their mother's genes and half of their father's genes. This provided a biological basis for the two types of twins observed by Galton; identical (monozygotic; MZ) twins developed from the same zygote and fraternal (dizygotic; DZ) twins developed from different zygotes (Hall, 2003).

The Classic Twin Design

The classic twin design is a method of partitioning phenotypic (P) variance into genes and environmental effects. Genetic effects might be *additive* or *non-additive*. Additive (A) genetic effects are the summative effect of alleles. Non-additive (NA) genetic effects include dominant

effects (interactive effects between alleles at a genetic locus) and epistatic effects (interactive effects between alleles at different genetic loci). Environmental effects might be shared or specific. Shared or common (C) environmental effects—for twins raised together—include for example home location, family socioeconomic status, and number of siblings. Specific or unique (E) environmental effects include environmental factors that are experienced by one twin but not their co-twin, for example different teachers, activities, or friends. Unique environmental effects also include impact of events that are shared by both twins but are experienced differently. Within this model, measurement error is also incorporated into E.

In order to partition phenotypic variance, the covariation of MZ twins on a trait is compared to that of DZ twins. This comparison rests on three premises. First, that phenotypic variance is due to genetic and/or environmental variation. Second, that, on average, MZ twins share all and DZ twins share half of their segregating genes. Third, that MZ and DZ twins share some environmental factors and not others. These three premises can be represented with the following equations:

$$\sigma^2_P = \sigma^2_A + \sigma^2_{NA} + \sigma^2_C + \sigma^2_E \qquad (1)$$

$$cov_{MZ} = \sigma^2_A + \sigma^2_{NA} + \sigma^2_C \qquad (2)$$

$$cov_{DZ} = \tfrac{1}{2}\sigma^2_A + k\sigma^2_{NA} + \sigma^2_C \qquad (3)$$

where k can hold a value between zero and ¼. Specically, when non-additive genetic effects are due solely to dominance, $k = ¼$, but as the number of genetic loci contributing to epistatsis increases k reduces to zero (Keller & Coventry, 2005). With five unknown parameters ($_{A, NA, C, E}$ and k) and only three observed statistics (i.e. σ^2_P, cov_{MZ}, cov_{MZ}) it is not possible to solve these equations, so the model is unidentified. Two parameters must be fixed. Typically it is assumed that non-additive effects are due to only dominance and epistatic effects are small or absent, thus $k = ¼$ (Evans, Gillespie, & Martin, 2002). As shared environmental effects will increase the covariation of DZ twins relative to MZ twins, where $cov_{DZ}/cov_{MZ} > ½$ then A, C, and E (an ACE model) is estimated. As non-additive genetic effects will decrease the covariation of DZ twins relative to MZ twins, where $cov_{DZ}/cov_{MZ} < ½$ then A, NA, and E (an ADE model) is estimated. Where an ACE model is selected, equations (1), (2), and (3) can be rearranged to give:

$$\sigma^2_A = 2(cov_{MZ} - cov_{DZ}) \qquad (4)$$

$$\sigma^2_C = 2\,cov_{DZ} - cov_{MZ} \qquad (5)$$

$$\sigma^2_E = \sigma^2_P - cov_{MZ} \qquad (6)$$

Alternatively, where an ADE model is selected, equation (5) is replaced with:

$$\sigma^2_{NA} = 2(cov_{MZ} - 2cov_{DZ}) \qquad (7)$$

The inability to estimate all sources of variance leads to bias in the estimates. Mathematically,

these expected biases are as follows (Keller & Coventry, 2005). In an ACE model, A will be overestimated, while NA and C will be underestimated. In an ADE model, the direction of bias in A and NA estimates depends on the extent of epistasis and shared environmental variance, and C will be underestimated. The biases, in practice, are apparent from more elaborate extended twin-family designs (Coventry & Keller, 2005). These suggest estimates of A in ACE models generally capture both additive and non-additive effects, and are thus reasonably representative of broad-sense heritability. By contrast, shared environment estimates are sometimes underestimated or overestimated in both ACE and ADE models, depending on the presence of other unmeasured effects, such as gene-environment correlations and assortative mating (as noted below). On average, at least, the shared environment effects are unbiased.

Several assumptions need to be met for these aforementioned premises to be true. The simplicity of the first premise assumes that *genetic and environmental influences are independent*. Genes, the environment, gene-environmental correlation, and gene-environment interactions can all influence phenotypic variation (Plomin, DeFries, & Loehlin, 1977). It is possible with specific measures of the environment to assess specific gene-environment correlations or gene-environment interactions. However, to the extent that gene-environment correlation and/or gene-environment interaction influences variation in the phenotype and is not specifically measured and modelled, estimates of genetic and environmental effects will be biased. The direction of bias will depend on whether the correlation or interaction is between genes and the shared environment or genes and the unique environment (Purcell, 2002).

For the second premise to be met there must be *no assortative mating*. To the extent that assortative mating is present and that genes influence that phenotype, DZ twins will be genetically more similar than 50% and the shared environment will be overestimated. Aside assortative mating, the first premise also requires that MZ twins share all genes that contribute to variation in the phenotype. Molecular genetic analyses have shown there is genetic variation between MZ twins, for example, though gene copy number variation (Bruder et al., 2008; Dear, 2009) and epigenetic mechanisms like DNA methylation (Kaminsky et al., 2009; note that these sources of genetic variation are present in all individuals, not just MZ twins). To the extent that MZ twins genetically vary in such a way that influences the phenotype of interest, genetic effects will be underestimated. The true extent of this bias is, to-date, unknown.

For the third premise to be true, the *equal environments assumption* requires MZ and DZ twins to correlate to the same degree on environmental factors that influence the phenotype of interest. To the extent that MZ twins share more similar environments that influence the phenotype, genetic effects will be overestimated. The twin method has been criticised for lack of

appropriate testing of the equal environments assumption (Richardson & Norgate, 2005). Yet recent analyses have found this assumption to hold for many phenotypes (Felson, 2014), and there is some evidence to suggest that it holds for educational outcomes (Conley, Rauscher, Dawes, Magnusson, & Siegal, 2013).

A final assumption, that relates more to interpretation than accurate partitioning of variance, is that the influences on twins are *generalizable* to the broader population. On average, twins have lower birth weight, shorter gestation, and greater likelihood of birth complications (Hall, 2003). Twins at age 2.5 years have less language development than singletons (Conway, Lytton, & Pysh, 1980). Although singletons have been found to outperform twins on IQ tests when young, their relative performance improves in early childhood (Myrianthopoulos, Nichols, & Broman, 1976). There has been no difference in academic performance in adolescence (Christensen et al., 2006), or in IQ in adults (Posthuma, De Geus, Bleichrodt, & Boomsma, 2000). Although lower birth weight has been associated with poorer academic performance, in a cohort of twins born in the 1980s Christensen et al. (2006) found low birth weight had a small effect and was only important when birth weight was low when compared to the average weight of the relevant group (i.e. twins or singletons).

Beyond these equations, computers allow us to partition the variance using more complicated models. Maximum likelihood estimation of twin covariances can be used to obtain estimates of genetic and environmental effects (Martin & Eaves, 1977). Maximum likelihood estimation essentially produces estimates that optimise the fit between observed data and a specified model (Boker et al., 2011). Optimisation is attained through successive iterations where estimates are adjusted. A likelihood statistic for the estimated model is obtained, which, under conditions of multivariate normality, is asymptotically distributed as a χ^2 distribution (Neale, Boker, Xie, & Maes, 2003). Various computer programs can be used in twin modelling, including OpenMx, Mplus, and LISREL. OpenMx, the program used for the analyses in this thesis, is robust to violations multivariate normality (Neale et al., 2003).

Contemporary computer programs used to estimate genetic and environmental effects provide several statistics to assess the fit of estimated parameters. For specific parameters, maximum likelihood based confidence intervals can be produced (Neale & Miller, 1997). Statistical significance of specific parameters can also be tested with nested models. Where one model is nested within another (i.e. one or more of the previously estimated parameters are fixed), the likelihood ratio test can be used to test if the nested model is a significantly poorer fit. The likelihood ratio test compares the difference in likelihoods from the nested models to a χ^2 distribution with degrees of freedom equal to the difference in number of estimated parameters

(Neale & Maes, 2004). Although not a statistical test, the comparative fit of models that are not nested can be assessed with Akaike's Information Criterion (AIC), with a lower AIC indicating a better fitting model (Akaike, 1987). Selection of the most appropriate model from which to report estimates of genetic and environmental influence needs to balance both parsimony and goodness-of-fit (Neale & Maes, 2004).

Aims and outline of the thesis

Broadly, the aim of this thesis is to explore the contribution of genes and the environment to individual differences in performance of Australian school students on literacy and numeracy. Previously, the International Longitudinal Twin Study (ILTS) has reported on the heritability of reading and reading-related skills on Australian children from their preschool year through to Grade 2 (Byrne et al., 2009; Byrne et al., 2002). Thus far, there are no studies on literacy and numeracy through the middle school years in Australia.

First and foremost, it is important to have confidence in the validity of the NAPLAN scales that I use throughout this thesis. Accordingly, Chapter 2 considers the validity of large-scale reading tests. One of the vociferous public complaints against the NAPLAN tests is that they are poor tests of reading. While it is clear they lack the ability to assess the component skills of reading, the purpose of the NAPLAN testing is not to diagnose specific reading problems. Rather the testing is an instrument with which to broadly assess reading ability. This first empirical paper used a criterion-based validity approach to consider the validity of large-scale reading tests. I used both phenotypic dyadic and behaviour genetic analyses to assess whether performance on large-scale reading tests converged with performance on well-accepted, individually administered tests of literacy skills.

Given this is the first behaviour genetic study to analyses these five domains of the NAPLAN: reading, spelling, grammar and punctuation, writing, and numeracy, Chapter 3 initially analyses the univariate estimates and tests for sex-limitation for each domain and grade. The second part of this chapter then assesses the source of covariation and independence among literacy and numeracy. The generalist gene hypothesis proposes that genes will mediate the covariation between academic domains and the environment will mediate their independence. While there is extensive support for genetic mediation of covariation, there is also evidence of genetic independence on reading and mathematical skills. To assess the overlap and independence in these Australian data, the genetic and environmental covariance structure across all five domains was tested with a sequence of independent pathway and a common pathway models.

Chapter 1

Genetic and environmental influences at a specific assessment point in time do not necessarily reflect their relative influence on variation in growth in performance. Chapter 4 assesses the influence of genes and the environment on stability and variation in growth in performance on NAPLAN tests. Genetic and environmental longitudinal correlations were used to assess stability, and based on studies in the United States of America (Betjemann et al., 2008; Wadsworth, Corley, Hewitt, & DeFries, 2001) and the United Kingdom (Harlaar, Dale, & Plomin, 2007; Kovas et al., 2007) genes were expected to substantially contribute to stability in performance over time. However, there are few studies on the etiology of variation in growth, and they have been conducted on reading skills. From these few studies, genes were generally important contributors to growth in reading while the shared environment was important in only some samples (Christopher et al., 2013a; Christopher et al., 2013b; Hart et al., 2013; Logan et al., 2013). The genetic influence on growth in reading in these studies belies the assumption that growth in performance is a direct consequence of teacher or school effects (Masters et al., 2008). This chapter extends this research on variation in growth to novel domains, namely grammar and punctuation, writing, and numeracy. Due to significant sex-limitation of genetic and environment variance components in some of the univariate analyses, a biometric latent growth curve model was adapted to include a test of sex-limitation.

Family socioeconomic status (SES), particularly in the United States of America, has been found to moderate the influence of genes and/or the environment on academic and cognitive ability (Rowe, Jacobson, & Van den Oord, 1999; Tucker-Drob, Rhemtulla, Harden, Turkheimer, & Fask, 2011; Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). Specifically, cognitive ability is more strongly influenced by the shared environment when SES is lower, and genes influence variation when SES is higher. This indicates a degree of social inequality that extends beyond the direct correlation between cognitive ability and SES, but this finding has not consistently been replicated internationally (Hanscombe et al., 2012; van der Sluis, Willemsen, de Geus, Boomsma, & Posthuma, 2008). Australia has legislated the intent that the quality of a student's education and opportunity to attain their potential should not be limited by where they live, their family wealth, or the school they attend. Chapter 5 first established the extent of the relationship between SES—both family and school—with literacy and numeracy performance in Australia. Second, this was extended to test if there was environmental constraint of genetic potential when family SES or school SES was less advantaged.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*: MIT press.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*(3), 317-332. doi:10.1007/BF02294359

Alamargot, D., & Fayol, M. (2009). Modelling the development of written composition. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE handbook of writing development*. London: SAGE Publications Ltd.

Anderson, J. R., & Fincham, J. M. (2014). Discovering the sequential structure of thought. *Cognitive Science, 38*(2), 322-352. doi:10.1111/cogs.12068

Anderson, J. R., Lee, H. S., & Fincham, J. M. (2014). Discovering the structure of mathematical problem solving. *Neuroimage, 97*, 163-177. doi:10.1016/j.neuroimage.2014.04.031

Athanasou, J. (2013, May 13). NAPLAN is fine but the way we use it is broken. *Sydney Morning Herald*. Retrieved from http://www.smh.com.au

Bantick, C. (2015, May 13). NAPLAN fails to show a child's progress in the education system. *The Courier-Mail*. Retrieved from http://www.couriermail.com.au

Baxter, A., & Farley, J. (1979). Mendel and meiosis. *Journal of the History of Biology, 12*(1), 137-173. doi:10.1007/BF00128138

Berninger, V., & Hart, T. (1992). A developmental neuropsychological perspective for reading and writing acquisition. *Educational Psychologist, 27*(4), 415-434. doi:10.1207/s15326985ep2704_2

Berninger, V., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing, 4*(3), 257-280. doi:10.1007/BF01027151

Betjemann, R., Willcutt, E., Olson, R., Keenan, J., DeFries, J., & Wadsworth, S. (2008). Word reading and reading comprehension: Stability, overlap and independence. *Reading and Writing, 21*(5), 539-558. doi:10.1007/s11145-007-9076-8

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306-317. doi:10.1007/s11336-010-9200-6

Bruder, C. E. G., Piotrowski, A., Gijsbers, A. A. C. J., Andersson, R., Erickson, S., Diaz de Ståhl, T., . . . Dumanski, J. P. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *The American Journal of Human Genetics, 82*(3), 763-771. doi:10.1016/j.ajhg.2007.12.011

Chapter 1

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology & Psychiatry, 46*(1), 3-18. doi:10.1111/j.1469-7610.2004.00374.x

Byrne, B. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle*. Hove, UK: Psychology Press.

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., . . . DeFries, J. C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to grade 2. *Journal of Neurolinguistics, 22*(3), 219-236. doi:10.1016/j.jneuroling.2008.09.003

Byrne, B., Delaland, C., Fielding-Barnsley, R., Quain, P., Samuelsson, S., Høien, T., . . . Olson, R. K. (2002). Longitudinal twin study of early reading development in three countries: Preliminary results. *Annals of Dyslexia, 52*(1), 47-73. doi:10.1007/s11881-002-0006-9

Chanquoy, L. (2009). Revision processes. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE Handbook of Writing Development.* (pp. 80-98). London: SAGE Publications Ltd.

Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2003). Schooling, literacy, numeracy and labour market success. *Economic Record, 79*(245), 165-181. doi:10.1111/1475-4932.t01-1-00096

Christensen, K., Petersen, I., Skytthe, A., Herskind, A. M., McGue, M., & Bingley, P. (2006). Comparison of academic performance of twins and singletons in adolescence: follow-up study. *BMJ, 333*(7578), 1095. doi:10.1136/bmj.38959.650903.7C

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., . . . Olson, R. K. (2013a). The genetic and environmental etiologies of individual differences in early reading growth in Australia, the United States, and Scandinavia. *Journal of Experimental Child Psychology, 115*(3), 453-467. doi:10.1016/j.jecp.2013.03.008

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., . . . Olson, R. K. (2013b). Modeling the etiology of individual differences in early reading development: Evidence for strong genetic influences. *Scientific Studies of Reading*, 1-19. doi:10.1080/10888438.2012.729119

Collins, J. W., & O'Brien, N. P. (2003). *The Greenwood dictionary of education*. Westport, Conn.: Greenwood Press.

Conley, D., Rauscher, E., Dawes, C., Magnusson, P. E., & Siegal, M. (2013). Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins. *Behavior Genetics, 43*(5), 415-426. doi:10.1007/s10519-013-9602-1

Conway, D., Lytton, H., & Pysh, F. (1980). Twin–singleton language differences. *Canadian Journal of Behavioural Science, 12*(3), 264-271. doi:10.1037/h0081061

Coulombe, S., Tremblay, J.-F., & Marchand, S. (2004). *Literacy scores, human capital and growth across fourteen OECD countries*. Statistics Canada: Ottawa.

Coulson, J. (2011, May 9). Why my kids won't sit the NAPLAN. *Sydney Morning Herald*. Retrieved from http://www.smh.com.au

Council of Australian Governments. (2013). *National Education Reform Agreement*. Retrieved from http://www.coag.gov.au/ode/501

Coventry, W. L., & Keller, M. C. (2005). Estimating the extent of parameter bias in the classical twin design: A comparison of parameter estimates from extended twin-family and classical twin designs. *Twin Research and Human Genetics, 8*(3), 214-223. doi:10.1375/twin.8.3.214

Dear, P. H. (2009). Copy-number variation: the end of the human genome? *Trends in Biotechnology, 27*(8), 448-454. doi:10.1016/j.tibtech.2009.05.003

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167-188. doi:10.1207/s1532799xssr0902_4

Evans, D. M., Gillespie, N. A., & Martin, N. G. (2002). Biometrical genetics. *Biological Psychology, 61*(1–2), 33-51. doi:10.1016/S0301-0511(02)00051-0

Felson, J. (2014). What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Social Science Research, 43*(0), 184-199. doi:10.1016/j.ssresearch.2013.10.004

Galton, F. (1875). The history of twins, as a criterion of the relative powers of nature and nurture. *Fraser's Magazine, 12*, 566-576.

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London, 45*(273-279), 135-145.

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*(4), 710-744,784-785. doi:10.17763/haer.81.4.t2k0m13756113566

Griffith, P. L., & Olson, M. W. (1992). Phonemic awareness helps beginning readers break the code. *The Reading Teacher, 45*(7), 516-523. Retrieved from http://www.jstor.org/stable/20200912

Hall, J. G. (2003). Twinning. *The Lancet, 362*(9385), 735-743. doi:10.1016/S0140-6736(03)14237-7

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PloS One, 7*(2), e30320. doi:10.1371/journal.pone.0030320

Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development, 78*(1), 116-131. doi:10.1111/j.1467-8624.2007.00988.x

Hart, S. A., Logan, J. A. R., Soden-Hensler, B., Kershaw, S., Taylor, J., & Schatschneider, C. (2013). Exploring how nature and nurture affect the development of reading: An analysis of the Florida Twin Project on Reading. *Developmental Psychology*. doi:10.1037/a0031348

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160.

Job, P. (2013, May 15). NAPLAN is driving our students backwards. *Sydney Morning Herald*. Retrieved from http://www.smh.com.au

Kaminsky, Z. A., Tang, T., Wang, S.-C., Ptak, C., Oh, G. H. T., Wong, A. H. C., . . . Petronis, A. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics, 41*(2), 240-245. doi:10.1038/ng.286

Keller, M. C., & Coventry, W. L. (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics, 8*(3), 201-213. doi:10.1375/twin.8.3.201

Kovas, Y., Haworth, C. M. A., Dale, P. S., Plomin, R., Weinberg, R. A., Thomson, J. M., & Fischer, K. W. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development, 72*(3), i-156. doi:10.2307/30163178

Kress, G. R. (2003). *Literacy in the new media age*. London: Routledge.

Logan, J. A. R., Hart, S. A., Cutting, L., Deater-Deckard, K., Schatschneider, C., & Petrill, S. (2013). Reading development in young children: Genetic and environmental influences. *Child Development, 84*(6), 2131-2144. doi:10.1111/cdev.12104

Martin, N., & Eaves, L. (1977). The genetical analysis of covariance structure. *Heredity, 38*(1), 79-95. doi:10.1038/hdy.1977.9

Masters, G., Rowley, G., Ainley, J., & Khoo, S. (2008). *Reporting and comparing school performances*. Retrieved from http://research.acer.edu.au/ar_misc/8

Mendel, G. (1866). Experiments in plant hybridization (1865). *Proceedings of the Natural History Society of Brünn, 4*.

Ministerial Council on Education, E., Training, and Youth Affairs,. (2008). *Melbourne Declaration on Educational Goals for Young Australians*. Retrieved from http://eric.ed.gov/?id=ED534449

Ministerial Council on Education, E., Training, and Youth Affairs,. (2009). *MCEETYA four-year plan 2009 – 2012: A companion document for the Melbourne Declaration on Educational Goals for Young Australians*. Retrieved from http://eric.ed.gov/?id=ED534447

Moss, B. J. (2009). Writing in the Wider Community. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE handbook of writing development* (pp. 561-574). London: SAGE Publications Ltd.

Myhill, D. (2008). Towards a linguistic model of sentence development in writing. *Language and Education, 22*(5), 271-288. doi:10.1080/09500780802152655

Myrianthopoulos, N. C., Nichols, P. L., & Broman, S. H. (1976). Intellectual development of twins—comparison with singletons. *Acta geneticae medicae et gemellologiae: twin research, 25*(01), 376-380. doi:10.1017/S0001566000014458

Neale, M. C., Boker, S., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry.

Neale, M. C., & Maes, H. H. (2004). *Methodology for the genetic studies of twins and families* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27*(2), 113-120. doi:10.1023/a:1025681223921

Organisation for Economic Co-operation and Development. (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. Retrieved from http://dx.doi.org/10.1787/9789264204256-en

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184-202. doi:10.2307/4151679

Parveva, T., De Coster, I., & Noorani, S. (2009). *National testing of pupils in europe: Objectives, organisation and use of results* (9292010360). Retrieved from http://www.eurydice.org

Pearson, K., & Lee, A. (1903). On the laws of inheritance in man I. Inheritance of Physical Characters. *Biometrika, 2*(4), 357-462.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84*(2), 309-322. doi:10.1037/0033-2909.84.2.309

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher, 40*(3), 103-116. doi:10.3102/0013189x11405038

Chapter 1

Posthuma, D., De Geus, E. J., Bleichrodt, N., & Boomsma, D. I. (2000). Twin–singleton differences in intelligence? *Twin Research and Human Genetics, 3*(02), 83-87. doi:doi:10.1375/twin.3.2.83

Purcell, S. (2002). Variance components models for gene x environment interaction in twin analysis. *Twin Research, 5*(6), 554-571. doi:10.1375/136905202762342026

Richardson, K., & Norgate, S. (2005). The equal environments assumption of classical twin studies may not hold. *British Journal of Educational Psychology, 75*(3), 339-350. doi:10.1348/000709904X24690

Rowe, D. C., Jacobson, K. C., & Van den Oord, E. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70*(5), 1151-1162. Retrieved from <Go to ISI>://WOS:000083197900008

Senate Standing Committee on Education and Employment. (2014). *Effectiveness of the National Assessment Program – Literacy and Numeracy*. Canberra: Senate Printing Unit Retrieved from http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Employment/Naplan13/Report/index.

Street, B. V. (2009). Multiple literacies and multi-literacies. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE handbook of writing development* (pp. 137-151). London: SAGE Publications Ltd.

Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E., & Fask, D. (2011). Emergence of a gene x socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychological Science, 22*(1), 125-133. doi:10.1177/0956797610392926

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*(6), 623-628. doi:10.1046/j.0956-7976.2003.psci_1475.x

UNESCO Institute for Statistics. (2008). *International literacy statistics: A review of concepts, methodology and current data*

van der Sluis, S., Willemsen, G., de Geus, E. C., Boomsma, D., & Posthuma, D. (2008). Gene-environment interaction in adults' IQ scores: Measures of past and present environment. *Behavior Genetics, 38*(4), 348-360. doi:10.1007/s10519-008-9212-5

Wadsworth, S. J., Corley, R. P., Hewitt, J. K., & DeFries, J. C. (2001). Stability of genetic and environmental influences on reading performance at 7, 12, and 16 years of age in the Colorado Adoption Project. *Behavior Genetics, 31*(4), 353-359. doi:10.1023/a:1012218301437

Chapter 2

Validity of large-scale reading tests: A phenotypic and behaviour-genetic analysis

Katrina L Grasby, Brian Byrne, and Richard K Olson

Candidate Signature

Principal Supervisor Signature

Chapter 2

Higher Degree Research Thesis by Publication
University of New England

STATEMENT OF AUTHORS' CONTRIBUTION

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research PhD candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated in the *Statement of Originality*.

	Author's Name (please print clearly)	Contribution
Candidate	Katrina Grasby	Responsible for study design; data analysis; writing of paper
Other Authors	Brian Byrne	Provided editorial comments; author of ILTS grant, which provided data for paper
	Richard Olson	Author of ILTS grant, which provided data for paper

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry

28/1/16
Date

>8/1/16
Date

35

Chapter 2

STATEMENT OF ORIGINALITY

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research Master/PhD candidate and the candidate's Principal Supervisor, certify that the following text, figures and diagrams are the candidate's original work.

Type of work	Page number/s
All text, tables, and figures	Chapter 2, pp27–43

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry
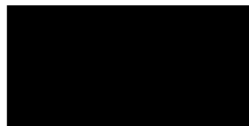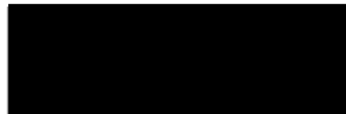
28/1/16
Date

28/1/16
Date

Chapter 3
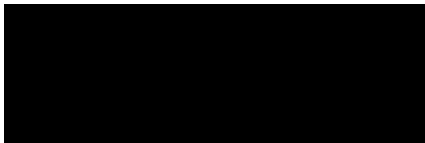
Genetic and Environmental Influences on Literacy and Numeracy Performance in Australian School Children in Grades 3, 5, 7, 9

Katrina L. Grasby, William L. Coventry, Brian Byrne, Richard K. Olson, and Sarah E. Medland

Candidate Signature

Principal Supervisor Signature

Chapter 3

The past decade has seen a burgeoning of research into the behaviour genetics of the key educational domains of reading and mathematics. Most of this has been conducted on several large twin projects in Western countries. These projects have targeted a variety of ages and abilities, have used a mix of assessment tools, and have shown that after a few years of schooling genes substantially and significantly influence both reading and mathematical ability. Each year in Australia, children in Grades 3, 5, 7, and 9 sit National Assessment Program in Numeracy and Literacy (NAPLAN) tests in reading, spelling, grammar and punctuation, writing, and mathematics. These tests were designed by educational authorities to objectively benchmark performance on state and territory curricula in English and Mathematics (Senate Standing Committee on Education and Employment, 2014). Since 2010, school results have been readily accessible through the *MySchool* website, and increasingly they are considered to be "high-stakes" tests. In this paper we will assess the heritability of performance in these tests and the degree to which genetic and environmental factors overlap or independently influence performance across the test domains in each grade. This is the first behaviour-genetic study on Australian children to target this transition from primary to high school, and one of few behaviour genetic investigations into writing ability.

Reading, Writing, Spelling, Grammar and Punctuation

Heritability estimates are influenced by genetic and environmental variation in the sample. This means the estimates from one country are not necessarily similar to the estimates from another. It is conceivable the environmental variation, especially with regards to education practices, might differ considerably from one country to another (e.g. Samuelsson et al., 2008). Given this, it is quite remarkable how similar heritability estimates can be. Take reading, for instance. After a year of formal instruction in reading, genes consistently and substantially influence performance on an extensive range of reading measures, including word reading, phonological decoding, reading fluency, and reading comprehension (Harlaar, Dale, & Plomin, 2005; Petrill et al., 2007; Samuelsson et al., 2008; Taylor & Schatschneider, 2010). In these studies, genes have typically explained more than 50% of variation in performance. These findings are reliable regardless of whether reading skills are assessed by a teacher (e.g. Harlaar, Dale, et al., 2005), a trained test administrator (e.g. Petrill et al., 2007), or via a phone or internet-based test (e.g. Haworth et al., 2009). A recent meta-analysis found no significant heterogeneity across 11 studies on reading skills, with an estimated heritability of .73 (de Zeeuw, de Geus, & Boomsma, 2015). Although heterogeneity was noted among studies on reading comprehension, only six studies were included in this category and a single UK study with a low heritability

estimate of .38 appears to be an exception to otherwise high heritability. Although we acknowledge the occasional exceptions, high heritability estimates—from a wide range of reading measures from populations in the USA, Scandinavia, Australia, and the UK—are evidence of the stability and extent to which genes are important contributors to variation in reading performance in the Western world.

Where studies from Western countries tend to differ is in the relative influence of the shared environment. In the UK the shared environment has consistently been found to have a modest influence on variation in reading, typically 15–20% (Harlaar, Dale, & Plomin, 2007; Haworth et al., 2009; Kovas, Harlaar, Petrill, & Plomin, 2005). In contrast, studies in Australia, thus far, have found the shared environment to have negligible impact (Byrne et al., 2007; Samuelsson et al., 2008). Meanwhile, results from the USA range more widely and span the findings in Australia and the UK (Friend, DeFries, & Olson, 2008; Hart, Petrill, & Thompson, 2010; Taylor & Schatschneider, 2010). The Twin Early Development Study (TEDS) in the UK is by far the largest twin study (Trouton, Spinath, & Plomin, 2002); as such, one theory is that it might be capturing a greater variety of family environments, especially when compared to the work in Australia. The Australian results come from the International Longitudinal Twin Study (ILTS), which drew a sample of twins from the Sydney metropolitan area (Byrne et al., 2002). The current study, by contrast, includes Australian twins from all states and both metropolitan and rural regions. Consequently, we expect a wider variety of environments to be captured in this sample and we might find a greater influence from the shared environment on reading compared to the ILTS.

Although the heritability of reading in children after Grade 3 has not been examined in Australia, we expect our measures of reading and spelling to be in line with the high heritability estimates of research in the USA and UK. However, there are no behaviour-genetic studies testing grammar and punctuation in the age range we are examining in this study. The ILTS assessed grammar in preschool children and at the end of Kindergarten and found grammar to be more influenced by the shared environment (.40) than genes (.21) (Byrne et al., 2005), but from the ILTS we also know that heritability estimates for reading skills can change markedly in the early years of formal education (Samuelsson et al., 2008). While investigations into the heritability of grammar are scarce, there are no studies, thus far, assessing that of punctuation. These data from the NAPLAN, which combines grammar and punctuation, will be the first of their kind to be assessed for heritability.

There are also few behaviour-genetic studies on writing. In TEDS, writing ability was assessed in 7-year-old twins by teachers rating ability against three descriptors of writing skills,

with options to rank students as not meeting Level 1 or exceeding Level 3 (Oliver, Dale, & Plomin, 2007). Writing was substantially heritable (.66) with negligible contribution of the shared environment (.07). In the USA, older participants, aged from 8 to 18 years, were assessed on several writing skills through the Colorado Learning and Disabilities Research Centre (CLDRC; Olson et al., 2013). Three measures of writing were assessed, one was a copying task and two required the participant to write a grammatically correct sentence from prompts. One of the sentence-writing tasks was timed and one untimed. They found a comparatively low heritability for the timed task (.33), compared to the untimed task (.66) and the copying task (.77). The writing measure in our study is akin to the untimed task employed in the Colorado study, but requires the student to write a substantial passage of several paragraphs in exam conditions. Based on these two studies, we expect genes will be important for performance in the writing task in our study.

Numeracy and Mathematics

In the UK the influence of genes and the environment on individual differences in mathematics is similar to the findings from studies on reading. The heritability of mathematical ability in 7- and 9-year-old twins was estimated to be .66 (Oliver et al., 2004) and .68 respectively (Haworth, Kovas, Petrill, & Plomin, 2007). In slightly older twins, the heritability of various mathematical skills was noticeably lower, ranging from .32–.45 at age 10 (Kovas, Haworth, Petrill, & Plomin, 2007) and .49 at age 12 (Haworth et al., 2009). In the younger twins mathematical ability was measured with teacher assessment but from age 10 this changed to a web-administered test, so this apparent decrease in heritability might not reflect age-specific differences but might reflect a change in test delivery. This estimate of approximately 50% appears to be quite stable, with heritability estimated at .55 on a nationwide school test at age 16 (Shakeshaft et al., 2013). Like the studies on reading, the influence of the shared environment is more modest than genes, estimated at .09 at age 9 (Haworth et al., 2007), .07–.23 at age 10 (Kovas et al., 2007), .19 at age 12 (Haworth et al., 2009), and .26 at age 16 (Shakeshaft et al., 2013). There is, perhaps, a slight increase in influence with age; although, this might reflect different measures at different ages. These studies show genes to be an important contributor to variation in mathematical ability throughout the school years with the shared environment having a more moderate contribution, at least in the UK.

In contrast to this consistent influence of genetic factors on mathematical achievement in the UK, estimates from the USA are wide-ranging. As part of the Western Reserve Reading and Math Project (WRRMP) in Ohio, the heritability of a variety of mathematical skills at age 8 was

negligible (0–.14; except for a test of mathematical fluency, which was .63; Hart, Petrill, Thompson, & Plomin, 2009). Although, subsequent analyses at age 10 showed heritability of mathematical skills to range from .34–.49 (Hart et al., 2010), and when mathematical ability was modelled as a latent variable from various mathematical skills, heritability estimates at age 10 was .41 and at age 11 was .34 (Petrill et al., 2012). These heritability estimates from ten- and eleven-year-olds in the WRRMP are close to those obtained from mathematical skills in the UK. However, markedly higher estimates come from the CLDRC where, using similar mathematical tests to the WRRMP and modelling mathematics as a latent variable, heritability estimates of .67 (Knopik & DeFries, 1999) and .90 (Alarcón, Knopik, & DeFries, 2000) have been reported. Although the CLDRC project have a sample of twins identified as having at least one twin with learning difficulties, these estimates are obtained from their control sample of twins without identified learning difficulties. Therefore, these higher estimates do not reflect differential heritability due to ability. The differences might result from different sampling techniques; the Colorado study included participants ranging in age from 8 to 20 years, while the WRRMP and TEDS project in the UK obtained estimates from participants similar in age. These studies also differ in estimates of the shared environment; while estimates from Colorado are similar to those from the UK (.21; Knopik & DeFries, 1999), estimates from the WRRMP were a more substantial. Shared environmental estimates of mathematical skills ranged from .15–.52 at age 9 (Hart et al., 2009) and .32–.46 at age 10 (Hart et al., 2010) and was .52 at both age 10 and 11 when mathematical ability was modelled as a latent trait (Petrill et al., 2012). The overall pattern from studies in mathematics suggest that genes are important in predicting variation in performance, with the shared environment having a somewhat less consistent influence across state and country samples. These results might reflect different educational environments and curricula between these states and countries.

Our study will assess participants from the middle of primary school through to the middle of high school. In Australia, the progression from primary to high school marks a transition where students, typically, change from having a single classroom teacher to having a different teacher, who is a specialist, in each subject. Grade 7 marks the beginning of high school for most states and territories in Australia, and in all states and territories from Grade 7 students take both a calculator and a non-calculator numeracy test. There is evidence that a teacher's pedagogical content knowledge in mathematics accounts for some of the variance in students' mathematical achievement (Baumert et al., 2010; H. C. Hill et al., 2008). In Australia, concerns have been raised around a high proportion of teachers who end up teaching "out-of-field", especially in lower secondary mathematics (Hobbs, 2015). This possibly results in increased diversity of

teaching expertise coinciding with mathematical content becoming more complex. Consequently, we might find a greater influence of the environment in high school compared to primary school. Although there are not similar concerns regarding teaching out-of-field in English, the increasing complexity of the comprehension and writing assessments might also be impacted by specialist teachers contributing to greater environmental variance in ways that meaningfully impact performance in the literacy assessments.

Sex Differences

Although a mean difference between girls and boys is widely reported in the educational literature for reading, writing, and mathematics (e.g. Hedges & Nowell, 1995; Hyde, 2005), it does not necessarily follow that the influences on variation in performance are different. Where different influences are present between the sexes, they might have a qualitative origin, or be quantitative in nature, or a combination of both. *Qualitative sex-effects* are present when the genetic and environmental factors that influence females and males differ; the presence of qualitative sex-effects are indicated when the genetic correlation between opposite-sex twins is significantly different to .5, which is the average portion of additive genetic effects shared between siblings who are not monozygotic twins. *Quantitative sex-effects* are present when the magnitude of genetic and environmental influence is different between the sexes; they are indicated when the genetic correlation between opposite-sex twins can be fixed to .5 but the genetic and environmental path estimates cannot be equated between females and males (Neale & Maes, 2004).

Thus far, results from studies on sex differences for literacy or mathematics are diverse and sex-effects found have not been replicated with any consistency. There is support for qualitative sex-effects in reading at age 7 (Harlaar, Spinath, Dale, & Plomin, 2005). Quantitative sex-effects have also been found. Greater heritability has been indicated in girls for reading but not mathematics (Petrill & Thompson, 1994), which contrasts with greater heritability in boys for reading and mathematics (Knopik, Alarcón, & DeFries, 1998; Shakeshaft et al., 2013). Other studies have found the relative influence of genes and the environment to be equivalent across sex, but the total variance was greater in boys in both reading and mathematics (Reynolds et al., 1996; Spinath, Spinath, & Plomin, 2008). Finally, there are numerous studies that find no significant sex-effects in various measures of reading and mathematics (e.g. Davis et al., 2008; Hart, Soden, Johnson, Schatschneider, & Taylor, 2013; Taylor & Schatschneider, 2010; Wadsworth & DeFries, 2005).

Although there is disparity in the findings, where sex-effects have been observed the effect sizes are generally small. This is consistent with the broader reality of sex differences, where mean differences are also of a small or negligible effect size (Hyde, 2007; Limbrick, Wheldall, & Madelaine, 2010). For our Australian study, we expect to find mean differences comparable with those reported in the national reports on the tests we have used (Australian Curriculum Assessment and Reporting Authority, 2008, 2009, 2010a, 2011, 2012, 2013, 2014a). In the national reports boys tend to have larger standard deviations across subjects and grades, which is consistent with greater variance in boys on tests of reading and mathematics in other countries (Machin & Pekkarinen, 2008). This difference in variance indicates that there might be sex-effects in these Australian data. If present, we will test if the origin of these sex-effects are genetic, environmental, or both.

Generalist Genes

A central proposal that has emerged from the extensive TEDS project has been the generalist genes hypothesis (Plomin & Kovas, 2005). It is based on pleiotropy (each gene affects multiple traits) and polygenicity (multiple genes affect each trait), such that many of the same genes influence various academic domains (Plomin, Kovas, & Haworth, 2007). This hypothesis has been extended to suggest that pleiotropic genes predominantly mediate the covariation of ability while unique environments mediate differences in ability (Kovas & Plomin, 2007). There is extensive support for genetic overlap across different domains of school achievement; high genetic correlations are invariably detected regardless of age, country, and subject (e.g. Davis, Haworth, & Plomin, 2009; Harlaar, Kovas, Dale, Petrill, & Plomin, 2012; Hart et al., 2010; Knopik & DeFries, 1999; Kovas et al., 2005; Oliver et al., 2007; Plomin et al., 2007). Evidence for the specificity of the unique environment is sparse. When unique environment correlations are low, they have typically not been modelled to exclude measurement error, and when heritability is modelled using latent variables unique environmental correlations have been substantial. Knopik et al. (1999) found a unique environment correlation of .50 between reading and mathematics. This was replicated by Davis et al. (2009) with a unique environmental correlation of .59 between reading and mathematics, and Harlaar et al. (2012) with a correlation of .71 between mathematics and reading comprehension. In a latent factor model, Hart et al. (2010) found no specific unique environmental contribution to a number of mathematical skills over-and-above those shared with reading. In our study, we use single measures in each achievement domain, so we expect that genes will predominantly mediate covariation between these domains, and that the unique environmental correlations will be small. As such we will be able to assess whether genes are

generalists, but not whether unique environments are specialists. In line with findings from the UK and USA we also hypothesise some unique genetic influences on mathematics and reading ability.

Method

Participants

Twins and triplets born from 1993–2006 were recruited through the voluntary Australian Twin Registry. For the 34 sets of triplets, a random pair from each set was selected for the analyses, and from hereon all multiple births are referred to as *twins*. Twins were invited to participate if they had sat (or would sit by 2014) a NAPLAN test. Of the 6853 families contacted, 2272 (33%) consented to participate with 2226 (32%) also providing zygosity information. For 1940 (28%) we successfully obtained NAPLAN results from state departments.

Zygosity was determined with a short questionnaire (Lykken, Bouchard, McGue, & Tellegen, 1990). For 896 twins we had both questionnaire responses and parent reports of DNA results. The questionnaire correctly classified 94.6% of them as either monozygotic (MZ) or dizygotic (DZ). The misclassifications of MZ and DZ were approximately equal (54% DZ). Where a phenotype is heritable, misclassified twins will have the effect of inflating estimates of the shared environment and decreasing estimates of additive genetic effects. The classification accuracy in our subsample is in line with, or higher than, other questionnaires used to determine zygosity (Jackson, Snieder, Davis, & Treiber, 2001). Our final sample by gender and zygosity comprised 455 female MZ pairs, 303 female DZ pairs, 412 male MZ pairs, 287 male DZ pairs, and 483 opposite-sex DZ pairs. The number of twin pairs at Grades 3, 5, 7, and 9 respectively, with percentage of overlap with at least an adjacent grade indicated in parentheses, was 1184 (66%), 1113 (98%), 1001(98%), and 819 (81%). At the time of testing the average ages were 8.6 years at Grade 3, 10.6 years at Grade 5, 12.5 years at Grade 7, and 14.5 years at Grade 9.

Materials

National Assessment Program in Literacy and Numeracy

The NAPLAN is a nationwide, standardised assessment introduced in 2008 in Australia. Each year, students in Grades 3, 5, 7, and 9 sit tests in reading, writing, language conventions, and numeracy. The test content is based on the "Statements of Learning for English" and the "Statements of Learning for Mathematics," which inform state and territory curricula. For each achievement domain students are given a score on a common scale from 0–1000. This common scale spans all years of the test and was designed to measure growth within cohorts and to

compare across cohorts. Technical information and test administration details were obtained from the Australian Curriculum Assessment and Reporting Authority (Australian Curriculum Assessment and Reporting Authority, 2010b, 2014b; 2015; R. Randall, personal communication, July 10, 2013). Example test papers and writing prompts are available at www.nap.edu.au.

Reading

The reading test is a comprehension test composed of 7–8 passages. The passages were extracts or adaptations from books, newspaper articles, posters or poems. Passage length varies from brief single paragraphs of about 100 words, to several paragraphs of about 450 words. There were 5–8 items relating to any given passage. Most items are multiple-choice format, with one or two short answer questions in each test. For Grades 3 and 5 there were 35–38 items to be completed in 45–50 minutes, and for Grades 7 and 9 there were 45–50 items to be completed in 65 minutes. A Cronbach's alpha of .85 (or above) for each test in each year indicates a high internal reliability.

Spelling

The spelling test presents misspelt words in simple sentences and requires students to identify and correct the spelling errors. For Grades 3 and 5 there were 23–25 items, and for Grades 7 and 9 there were 25–30 items. The spelling test is administered in the same paper as the grammar and punctuation test, and students are given 40–45 minutes to complete both of these question sets. For the spelling test, a Cronbach's alpha of .90 (or above) for each test in each year indicates a high internal reliability.

Grammar and punctuation

The grammar questions ask students to choose the correct word(s) to complete a sentence. This form of question is used in early grades to identify correct tense, pronouns, conjunctions, and verb forms. In later grades relative pronouns, clauses, and comparative adjectives are also assessed. The punctuation questions ask students to insert or identify punctuation marks at the correct location in a sentence. For all grades there were 23–28 items. A Cronbach's alpha of .71 to .87 for each test in each year indicates an acceptable internal reliability (average .80).

Writing

The writing test is composed of a writing stimulus, which provides an idea or topic, and students are asked to write a response in a specified writing style (i.e. narrative, informative, or persuasive). For example, "It is cruel to keep animals in cages. What do you think? Do you agree or disagree? Perhaps you can think of ideas for both sides of this topic." The same prompt and style is used for all grade levels in a given year. Students have 40 minutes of writing time. Marks

are awarded on 10 criteria: audience, text structure, ideas, vocabulary, cohesion, paragraphing, sentence structure, punctuation, spelling, and the final criterion depended on the writing style specified. For persuasive writing the criterion was persuasive devices (2011–2014), and for narrative writing the criterion was character and setting (2008–2010). From 2008–2010 the maximum score was 47, from 2011–2014 the maximum score was 48. A Cronbach's alpha, calculated using pooled data from all grades, of .93 (or above) for each test year indicates a high internal reliability; unfortunately we were unable to obtain inter-rater reliability information.

Numeracy

The numeracy test assesses five aspects of mathematics. *Working mathematically* includes problem solving, reasoning and interpretation. *Number* includes counting and computation. *Algebra, function and pattern* includes working with functions and relationships, graphs, equations, and rules. *Measurement, chance and data* includes working with units, likelihood and inference. *Space* includes shape and location. Most items are multiple-choice format, with a few short answer questions in each test. For Grade 7 and Grade 9 the students sit a calculator-allowed and a non-calculator numeracy test. For Grades 3 and 5 there were 35–40 items to be completed in 45–50 minutes. For Grades 7 and 9 there were 62–64 items from the combined calculator and non-calculator papers, with each paper to be completed in 40 minutes. A Cronbach's alpha of .84 (or above) for each test in each year indicates a high internal reliability.

Procedure

After receiving parental consent, the state and territory departments of education provided NAPLAN test results. The NAPLAN tests are administered in the morning over three consecutive days each year in the second full week of May (approximately 3.5 months into the school year). On the first day the language conventions test (comprising of the spelling and grammar and punctuation domains) is administered and, after a minimum 20 min break, is followed by the writing test. On the second day the reading test is administered. On the third day the numeracy tests are administered; for Grades 7 and 9 the first test permits use of a calculator and the second test does not. Support within specific constraints can be provided for students with disability, such as scribing or reading questions in the numeracy test. Across the nation 96% of students participate in the tests.

Analyses

Raw data were fitted to structural equation models to obtain parameter estimates, confidence intervals, and fit indices. Models were estimated using full information maximum likelihood in OpenMx (Boker et al., 2011). To test assumptions of equal means and variances

across different zygosity groups, we ran a series of nested models. Beginning with a saturated model that included age, age-squared, age-by-sex, age-squared-by-sex, and cohort effects as covariates on the means, parameters were equated or covariates dropped and the likelihood ratio test (LRT) used to assess if the loss of fit was significant. The LRT compares the difference in log likelihood from the nested models to a χ^2 distribution with degrees of freedom equal to the difference in estimated parameters from the nested models (Neale & Maes, 2004). Given the large number of tests—20 phenotypes—it was probable that some assumptions would be violated by chance. This was found to be the case when after randomly allocating twins to Twin 1 and Twin 2 their means could not be equated with an alpha of .05; therefore we used an alpha of .01, which is consistent with the procedure employed by Evans, Frazer, and Martin (1999). Where covariates could not be dropped without a significant loss of fit they were included in subsequent analyses.

We also ran a series of sex-limitation models to test if means, variance, and covariances could be pooled across female and male twins. The *full sex-limitation* model allows for both qualitative and quantitative sex-effects by allowing genetic and environmental parameter estimates to differ between females and males and either the genetic correlation (r_A) between opposite-sex DZ twins to differ from .5 or the shared environment correlation (r_C) to differ from 1 (see Figure 3.1). When twins are raised together, the difference in same-sex DZ and opposite-sex DZ correlations informs the calculation of both r_A and r_C, thus only one can be estimated in any given model. As these models are not nested, Akaike's Information Criterion (AIC) was compared across models, with a lower AIC indicating a better-fitting model (Akaike, 1987). After comparing both r_A and r_C models, a series of nested models were compared to the full r_A sex-limitation model using the LRT (Neale & Maes, 2004). Fixing r_A to .5 tested a *common-effects* model that does not allow qualitative sex difference but does allow the magnitude of genetic and environmental influences to differ for females and males. Fixing the male paths a_m, c_m, and e_m to be a scalar multiple of the respective female paths a_f, c_f, and e_f, tested a *scalar* model. Finally, equating female and male path estimates and fixing the scalar to 1 tested a *null* model with no sex-effects. Parameter estimates are reported for the most parsimonious model that did not result in a significant loss of fit from the full sex-limitation model.
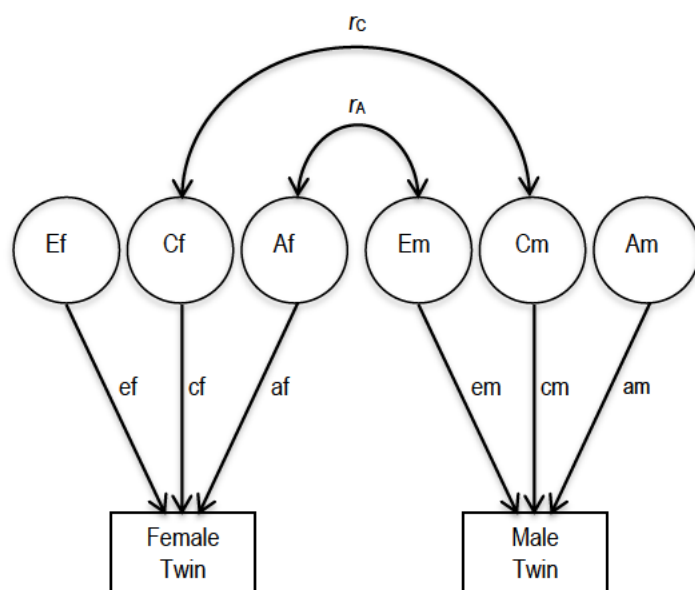
Figure 3.1 Path diagram of full sex-limitation model depicting opposite-sex twin pair. Either the shared environment correlation (r_C) or the genetic correlation (r_A) is estimated. Additive genetic (A), shared environment (C), and unique environment (E) variance is estimated separately for females (f) and males (m). Paths af, cf, ef, am, cf, ef = A, C and E effects on the trait for females and males respectively.

For multivariate modelling, initially a nonscalar sex-limitation correlated factors model was fit to the data to test for significant multivariate sex-effects. This is a saturated model where *a*, *c*, and *e* paths are free to differ by sex and all correlation paths are estimated (Neale, Røysamb, & Jacobson, 2006). To test for equivalence between the sexes across multiple domains, path estimates and correlations were equated for females and males and the LRT was used to compare the constrained model to the saturated model. Where there was no significant loss of fit, females and males were combined for further multivariate models. To test the structure of A, C, and E we fitted a sequence of independent pathway models and a common pathway model. The first independent pathway model contained two additive genetic common factors (A1, A2), two shared environmental common factors (C1, C2), two unique environmental common factors (E1, E2), and specific factors (Asp, Csp, Esp). The first common factor was loaded onto all five domains and represented genetic or environmental factors that influence performance on all NAPLAN tests. The second factor was loaded onto only the literacy variables; these second common paths were systematically dropped and the LRT used to test if there were literacy-specific influences on test performance. The common pathway model was even more stringent, and tested if a single latent factor explained the covariance among the achievement domains. More parsimonious models that provided no loss of fit from the saturated model were selected and the parameter estimates from these models are reported.

Chapter 3

Results

Assumption Testing

Distributions for each domain in each grade were normal. Scores were truncated at $\pm$ 3 standard deviations from the mean, resulting in less than 1% of scores dropped in any given domain and grade. Three multivariate outliers were identified and removed, one twin pair from each of Grade 3 reading, Grade 7 reading, and Grade 9 writing. Age was a significant covariate for Grades 3 and 5, except for Grade 5 grammar and punctuation. In addition to age, age-squared was a significant covariate for spelling in Grades 3 and 5. For the latter years age ceased to be significant, but in Grade 9 age-by-sex was important for spelling and age-squared was important for writing. Cohort effects were present in writing from Grade 5 onwards. Sex significantly influenced means, except for reading in Grades 7 and 9, grammar and punctuation in Grade 7 and spelling in Grade 9. (Model fit statistics from assumption testing are reported in the Appendix, see Tables A1–A5.) To keep models parsimonious, covariates were included only when they had a significant effect on the means.

Table 3.1   Descriptive statistics by sex for NAPLAN domains at each grade

Variable	Females			Males			Cohen's d
	M	SD	n	M	SD	n	
Reading							
Grade 3	452	84	1207	440	89	1127	0.14
Grade 5	526	77	1091	521	76	1088	0.07
Grade 7	580	66	999	576	67	958	0.05
Grade 9	617	64	831	619	65	766	-0.03
Spelling							
Grade 3	432	72	1209	417	78	1129	0.21
Grade 5	512	65	1094	498	70	1087	0.20
Grade 7	568	59	999	556	65	951	0.19
Grade 9	606	63	831	599	65	767	0.11
Grammar & Punctuation							
Grade 3	460	88	1203	441	89	1121	0.22
Grade 5	534	82	1091	519	82	1085	0.18
Grade 7	576	72	1001	568	77	957	0.11
Grade 9	612	70	826	604	72	767	0.12
Writing							
Grade 3	444	56	1206	422	60	1120	0.38
Grade 5	512	61	1089	493	62	1078	0.31
Grade 7	563	66	999	539	71	952	0.35
Grade 9	610	75	827	584	78	765	0.33
Numeracy							
Grade 3	416	68	1205	426	75	1118	-0.15
Grade 5	500	66	1093	519	74	1086	-0.27
Grade 7	565	65	1001	586	76	952	-0.30
Grade 9	609	65	822	634	71	756	-0.36

Note. MZ = monozygotic. DZ = dizygotic. *n* = individuals. Cohen's *d*

calculated as females minus males; positive values indicate a higher score

in females and negative values indicate a higher score in males.

Chapter 3

Univariate Analyses

The mean scores were higher in females than males on the four literacy domains of reading, spelling, grammar and punctuation, and writing, while males scored higher on the numeracy domain (see Table 3.1). As indicated by Cohen's *d*, the differences are small in effect. For reading, spelling, grammar and punctuation sex differences tended to reduce as age increased. Writing showed consistent differences, with girls scoring about one third of a standard deviation higher than boys. For numeracy, the mean differences increased over time, with boys scoring about one third of a standard deviation higher than girls in Grade 9.

Intraclass correlations were substantially larger for MZ twins compared to DZ twins for achievement in all five domains at each grade level for both sexes, indicating genetic influences in all domains (see Table 3.2). The DZ correlations were generally greater than half the MZ correlations, indicating some shared environmental effects. Some opposite-sex DZ correlations were lower than the same-sex DZ correlations and sex-effects were explicitly modelled and tested for each domain and grade (reported below).

Table 3.2   Intraclass correlations by zygosity and sex for NAPLAN domains at each grade

| | Intraclass Correlations | | | | | | | | | |
| | MZF | | MZM | | DZF | | DZM | | DZOS | |
Variable	r	n	r	n	r	n	r	n	r	n
Reading										
Grade 3	.75	262	.75	245	.42	181	.44	162	.43	303
Grade 5	.68	228	.68	221	.54	166	.46	170	.30	284
Grade 7	.70	210	.72	201	.52	164	.48	153	.38	230
Grade 9	.72	195	.73	169	.46	123	.46	120	.43	168
Spelling										
Grade 3	.78	262	.76	245	.45	180	.48	163	.36	305
Grade 5	.78	230	.76	222	.40	166	.51	172	.41	281
Grade 7	.78	213	.74	197	.44	163	.31	155	.36	223
Grade 9	.80	196	.73	166	.39	125	.29	121	.24	166
Grammar and Punctuation										
Grade 3	.72	260	.65	242	.46	179	.49	161	.41	299
Grade 5	.73	231	.67	221	.43	165	.46	168	.33	280
Grade 7	.63	213	.70	199	.42	163	.38	153	.38	229
Grade 9	.68	193	.62	166	.33	124	.42	123	.40	171
Writing										
Grade 3	.51	262	.49	242	.32	177	.36	160	.34	302
Grade 5	.61	228	.53	220	.40	164	.39	168	.29	277
Grade 7	.50	212	.53	197	.36	162	.27	152	.20	228
Grade 9	.54	194	.45	164	.22	124	.38	123	.18	171
Numeracy										
Grade 3	.75	259	.73	239	.55	182	.46	162	.38	301
Grade 5	.79	229	.73	219	.55	167	.49	171	.42	284
Grade 7	.84	211	.81	199	.48	164	.45	152	.37	229
Grade 9	.79	196	.79	166	.44	121	.43	115	.45	168

Note. MZ = monozygotic. DZ = dizygotic. F = female. M = male. OS = opposite sex. *n* = complete twin pairs.

Chapter 3

No qualitative sex-effects were significant in any achievement domain at any grade level; this was evident from no significant loss of fit when dropping either r_A or r_C from any of the full sex-limitation models (model fit statistics are reported in Tables A6–A10). For reading, in Grades 3, 7, and 9 no sex-effects were significant, but in Grade 5 the common-effects model fit the data best, female and male *a*, *c*, and *e* parameter estimates could not be equated. For spelling, sex-effects were present at each grade level. Scalar models were the best-fitting models in Grade 3, 5, and 7 with greater total variance in males, but a common-effects model was the best fit in Grade 9. For grammar and punctuation, and writing, in each grade, the most parsimonious models tested were those that allowed no sex-effects. For numeracy, the common-effects model was the best-fitting model in Grade 3, the scalar model with more variance for males was best in Grades 5 and 7, and no sex-effects was the best model in Grade 9.

Table 3.3 Univariate estimates from the best-fitting model for each NAPLAN domain and grade

	Model	A	C	E	Scalar
Reading					
Grade 3	Null	.71 [.58, .79]	.05 [0, .17]	.24 [.21, .27]	
Grade 5	Common F	.25 [.04, .50]	.43 [.20, .61]	.32 [.26, .39]	
	M	.71 [.60, .77]	.02 [0, .11]	.27 [.22, .33]	
Grade 7	Null	.58 [.44, .73]	.14 [0, .26]	.28 [.24, .32]	
Grade 9	Null	.61 [.46, .77]	.13 [0, .27]	.26 [.22, .31]	
Spelling					
Grade 3	Scalar	.76 [.64, .81]	.01 [0, .13]	.22 [.19, .26]	1.08 (.03)
Grade 5	Scalar	.77 [.65, .83]	.03 [0, .15]	.20 [.17, .23]	1.08 (.03)
Grade 7	Scalar	.78 [.68, .81]	0 [0, 0]	.22 [.19, .26]	1.14 (.04)
Grade 9	Common F	.68 [.48, .81]	.12 [0, .31]	.20 [.16, .25]	
	M	.68 [.49, .78]	.06 [0, .23]	.26 [.21, .33]	
GP					
Grade 3	Null	.51 [.38, .65]	.19 [.07, .31]	.30 [.26, .34]	
Grade 5	Null	.66 [.52, .75]	.06 [0, .19]	.28 [.24, .32]	
Grade 7	Null	.62 [.47, .73]	.06 [0, .20]	.31 [.27, .36]	
Grade 9	Null	.54 [.36, .70]	.12 [0, .28]	.34 [.29, .40]	
Writing					
Grade 3	Null	.39 [.21, .56]	.12 [0, .26]	.49 [.43, .55]	
Grade 5	Null	.52 [.34, .64]	.07 [0, .21]	.42 [.36, .48]	
Grade 7	Null	.50 [.30, .58]	.01 [0, .17]	.48 [.42, .55]	
Grade 9	Null	.42 [.20, .55]	.05 [0, .24]	.52 [.45, .60]	
Numeracy					
Grade 3	Common F	.39 [.20, .61]	.37 [.15, .54]	.24 [.20, .30]	
	M	.73 [.63, .79]	.03 [0, .11]	.24 [.20, .30]	
Grade 5	Scalar	.61 [.48, .73]	.17 [.04, .28]	.23 [.20, .26]	1.14 (.03)
Grade 7	Scalar	.79 [.66, .85]	.04 [0, .16]	.17 [.15, .20]	1.19 (.04)
Grade 9	Null	.65 [.50, .80]	.13 [0, .28]	.22 [.18, .25]	

Note. A = standardised additive genetic variance. C = standardised shared environmental variance. E = standardised unique environmental variance. GP = Grammar and Punctuation; Common = common effects sex-limitation model. Scalar = scalar effects sex-limitation model. Null model allows no sex-effects. 95% confidence intervals are in square brackets. F = females. M = males.

For reading, heritability estimates were large for most domains in each grade (see Table 3.3). Variation in reading was substantially influenced by genetic factors (.58–.71) and showed little influence from the shared environment (.02–.14), with the exception of girls in Grade 5. For girls in Grade 5, genetic factors contributed to only 25% of the variation in Grade 5 reading with the shared environment contributing 43%. In Grade 5, 98% of the sample also provided results for Grade 3 or Grade 7; as such, having a markedly different sex-effect and heritability estimates in Grade 5 compared to Grades 3 and 7 is unusual.

To explore this further, instead of comparing estimates from the most parsimonious models, we compared estimates from the common-effects models in each grade. Although there was no statistical loss of fit in the other grades, there might have been a pattern that did not reach significance to make some sense of the unusual Grade 5 results. In Grades 3, 5, 7, and 9, the heritability estimates for girls were quite varied, being .72, .25, .35, and .56 respectively, while for boys they were more stable, being .68, .72, .69 and .66. The shared environment estimates for girls were .04, .43, .36, and .17, and for boys they were again stable at .09, .02, .05, and .09. These estimates show a marked increase in the shared environment with a concomitant drop in heritability in girls in Grade 5, with the relative importance of genes and the shared environment gradually reversing with increasing grade. Meanwhile for boys the estimates remain quite stable over the grades. From looking at the intraclass correlations it seems that the significant sex-effects in Grade 5 might be driven by the low opposite-sex DZ correlation. However, we ran models separately for males and females with only same-sex twins and the estimates obtained showed the same pattern as the common-effects models (i.e. a sharp drop in heritability for girls in Grade 5 that gradually increased over time).

The other literacy domains showed more consistent results. Spelling was strongly influenced by genes with negligible shared environmental effects in Grades 3, 5, and 7. In Grade 9, when a, c, and e parameter estimates could not be equated across sex, there was no difference in heritability between girls and boys and the inability to equate the sexes resulted from differences in the contribution of the environment. For grammar and punctuation, genes explained between 52% and 66% of the variation in scores. At 18% there was a modest influence from the shared environment in Grade 3, but the shared environment was not significant in the latter grades. In each grade, the heritability of writing was a little lower than the other domains, explaining 39–52% of the variation in scores. The contribution of the unique environment was substantially higher for writing than the other domains, explaining close to half of the total variance.

For numeracy boys had larger variation overall than girls. The variance ratio of boys to girls in numeracy was 1.26:1 (averaged across all four grades). Across most grades there was a strong influence from genetic factors and smaller shared environmental effects, with Grade 3 girls as an exception. For girls in Grade 3, genetic and shared environmental influences were 39% and 37% respectively, thus equally influential on individual differences in numeracy scores. By Grade 5 the relative influence was equatable between girls and boys, with genes explaining 60% of the variance with a small effect (17%) from the shared environment.

Transition from primary to high school

Aside from the anomalous sex-effects in Grade 5 reading and Grade 3 numeracy, the standardised variance components were largely stable across grades within domains. In literacy, smaller standard deviations in latter grades indicated a trend for phenotypic variance to decrease with increasing grade, except writing, which increased in variance. Variance in numeracy was stable across the grades. Unstandardised parameter estimates show both genetic and unique environmental parameter estimates generally reduced in size (details in Table A11). The changes to the shared environmental variation over the grades show no discernable pattern, possibly due to less power to detect these generally smaller effect sizes, indicating that the more general changes in variance over time were not mediated by particular changes in either genetic or environmental influences.

Multivariate Analyses

Phenotypic correlations were high among all domains, ranging from .51 between writing and numeracy in Grade 3 to .72 between reading and grammar and punctuation in Grade 3. The correlations were very similar across the four grades (details in Table A12). The lowest correlations were with writing, ranging from .51–.60. Numeracy was as highly correlated with each of the literacy domains (.51–.69) as those literacy domains were with each other (.54–.72). Compared to the nonscalar sex-limitation correlated factors model, there was no significant loss of fit when path estimates and correlations were constrained to be equal between females and males in any grade (model fit statistics in Table A13). Given this, females and males were combined for multivariate analyses.

Genetic and environmental correlations

Genetic correlations were high among all of the variables, ranging from .65–.96 with an average genetic correlation of .81 across all domains and grades (see Tables 2.4–2.7 for specific genetic and environmental correlations). To the extent that genetic correlations are estimated as

less than one, there is some genetic independence on performance in bivariate pairings. How much of these shared genetic influences are due to genes common to all five domains or a single underlying construct will be addressed with fitting the independent pathway and common pathway models.

Unlike the consistency in the genetic correlations, the shared environmental correlations range widely; this is partly due to there being less power, which is evident in the wide confidence intervals. By contrast, the unique environmental correlations are stable and much smaller in size than the genetic correlations. They ranged from .05–.37 with an average correlation of .22 across all domains and grades. Unique environmental correlations, or twin-specific environmental correlations, typically indicate the presence of genuine environmental influences over and above measurement error. In these data there is a caveat. The tests were administered over three consecutive days, which likely introduces correlated errors within an individual across the tests, for example illness. Therefore, we cannot assume that these unique environmental correlations are free from measurement error.

The degree that a phenotypic correlation is mediated by genetic or environmental factors is a product of the variance in each trait that is explained by genes or the environment and the size of the genetic or environmental correlations between the traits. For example, the genetic mediation of the phenotypic correlation between reading and numeracy in Grade 7 is calculated from the square root of the heritability of reading multiplied by the genetic correlation between reading and numeracy multiplied by the square root of the heritability of numeracy. The portions of phenotypic correlations that are mediated by genes or the environment are reported above the diagonal in Tables 2.4–2.7. Comparing the portion to the phenotypic correlation, we can then calculate the proportion, or percentage, of covariation that is accounted for by genes or the environment. On average, 75–80% of the covariation among performance on NAPLAN tests is due to genetic factors. The remaining 20–25% is fairly evenly attributed to environmental factors that are either common to twins and affect performance on multiple tests, or are twin-specific environmental factors that affect performance on multiple tests. Although we cannot assess from these models if these genetic and environmental factors are stable in their contribution to each grade, the relative influence of genes and the shared and unique environments is consistent at each grade level.

Independent pathway models

To examine the covariance structure, initially an independent pathway model was fitted with two sets of genetic and environmental factors. The first factor loaded onto all domains (A1,

C1, E1), the second factor loaded onto only the literacy domains (A2, C2, E2). Sequentially, A2, C2, and E2 were dropped and the model fit compared with the saturated model. A common pathway model was also compared. In Grades 3, 5, and 7, both of the environmental second factors could be dropped; however, the second genetic factor could not. Interestingly, this second (literacy only) factor did not load significantly onto reading; it only held significant loadings for spelling, grammar and punctuation, and writing. For Grade 9, the second genetic factor could be dropped. In each grade the common pathway model was a significantly poorer fit than the saturated model. Figures 3.2–3.5 show the standardised path estimates of the best-fitting independent pathway models for each grade; dashed lines indicate paths with 95% confidence intervals that cross zero. Model fit comparisons are detailed in Table A14.

The structure of genetic influences on performance in the NAPLAN domains was very similar across Grade 3, 5, and 7. There was a general genetic factor with a substantial influence on all domains; when averaged across domains and grade this general factor contributed to 48% of the variation in performance (calculated by averaging the squared standardised path estimates from the measured traits to A1). The second genetic factor, which was allowed to load onto all of the literacy variables, might better be described as a spelling factor. Spelling had the strongest loading on this second factor; approximately 30% of the variation in spelling is accounted for by this second set of genes. About 9% of the variation in writing in Grade 3 is accounted for by this second set of genes, this diminished to 6% in Grade 5 and a mere 2% in Grade 7. Grammar and punctuation had small, but significant, loadings onto this second factor in Grades 3 and 5. Given that spelling, grammar, and punctuation skills contribute to the marking criteria of the NAPLAN writing task this small overlap is perhaps unsurprising. However, it is notable that this independent covariance among spelling, grammar, punctuation, and writing is genetic in origin and not environmental. There were no specific genetic influences on spelling, grammar and punctuation, or writing over-and-above those shared across all domains, or those from the second literacy factor. In contrast, reading did not load onto this second factor in any grade. In Grades 3 and 5 there were no significant reading-specific genetic effects, but in Grade 7 specific genetic influences contributed a small (6%) amount to the variance in reading.

The genetic structure in Grade 9 was a little different to the preceding grades, as there was no second genetic factor. There was still a substantial genetic influence on spelling independent of reading and numeracy (18%), but it now loaded on a spelling-specific genetic factor. Compared to the previous grades, the influence on reading from genes independent of the other tests was a little more substantial; specific genetic effects contributed 14% to the variation in reading.
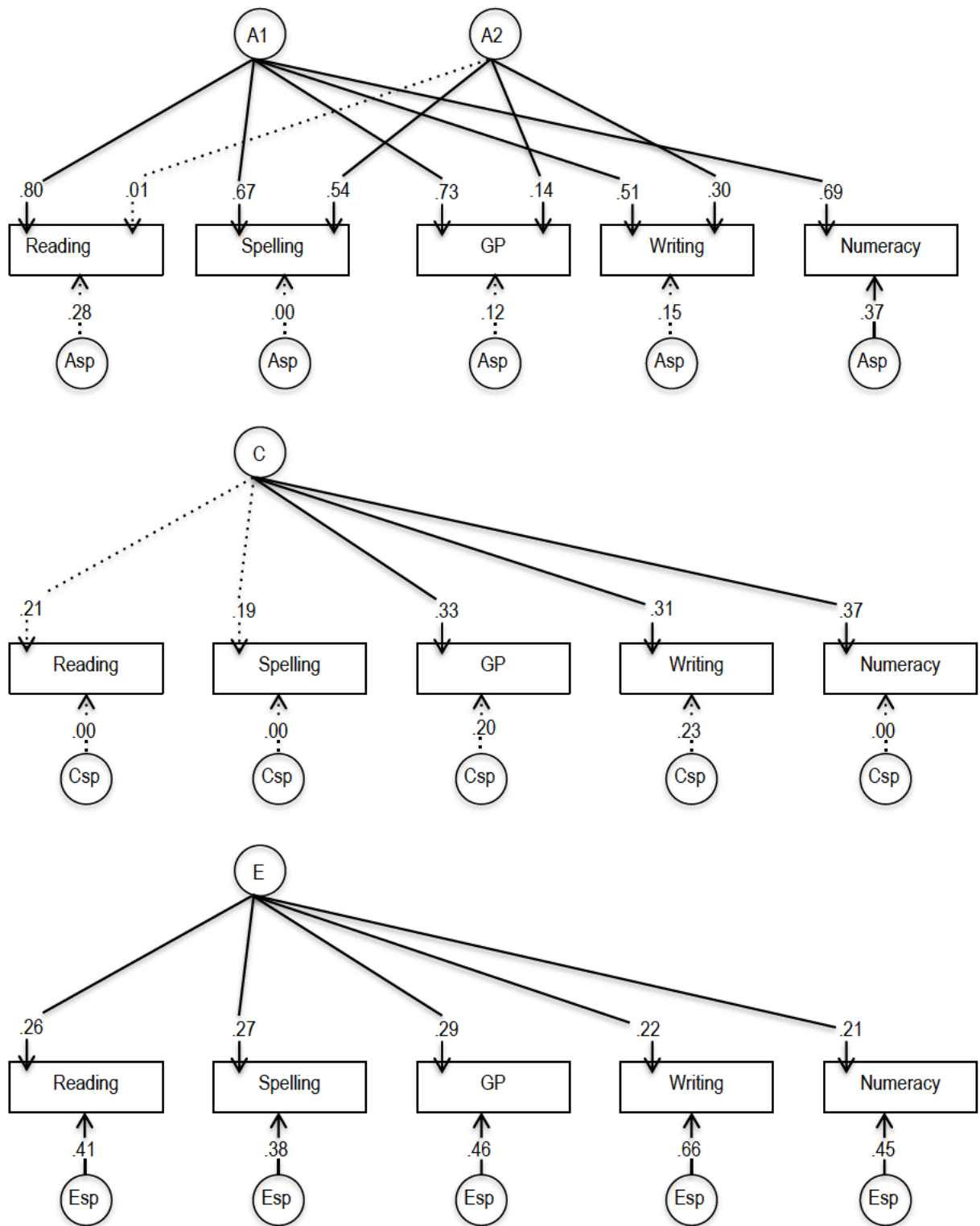
Figure 3.2 Path diagram of Grade 3 independent pathway model with standardised parameter estimates. Solid lines indicate significant paths, where 95% confidence intervals did not include zero. The model has a general additive genetic factor (A1), a literacy only factor (A2), a general shared environmental factor (C), a general unique environmental factor (E1), and domain specific factors (Asp, Csp, Esp). Most genetic influences were shared across domains; a second genetic factor influenced spelling, grammar and punctuation, and writing. Some independent genetic effects influenced numeracy. All shared environmental influences were general.
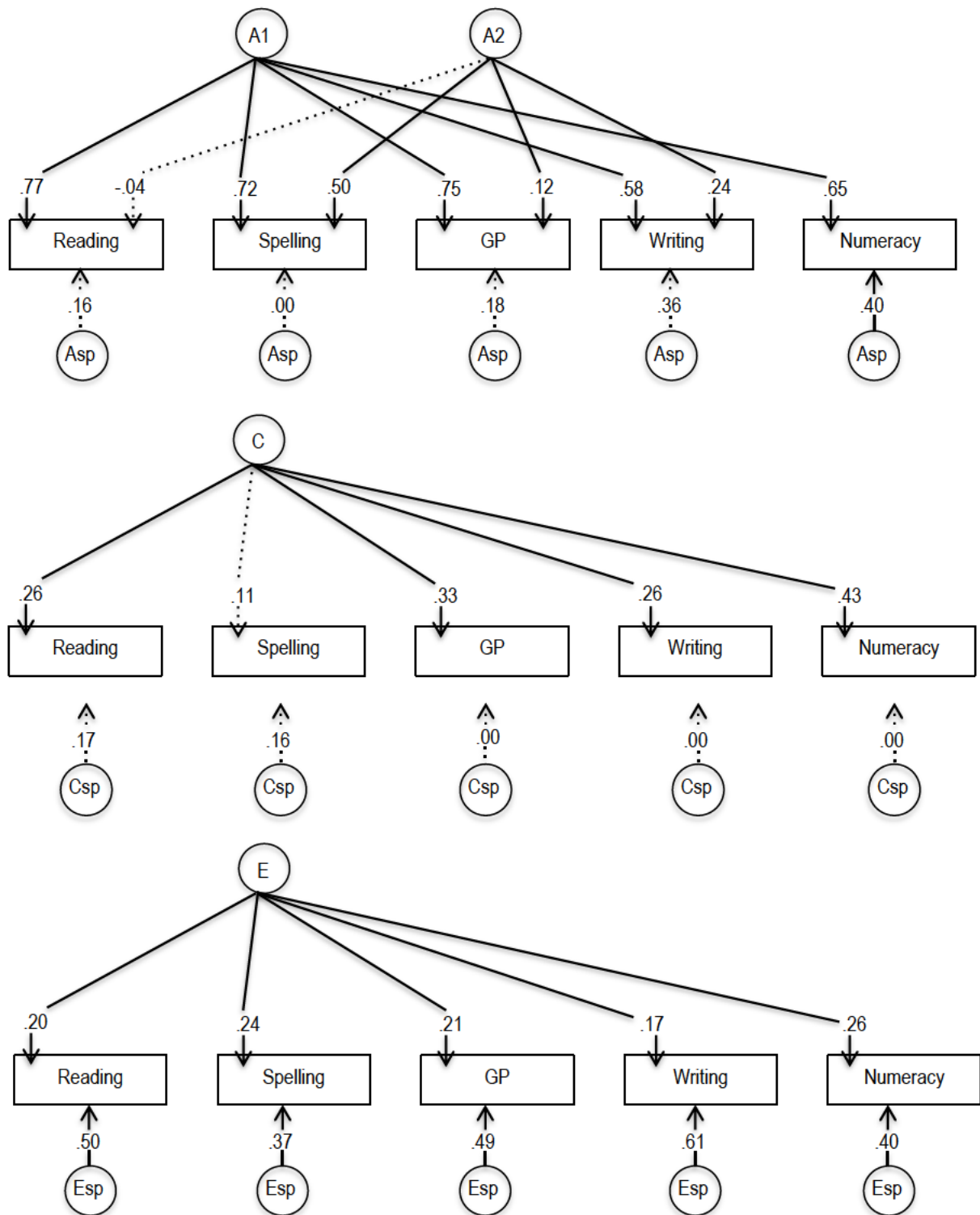
Figure 3.3 Path diagram of Grade 5 independent pathway model with standardised parameter estimates. Solid lines indicate significant paths, where 95% confidence intervals did not include zero. The model has a general additive genetic factor (A1), a literacy only factor (A2), a general shared environmental factor (C), a general unique environmental factor (E1), and domain specific factors (Asp, Csp, Esp). Most genetic influences were shared across domains; a second genetic factor influenced spelling, grammar and punctuation, and writing. Some independent genetic effects influenced numeracy. All shared environmental influences were general.
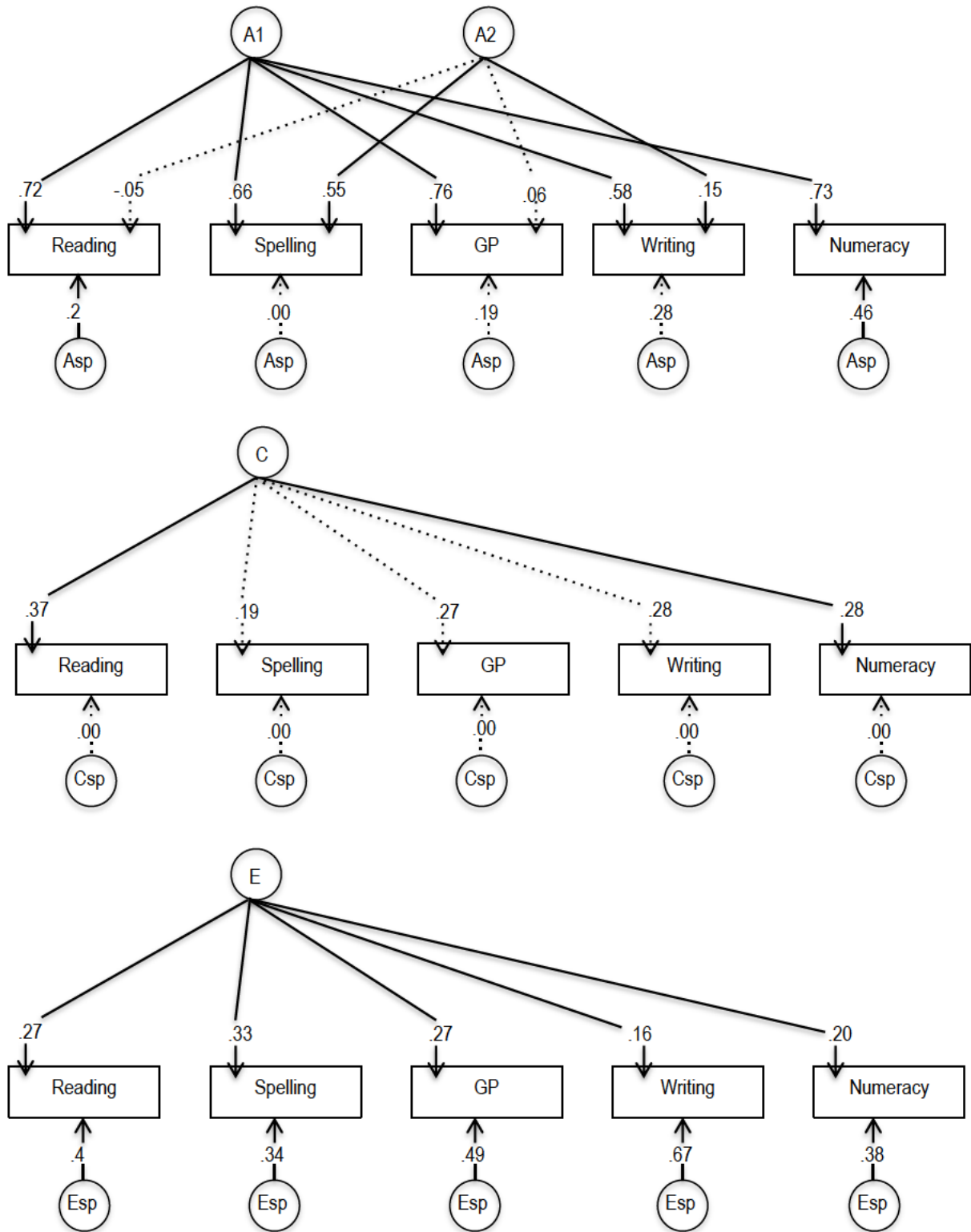
Figure 3.4 Path diagram of Grade 7 independent pathway model with standardised parameter estimates. Solid lines indicate significant paths, where 95% confidence intervals did not include zero. The model has a general additive genetic factor (A1), a literacy only factor (A2), a general shared environmental factor (C), a general unique environmental factor (E1), and domain specific factors (Asp, Csp, Esp). Most genetic influences were shared across domains; a second genetic factor influenced spelling and writing. Some independent genetic effects influenced numeracy and reading. All shared environmental influences were general.
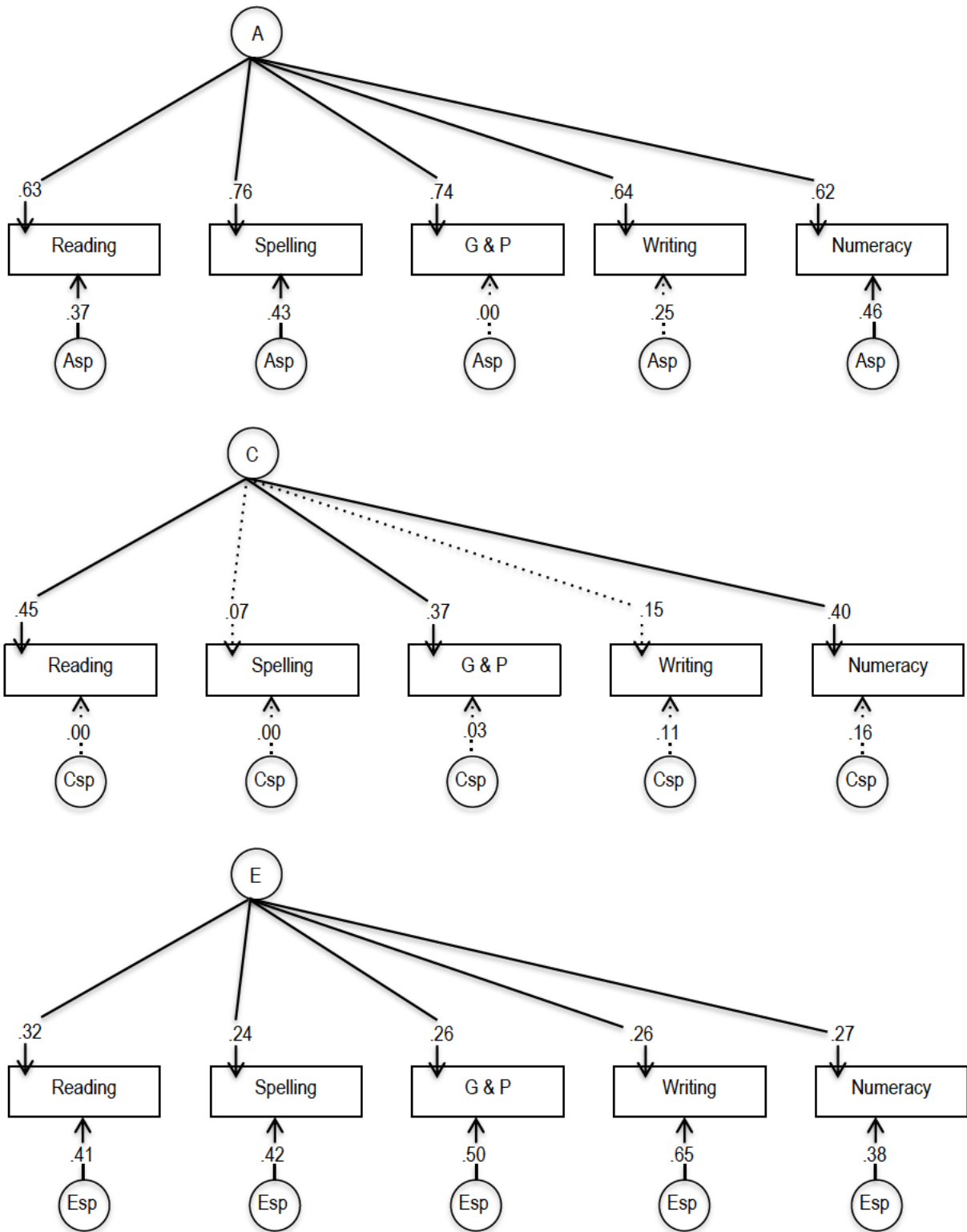
Figure 3.5 Path diagram of Grade 9 independent pathway model with standardised parameter estimates. Solid lines indicate significant paths, where 95% confidence intervals did not include zero. The model has a general additive genetic factor (A), a general shared environmental factor (C), a general unique environmental factor (E1), and domain specific factors (Asp, Csp, Esp). Most genetic influences were shared across domains; there were some literacy specific genetic influences on spelling, grammar and punctuation, and writing. Some independent genetic effects influenced numeracy, reading, and spelling. All shared environmental influences were general.

In contrast to the literacy variables, numeracy was only modelled to share genetic influences across all of the domains and to allow for specific genetic effects. Consistent with the generalist genes hypothesis, much of the variation in numeracy was due to shared genes, 48%, 43%, 53%, and 39% in Grades 3, 5, 7, and 9 respectively. However, in each grade, there was also a significant proportion of variation in numeracy performance due to genes that were not influencing performance in the other tests. Genes independent of the other domains influenced 13%, 16%, 22%, and 21% of the variation in numeracy performance in Grades 3, 5, 7, and 9 respectively.

Overall, the effect of the shared environment was small. There were no domain-specific shared environmental effects and the shared environment did not significantly influence spelling at any grade. However, the structure of the general shared environmental factor, allowed to load onto all domains, changes a little from grade to grade. The general shared environmental factor in Grade 3 influenced 10–14% of the variation in grammar and punctuation, writing, and numeracy. In Grade 5, reading was also influenced by this general factor, with 7–18% of the variation in reading, grammar and punctuation, writing, and numeracy due to common environmental factors. In Grade 7, only reading (14%) and numeracy (8%) were influenced by shared environmental influences common to both subjects. Performance in Grade 9 showed the general shared environmental factor to again have an influence on grammar and punctuation (14%) along with reading (20%), and numeracy (16%). Overall, the general shared environmental factor had a small influence on all domains, except for spelling, in most of the grades.

Unlike the shared environment, both the size and the structure of the unique environmental influences were consistent in each grade. Specific environmental influences unique to each twin and unique to each domain were important for variation in each grade. A large proportion of the variance in writing was due to these specific environmental effects (38–45%). It is, unfortunately, not possible to tell how much of this variance is due to measurement error. For the other domains, less variance was due to specific environmental effects (12–25%). However, in each grade, there was a consistent and small contribution from environmental factors that were unique to each twin but extended an influence on performance in each of the five domains. These twin-specific environments, which influenced each domain in a given grade, contributed 6% (on average) to the variation in performance on NAPLAN tests.

Discussion

The main aim of this study was to examine the relative influence of genes and the environment on individual differences in academic achievement in Australian school children.

Chapter 3

This is the first behaviour genetic study, in Australia, to include children from primary through to high school, and the first to explore grammar and punctuation, writing and numeracy. With very few exceptions, genes explain more of the variation in academic achievement in Australia than the environment. We found evidence of some sex-effects with more variation in boys than girls in both spelling and numeracy. The shared environment had a more substantial influence, and genes a smaller effect, in girls than in boys for numeracy in Grade 3 and for reading in Grade 5. These differences in magnitude diminished over time such that the heritability of girls increased to be on par with the boys. Contrary to expectation, we did not find a greater influence of the environment in high school compared to primary school.

Univariate Findings

Given the general findings of small to negligible shared environmental effects, the sex-effect in Grade 5 reading, which showed a small genetic and large shared environmental effect in girls, was remarkable. Such a large deviation from a general pattern in data is unexpected, particularly as we have a large number of longitudinal participants. Almost all of those in Grade 5 also participated in an adjacent grade and the tests are the same in style merely differ in difficulty as the grades progress. The dramatic changes in heritability estimates of girls from 71% in Grade 3 to 25% in Grade 5 and up again to 58% in Grade 7 warranted further exploration and required explanation, as we do here.

On examination, there was a slight decrease in the intraclass correlations of both female and male MZ twins in Grade 5 compared to Grades 3, along with a concomitant increase in the female DZ twins in Grade 5. From Grade 5 the intraclass correlations of the MZ twins gradually increased, while that of the female DZ twins gradually decreased. Although this difference did not seem remarkable on first viewing the intraclass correlations, it is this relative increase in covariance in the DZ females to the MZ females that is driving the sex-effects in reading in our Grade 5 data. The change in intraclass correlations appear to be an incremental, not disjointed, pattern. Thus, we compared the heritability results from the common-effects model from all grades instead of comparing across grades the heritability results from the best-fitting model in a grade. These estimates continued to indicate a sharp drop in heritability from Grade 3 to Grade 5 for girls, but then a gradual rise from Grade 5 through to Grade 9. It is possible sex-effects diminish gradually from Grade 5 to Grade 9. The absence of any sex-effects in Grade 3 reading, even in the common-effects model, means we are looking for a sex-effect that becomes important between the ages of 8 and 10. In Australia, in the national curriculum for English, Grade 3 marks the transition from 'learning to read' to 'reading to learn', which brings with it a decreasing

reliance on basic decoding mechanisms and an increased demand of inferential instead of literal comprehension. It is possible that girls and boys are being influenced to a different degree by their genes and environments as they bring this higher-level of interpretation to comprehending text. The problem with this explanation is there is no evidence of different magnitudes in genetic and shared environmental effects in other countries where children of a similar transition age have been assessed. In fact, our results are in the opposite direction to those noted by Petrill and Thompson (1994) who, in a sample of 6–12 year-old twins from Ohio, found higher heritability for girls and higher shared environmental influences in boys. Furthermore, considering the specificity suggested in the intraclass correlations and the absence of sex-effects internationally, to explain these data we are looking for factors unique to the Australian culture or education system that have a differential and remarkable effect on reading comprehension in female DZ twins in middle-late primary school. This sounds unlikely. Certainly, investigating possible sources of this specific source of covariance in our female DZ twins is a potential future study; however, the increased covariance of our female DZ twins in Grade 5 is perhaps an artefact of our sample, which might dissipate with the inclusion of more waves of data. Equating females and males in a multivariate model that included these Grade 5 reading data did not result in a significant loss of model fit; this adds weight to the option that the sex-effect in these Grade 5 reading data is perhaps a chance event. Replication is desirable before over-interpreting such an unusual finding.

Unlike reading, the results from spelling, grammar and punctuation, and writing are more stable over the grades for both sexes. Spelling had particularly high and stable genetic influences, which is comparable to findings from previous research on younger Australian school children (Samuelsson et al., 2008). Males did have greater overall variance for spelling than females, and this is consistent with other studies that have used a measure of spelling (Hawke, Olson, Willcut, Wadsworth, & DeFries, 2009) and wider research on academic domains (Hedges & Nowell, 1995; Machin & Pekkarinen, 2008). Genes were also the strongest influence on the variation in grammar and punctuation performance. The shared environment had a small, but significant, influence only in Grade 3. The unique environmental influences were also a little higher for grammar and punctuation than they were for spelling.

Writing has the most marked influence from the unique environment. For writing, the estimates of A and E are about on par, the only domain to have such a strong impact from the unique environment. There might be many reasons for this, but we will propose two. The nature of the writing task is certainly one of the most complex of the literacy tests. The reading, spelling, grammar and punctuation tests are all predominantly multiple choice. In contrast, the writing test

requires a student to generate their answer, and they are scored on how accurate, comprehensive, and integrated their passage of writing is. Even at the Grade 3 level there is an expectation that the passage of writing will be several paragraphs in length with a cogent flow of ideas. It seems possible that how well a student executes their writing task might in part result from how they personally identify with or are inspired by the writing prompt. Alternatively, it is probable there is more measurement error in the writing task. There are 10 criteria that papers are marked on, and although the developers report a high Cronbach's alpha, no inter-rater reliability data are provided. Inter-rater reliability is not a source of error on any of the other test domains. There is, perhaps, an indication of this higher measurement error in the unique environmental correlations. Even though spelling, grammar and punctuation all contribute to the marking criteria of the writing test, the unique environmental correlations are smaller between writing and these literacy domains than those between numeracy and these literacy domains. This argument is further supported by comparing our low unique environmental correlation between writing and reading with the moderate one (.37) found by Oliver et al. (2007) in the UK. Oliver et al. also had a much lower unique environmental variance component for writing (.27) than our data, possibly indicative of less error in the measurement of their writing task. The nature of their five-choice teacher rating scale makes the likelihood of reduced measurement error plausible. Given our writing assessment is, arguably, more susceptible to measurement error, then our finding that genes still account for about half of the variation in scores is impressive.

We found heritability estimates of numeracy were comparable to those from the literacy domains, and contrary to expectation, we did not find an increase in environmental influences in the high school grades. Instead we found the unique environment to hold quite steady and the shared environment to only be significantly contributing to variation in numeracy performance in the primary grades; it was most pronounced in girls in Grade 3. This would suggest that variation of teacher's pedagogical content knowledge in mathematics and the effects of high school teachers teaching out-of-field are not contributing substantially to variation in numeracy performance, at least as measured by the NAPLAN. The sex-effect present in Grade 3 was in the same direction as that found in Grade 5 reading, with higher heritability in boys than girls and a greater influence from the shared environment in girls than boys. As this sex-effect is not internationally pervasive, it would seem unlikely to be a consequence of an age- and sex-specific interaction with genes, and more likely that the effect is due to the environment of our Australian sample. Furthermore, since there was no significant qualitative difference in the shared environment, our results imply that girls and boys are responding differently to the same environments and in such a way as to affect numeracy performance. There is some evidence that

girls in primary school are differentially influenced by female teachers' mathematical anxiety depending on the girls' traditional gender beliefs. Beilock et al. (2010) found that for girls who agreed with gender stereotypes and were taught by female teachers, then the mathematical anxiety of the teacher was related to the girls' mathematical achievement at the end of the year. This was not the case for boys. In a similar vein, Bleeker and Jacobs (2004) found that mothers' beliefs about mathematics and gender had a differential influence on girls and boys, albeit in older participants. These are examples of possible environmental sources of variation in mathematical performance that are specifically relevant to girls; unfortunately we cannot tell from our data what is causing the differential influences on girls and boys in Grade 3. However, we do note that as schooling progresses this interaction that contributes to variation in performance in girls diminishes in effect.

Multivariate Findings

Our multivariate findings are clearly in support of the generalist genes hypothesis. High bivariate genetic correlations are evident among all of the academic domains, averaging .85 among the literacy variables and .75 between numeracy and the literacy variables. Most of the genes that affect performance in one academic domain are also affecting performance in at least one other domain. From our independent pathway models we can extend on from the bivariate relatedness and conclude that a set of genes shared across all of the academic domains accounted for about half of the variation in each of reading, spelling, grammar and punctuation, and numeracy, and about one third of the variation in writing. These same genes mediated three quarters of the phenotypic correlations among these achievement domains. Certainly, each test requires some reading ability. There are allowances in the test administration protocols for questions to be read allowed, but no doubt some of the correlations, both phenotypic and genetic, are due in part to reading skill. However, given the second genetic factor loaded on the literacy variables except reading, the genetic covariation across domains is not solely due to reading.

The magnitude of overlap across all five literacy and numeracy domains is, perhaps, unsurprising when considering possible functional pathways of genetic influence, such as genes responsible for the expression or regulation of neurotransmitters or receptors involved in the neural correlates of learning and memory. Some of the links between heritability and these functional pathways are being made; researchers investigating both specific genes and sets of genes have found evidence that variation in the set of genes responsible for expression of the *N*-methyl-D-aspartate receptor complex contributes to variation in fluid cognitive ability (W. D. Hill et al., 2014). From the study by Hill et al. (2014), a particularly pertinent finding to the generalist

genes hypothesis is that the set of genes, instead of a single gene, was preferentially associated with variation at the level of fluid cognitive ability and general cognitive ability. Academic achievement has been shown to have genetic overlap with a range of other traits including intelligence, self-efficacy, personality, and behaviour (Krapohl et al., 2014). With such diverse areas of genetic overlap, there are many potential neural, endocrine, and metabolic pathways yet to be investigated that might have a bearing on academic ability, with the challenge that any given effect size is likely to be small. Although we are far from a comprehensive understanding of the genes, or sets of genes, that are influencing various cognitive and academic abilities, it is encouraging to see the conceptual findings of common genetic influence, such as ours, are mirrored in physical genome studies.

An important finding from this study is that genetic influence across academic achievement as measured by NAPLAN performance is not solely through the actions of a single set of genes. The bivariate genetic correlations were, generally, significantly less than one, indicating unique genetic influences. From the independent pathway models there was genetic contribution to numeracy performance independent of the literacy domains, and as the grades progressed there was an increasing amount of genetic influence on reading that was independent of the other literacy domains. It is probable that some of the variance modelled as independent is shared with other unmeasured traits, and, as such, the genes are still generalists is some way. However, where achievement tests of mathematics and reading have been employed, there is both genetic overlap and independence (e.g. Hart et al., 2010; Hart et al., 2009; Knopik & DeFries, 1999). This has also been the case when using teacher ratings in the UK (e.g. Kovas et al., 2005). That we have replicated these findings in NAPLAN data is encouraging; it confirms a degree of specificity to the assessment of the domains, which we explored in a much smaller sample (Grasby, Byrne, & Olson, 2015). This independence of test domain is also apparent in the poorer fit to the data of a common pathway model. A common pathway model would have suggested that a single latent construct was appropriate to explain individual differences in NAPLAN performance, which was not the case with these data.

The evidence is strong for the importance of genetic variation in explaining individual differences in academic performance. High heritability does not mean students are at the mercy of their genotype. Heritability estimates are in direct proportion to the presence of variation in environmental factors that influence the phenotype. That heritability is a strong influence on individual differences in NAPLAN performance means the environment, in so far as it influences performance, is reasonably constant. Although reasonably constant, the environmental range in our sample still contributed a significant amount to variation and to covariation in performance

across domains, meaning that after controlling for genes there are some environmental factors influencing some individuals to perform higher than average and some factors influencing individuals to perform below average. If we consider only the covariation among all five of our domains, to exclude any measurement error, then the combined influence of the shared and unique environment is (on average) 8% of the variation in any given domain. The size of this environmental covariance effect is of a similar magnitude to "teacher effects" or classroom effects in the early school years, where classroom variation contributes 5–10% of the variation in student performance in reading and mathematics (Byrne et al., 2010; Nye, Konstantopoulos, & Hedges, 2004; Taylor, Roehrig, Hensler, Connor, & Schatschneider, 2010). Although smaller in size than that of genetic influences, this variation is still considered of practical importance. It is within this environmental portion of variation that we look for factors that can be altered so as to bring greater equality of opportunity to students.

Limitations

There are a number of limitations with our study. Firstly, on average, our sample scored .39 of a standard deviation higher than the Australian population. We did, however, have representation across the entire distribution; the national results are divided into 10 bands and our sample included participants in each band in each domain in each grade. Our participants were predominantly recruited through the Australian Twin Registry, a volunteer registry for twins interested in participating in research. As such, we expect there is some restriction of environmental range in our sample. This is particularly important to consider in view of the generally small effects found from the shared environment. Secondly, our measures do not allow for analysis of sub-domains of these broad academic areas; for example we cannot explore the relative contribution of word decoding and language comprehension to our reading measure, or assess the five aspects of mathematics that the numeracy test is composed of. However, the domain-level scores that we have analysed in this study are a measure of performance against curricula outcomes, which supports the use of these data as measures of academic achievement in these domains. Finally, although we have reported measures across four grade levels, there is a high proportion of overlap in participants across grades. This is a necessary property for longitudinal analyses, but in this paper, where we have reported from within each grade level, we acknowledge that the probability of similar findings across grades is greater than if the participants at each grade level were independent samples.

Chapter 3

Conclusions

Bearing these limitations in mind, our key findings from this study are substantial heritability of individual differences in performance on literacy and numeracy tests in Australia with generally a small contribution of the shared environment. We found an exception to this pattern in Grade 3 numeracy, where girls were equally influenced by genes and the environment, and in Grade 5 reading, where the shared environment primarily influenced girls, although, we habour doubts for the robustness of, primarily, the Grade 5 reading sex-effect. Overall, boys were found to have greater variation than girls. This was significant for numeracy and spelling performance and is consistent with international findings on academic and cognitive measures. Averaged across domains and grade, genetic factors explained 60%, shared environment 10%, and unique environment 30% of the variation in achievement; these results are very similar to findings from the UK. We found evidence of generalist genes, with 80% of genetic variation in any given domain due to the influence of a set of genes that affected performance in all domains. There were also small, but significant, environmental effects influencing all domains. This finding of small environmental effects is not to assert that environmental factors are unimportant. It matters that children are housed, clothed, fed, and attend school—typically shared environmental factors. However, the amount of variation in these factors (e.g., where children live and which school they attended) did not contribute very much to the variation in performance after controlling for genetic effects. These results might be considered encouraging, in that it describes a relative degree of environmental parity, at least within our sample.

Acknowledgements

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*(3), 317-332. doi:10.1007/BF02294359

Alarcón, M., Knopik, V. S., & DeFries, J. C. (2000). Covariation of mathematics achievement and general cognitive ability in twins. *Journal of School Psychology, 38*(1), 63-77. doi:10.1016/S0022-4405(99)00037-0

Australian Curriculum Assessment and Reporting Authority. (2008). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2008*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2009). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2009*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2010a). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2010*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2010b). *Submission to the senate education, employment and workplace relations committee: Inquiry into the administration and reporting of NAPLAN testing [Appendix C: NAPLAN technical paper]* (Submission 261). Retrieved from http://www.aph.gov.au/Parliamentary_Business/Committees/Senate_Committees?url=eet_ctte/completed_inquiries/2010-13/naplan/submissions.htm

Australian Curriculum Assessment and Reporting Authority. (2011). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2011*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2012). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2012*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2013). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2013*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2014a). *NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2014*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2014b). *National Assessment Program—Literacy and Numeracy 2013: Technical report*. Retrieved from http://www.nap.edu.au/results-and-reports/national-reports.html

Australian Curriculum Assessment and Reporting Authority. (2015). *National Assessment Program—Literacy and Numeracy 2014: Technical report*

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133-180. doi:10.3102/0002831209345157

Beilock, S. L., Gunderson, E. A., Ramirez, G., Levine, S. C., & Smith, E. E. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 107*(5), 1860-1863. doi:10.2307/40536499

Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96*(1), 97-109. doi:10.1111/1467-8624.00273

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306-317. doi:10.1007/s11336-010-9200-6

Byrne, B., Coventry, W. L., Olson, R. K., Wadsworth, S. J., Samuelsson, S., Petrill, S. A., . . . Corley, R. (2010). "Teacher effects" in early literacy development: Evidence from a study of twins. *Journal of Educational Psychology, 102*(1), 32-42. doi:10.1037/a0017288

Byrne, B., Delaland, C., Fielding-Barnsley, R., Quain, P., Samuelsson, S., Høien, T., . . . Olson, R. K. (2002). Longitudinal twin study of early reading development in three countries: Preliminary results. *Annals of Dyslexia, 52*(1), 47-73. doi:10.1007/s11881-002-0006-9

Byrne, B., Samuelsson, S., Wadsworth, S., Hulslander, J., Corley, R., DeFries, J., . . . Olson, R. (2007). Longitudinal twin study of early literacy development: Preschool through Grade 1. *Reading and Writing, 20*(1), 77-102. doi:10.1007/s11145-006-9019-9

Byrne, B., Wadsworth, S., Corley, R., Samuelsson, S., Quain, P., DeFries, J. C., . . . Olson, R. K. (2005). Longitudinal twin study of early literacy development: Preschool and kindergarten phases. *Scientific Studies of Reading, 9*(3), 219-235. doi:10.1207/s1532799xssr0903_3

Davis, O. S. P., Haworth, C. M. A., & Plomin, R. (2009). Learning abilities and disabilities: Generalist genes in early adolescence. *Cognitive Neuropsychiatry, 14*(4-5), 312-331. doi:10.1080/13546800902797106

Davis, O. S. P., Kovas, Y., Harlaar, N., Busfield, P., McMillan, A., Frances, J., . . . Plomin, R. (2008). Generalist genes and the Internet generation: Etiology of learning abilities by web testing at age 10. *Genes, Brain and Behavior, 7*(4), 455-462. doi:10.1111/j.1601-183X.2007.00370.x

de Zeeuw, E. L., de Geus, E. J. C., & Boomsma, D. I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education, 4*(3), 69-76. doi:10.1016/j.tine.2015.06.001

Evans, D. M., Frazer, I. H., & Martin, N. G. (1999). Genetic and environmental causes of variation in basal levels of blood cells. *Twin Research and Human Genetics, 2*(04), 250-257. doi:doi:10.1375/twin.2.4.250

Friend, A., DeFries, J. C., & Olson, R. K. (2008). Parental education moderates genetic influences on reading disability. *Psychological Science, 19*(11), 1124-1130. doi:10.1111/j.1467-9280.2008.02213.x

Grasby, K. L., Byrne, B., & Olson, R. K. (2015). Validity of large-scale reading tests: A phenotypic and behaviour–genetic analysis. *Australian Journal of Education, 59*(1), 5-21. doi:10.1177/0004944114563775

Harlaar, N., Dale, P. S., & Plomin, R. (2005). Telephone testing and teacher assessment of reading skills in 7-year-olds: II. strong genetic overlap. *Reading and Writing, 18*(5), 401-423. doi:10.1007/s11145-005-0271-1

Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development, 78*(1), 116-131. doi:10.1111/j.1467-8624.2007.00988.x

Harlaar, N., Kovas, Y., Dale, P. S., Petrill, S. A., & Plomin, R. (2012). Mathematics is differentially related to reading comprehension and word decoding: Evidence from a genetically sensitive design. *Journal of Educational Psychology, 104*(3), 622-635. doi:10.1037/a0027646

Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: A study of 7-year-old twins. *Journal of Child Psychology and Psychiatry, 46*(4), 373-384. doi:10.1111/j.1469-7610.2004.00358.x

Hart, S. A., Petrill, S. A., & Thompson, L. A. (2010). A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learning and Individual Differences, 20*(2), 63-69. doi:10.1016/j.lindif.2009.10.004

Hart, S. A., Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology, 101*(2), 388. doi:10.1037/a0015115

Hart, S. A., Soden, B., Johnson, W., Schatschneider, C., & Taylor, J. (2013). Expanding the environment: Gene × school-level SES interaction on reading comprehension. *Journal of Child Psychology and Psychiatry, 54*(10), 1047-1055. doi:10.1111/jcpp.12083

Hawke, J. L., Olson, R. K., Willcut, E. G., Wadsworth, S. J., & DeFries, J. C. (2009). Gender ratios for reading difficulties. *Dyslexia, 15*(3), 239-242. doi:10.1002/dys.389

Haworth, C. M. A., Kovas, Y., Harlaar, N., Hayiou-Thomas, M. E., Petrill, S. A., Dale, P. S., & Plomin, R. (2009). Generalist genes and learning disabilities: A multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. *Journal of Child Psychology and Psychiatry, 50*(10), 1318-1325. doi:10.1111/j.1469-7610.2009.02114.x

Haworth, C. M. A., Kovas, Y., Petrill, S. A., & Plomin, R. (2007). Developmental origins of low mathematics performance and normal variation in twins from 7 to 9 years. *Twin Research and Human Genetics, 10*(1), 106-117. doi:10.1375/twin.10.1.106

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45. doi:10.1126/science.7604277

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511. doi:10.1080/07370000802177235

Hill, W. D., Davies, G., van de Lagemaat, L. N., Christoforou, A., Marioni, R. E., Fernandes, C. P. D., . . . Deary, I. J. (2014). Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Translational Psychiatry, 4*, e341. doi:10.1038/tp.2013.114

Hobbs, L. (2015, April 13). Too many teachers teaching outside their area of expertise. *The Conversation*. Retrieved from http://www.theconversation.com/au

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science, 16*(5), 259-263. doi:10.1111/j.1467-8721.2007.00516.x

Jackson, R. W., Snieder, H., Davis, H., & Treiber, F. A. (2001). Determination of twin zygosity: A comparison of DNA with various questionnaire indices. *Twin Research, 4*(1), 12-18. doi:10.1375/1369052012092

Knopik, V. S., Alarcón, M., & DeFries, J. C. (1998). Common and specific gender influences on individual differences in reading performance: A twin study. *Personality and Individual Differences, 25*(2), 269-277. doi:10.1016/S0191-8869(98)00040-3

Knopik, V. S., & DeFries, J. C. (1999). Etiology of covariation between reading and mathematics performance: A twin study. *Twin Research and Human Genetics, 2*(3), 226-234. doi:10.1375/twin.2.3.226

Kovas, Y., Harlaar, N., Petrill, S. A., & Plomin, R. (2005). "Generalist genes" and mathematics in 7-year-old twins. *Intelligence, 33*(5), 473-489.

Kovas, Y., Haworth, C. M. A., Petrill, S. A., & Plomin, R. (2007). Mathematical ability of 10-year-old boys and girls. *Journal of Learning Disabilities, 40*(6), 554-567. doi:10.1177/00222194070400060601

Kovas, Y., & Plomin, R. (2007). Learning abilities and disabilities: Generalist genes, specialist environments. *Current Directions in Psychological Science, 16*(5), 284-288. doi:10.2307/20183217

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., . . . Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences of the United States of America, 111*(42), 15273-15278. doi:10.1073/pnas.1408777111

Limbrick, L., Wheldall, K., & Madelaine, A. (2010). Estimating gender ratios of poor reading using large-scale assessments. *Australian Journal of Education, 54*(2), 190-222. doi:10.1177/000494411005400206

Lykken, D. T., Bouchard, T. J., McGue, M., & Tellegen, A. (1990). The Minnesota Twin Family Registry: Some initial findings. *Acta Geneticae Medicae et Gemellologiae, 39*(1), 35-70. doi:10.1017/S0001566000005572

Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science, 322*(5906), 1331-1332. doi:10.2307/20176888

Neale, M. C., & Maes, H. H. (2004). *Methodology for the genetic studies of twins and families* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Neale, M. C., Røysamb, E., & Jacobson, K. (2006). Multivariate genetic analysis of sex limitation and G x E interaction. *Twin Research and Human Genetics, 9*(4), 481-489. doi:10.1375/183242706778024937

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257. doi:10.2307/3699577

Oliver, B., Dale, P. S., & Plomin, R. (2007). Writing and reading skills as assessed by teachers in 7-year olds: A behavioral genetic approach. *Cognitive Development, 22*(1), 77-95. doi:10.1016/j.cogdev.2006.07.003

Oliver, B., Harlaar, N., Hayiou Thomas, M. E., Kovas, Y., Walker, S. O., Petrill, S. A., . . . Plomin, R. (2004). A twin study of teacher-reported mathematics performance and low performance in 7-year-olds. *Journal of Educational Psychology, 96*(3), 504-517. doi:10.1037/0022-0663.96.3.504

Olson, R. K., Hulslander, J., Christopher, M., Keenan, J. M., Wadsworth, S. J., Willcutt, E. G., . . . DeFries, J. C. (2013). Genetic and environmental influences on writing and their relations to language and reading. *Annals of Dyslexia, 63*(1), 25-43. doi:10.1007/s11881-011-0055-z

Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., DeThorne, L., & Vandenbergh, D. J. (2007). Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. *Reading and Writing, 20*(1-2), 127-146. doi:10.1007/s11145-006-9021-2

Petrill, S. A., Logan, J., Hart, S. A., Vincent, P., Thompson, L. A., Kovas, Y., & Plomin, R. (2012). Math fluency is etiologically distinct from untimed math performance, decoding fluency, and untimed reading performance: Evidence from a twin study. *Journal of Learning Disabilities, 45*(4), 371. doi:10.1177/0022219411407926

Petrill, S. A., & Thompson, L. A. (1994). The effect of gender upon heritability and common environmental estimates in measures of scholastic achievement. *Personality and Individual Differences, 16*(4), 631-640. doi:10.1016/0191-8869(94)90190-2

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131*(4), 592-617. doi:10.1037/0033-2909.130.6.858

Plomin, R., Kovas, Y., & Haworth, C. M. A. (2007). Generalist genes: Genetic links between brain, mind, and education. *Mind, Brain, and Education, 1*(1), 11-19. doi:10.1111/j.1751-228X.2007.00002.x

Reynolds, C. A., Hewitt, J. K., Erickson, M. T., Silberg, J. L., Rutter, M., Simonoff, E., . . . Eaves, L. J. (1996). The genetics of children's oral reading performance. *Journal of Child Psychology and Psychiatry, 37*(4), 425-434.

Samuelsson, S., Byrne, B., Olson, R. K., Hulslander, J., Wadsworth, S., Corley, R., . . . DeFries, J. C. (2008). Response to early literacy instruction in the United States, Australia, and Scandinavia: A behavioral-genetic analysis. *Learning and Individual Differences, 18*(3), 289-295. doi:10.1016/j.lindif.2008.03.004

Senate Standing Committee on Education and Employment. (2014). *Effectiveness of the National Assessment Program – Literacy and Numeracy*. Canberra: Senate Printing Unit Retrieved from

http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Empl oyment/Naplan13/Report/index.

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M., . . . Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PloS One, 8*(12), e80341. doi:10.1371/journal.pone.0080341

Spinath, F. M., Spinath, B., & Plomin, R. (2008). The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality, 22*(3), 211-229. doi:10.1002/per.677

Taylor, J., Roehrig, A. D., Hensler, B. S., Connor, C. M., & Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science, 328*(5977), 512-514. doi:10.1126/science.1186149

Taylor, J., & Schatschneider, C. (2010). Genetic influence on literacy constructs in kindergarten and first grade: Evidence from a diverse twin sample. *Behavior Genetics, 40*(5), 591-602. doi:10.1007/s10519-010-9368-7

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research and Human Genetics, 5*(05), 444-448. doi:doi:10.1375/twin.5.5.444

Wadsworth, S. J., & DeFries, J. C. (2005). Genetic etiology of reading difficulties in boys and girls. *Twin Research and Human Genetics, 8*(06), 594-601.

Higher Degree Research Thesis by Publication
University of New England

STATEMENT OF AUTHORS' CONTRIBUTION

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research Master/PhD candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated in the *Statement of Originality*.

	Author's Name (please print clearly)	Contribution
Candidate	Katrina Grasby	Responsible for study design; data cleaning and analysis; writing of paper
Other Authors	William Coventry	Provided editorial comments; author of grant, which provided data for paper
	Brian Byrne	Provided editorial comments; author of grant, which provided data for paper
	Richard Olson	Provided editorial comments on introduction; author of grant, which provided data for paper
	Sarah Medland	Provided assistance with data analysis and editorial comments

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry

27/1/16
Date

28/1/16
Date

Higher Degree Research Thesis by Publication
University of New England

STATEMENT OF ORIGINALITY

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research Master/PhD candidate and the candidate's Principal Supervisor, certify that the following text, figures and diagrams are the candidate's original work.

Type of work	Page number/s
All text, tables, and figures	Chapter 3, pp46–83

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry
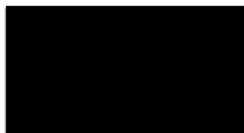
28/1/16
Date

25/1/16
Date

Chapter 4
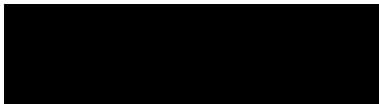
Longitudinal Stability and Growth in Literacy and Numeracy in Australian School Students

Katrina L. Grasby and William L. Coventry

Candidate Signature

Principal Supervisor Signature

Chapter 4

Literacy and numeracy are two key learning outcomes of formal education. Given the societal importance of acquiring functional ability in these areas, there is considerable research invested in understanding the development of these skills and why people differ in their ability. Numerous factors influence variation in performance on assessments of literacy and numeracy, including socioeconomic status (SES; Sirin, 2005), home environments (Smyth, Whelan, McCoy, Quail, & Doyle, 2010), classroom and teacher (Byrne et al., 2010; Nye, Konstantopoulos, & Hedges, 2004), study (Cooper, Robinson, & Patall, 2006), confidence (Stankov, Morony, & Lee, 2014), and genetic variation (Kovas et al., 2007). Genetic variation has generally been found to account for more than half of the variation in literacy and numeracy performance in school students in developed countries (de Zeeuw, de Geus, & Boomsma, 2015; Olson, Keenan, Byrne, & Samuelsson, 2014). Yet time-point specific assessment of ability is only one aspect of assessing the development of these skills. Students also vary in the rate they acquire literacy and numeracy skills. Depending on the skill measured and ability at initial performance, variation in growth trajectories has been found to differ and result in either increasing or decreasing achievement gaps (Pfost, Hattie, Dörfler, & Artelt, 2014; Shin, Davison, Long, Chan, & Heistad, 2013). Variation in growth trajectories has also been linked with SES, cognitive resources, and behaviour (Baumert, Nagy, & Lehmann, 2012; Morgan, Farkas, & Wu, 2011). So far, relatively few studies have investigated the influence of genetic variation on growth, or investigated the extent of environmental influences after controlling for genetic influences.

In this paper we examine the relative influence of genes and the environment on both stability and growth in literacy and numeracy performance in Australian twins assessed on the National Assessment Program in Literacy and Numeracy (NAPLAN). Since 2008, each year students in Grades 3, 5, 7, and 9 sit nationwide standardised tests of reading, spelling, grammar and punctuation, writing, and numeracy (Senate Standing Committee on Education and Employment, 2014). Student growth is one aspect of the reporting available on these tests, and has been considered a basis to judge the value added by schools (Masters, Rowley, Ainley, & Khoo, 2008). However, it is important to assess the extent of genetic influence on growth before assuming that growth in scores indicates school value. Grasby et al. (2015) found genes were important contributors to variation in each of the literacy and numeracy domains at each grade level in these Australian tests. In this paper we test the genetic and environmental etiology of individual differences in both stability and growth of literacy and numeracy.

Past performance is one of the best predictors of achievement in an academic domain; students who achieve highly in one year will typically achieve highly in a later year, and those who score lower in one year will typically score lower in another (Hattie, 2008). For reading,

genes have been found to be the strongest contributor to this stability in relative performance between students in the early years of school (Byrne et al., 2005; Olson et al., 2011; Samuelsson et al., 2008; Soden et al., 2015), and through middle school to high school (Betjemann et al., 2008; Harlaar, Dale, & Plomin, 2007; Wadsworth, Corley, Hewitt, & DeFries, 2001). These studies comprise a range of measures of reading skills, including phonological awareness, word reading, spelling, vocabulary, comprehension, and reading fluency. Consistently, many of the same genes are influencing variation in performance at multiple ages with genetic correlations ranging from .65 (Harlaar et al., 2007) to 1.0 (Betjemann et al., 2008; Olson et al., 2011; Wadsworth et al., 2001). Similar results have been found for mathematics, with genetic correlations of .62, .68, and .73 reported among participants tested at ages 7, 9 and 10 (Kovas et al., 2007). From these studies, genes have explained at least 75% of the phenotypic correlation in performance across time in both reading skills and mathematics.

Although genes substantially contribute to correlated performance over time in literacy and numeracy, it does not necessarily follow that variation in growth of performance is primarily due to genetic influences. A possible scenario could be that genes maintain the relative performance of students in a way that is largely similar among individuals and environmental factors deviate growth from this trajectory. A scenario such as this would result in high phenotypic and genetic correlations between assessments but most of the variation in growth would arise from the environment. Although such a scenario is theoretically possible, it is more likely that both genetic and environmental factors will influence variation in growth of literacy and numeracy.

Biometric growth curve models have been published on three twin studies of reading: The Western Reserve Reading and Math Project (WRRMP; Logan et al., 2013; Petrill et al., 2010), The International Longitudinal Twin Study (ILTS; Christopher et al., 2013a; 2013b), and The Florida Twin Project on Reading (FTPR; Hart et al., 2013). All of these studies model growth of reading ability from the beginning of school to early or middle school and, where sufficient measurement times were available, showed a deceleration in rate of reading growth over time. There are some differences in the relative influence of genes and the shared environment among these studies. Both genes and the shared environment were significant contributors to variation in growth of oral reading fluency among participants in the FTPR (Hart et al., 2013). Similarly, both genes and the shared environment were significant contributors to variation in growth of word reading among participants in the WRRMP (Logan et al., 2013) and among Scandinavian participants in the ILTS (Christopher et al., 2013a). However, the shared environment was not a significant contributor to variation in growth of word reading among Australian or Colorado

participants in the ILTS (Christopher et al., 2013a; 2013b). For reading comprehension, only the shared environment substantially and significantly influenced growth among participants in the WRRMP (Logan et al., 2013), while genes, the shared environment, and the unique environment were comparable in size of influence among Colorado participants in the ILTS but only genes and the unique environment were significant (Christopher et al., 2013b). The only study to include a measure of spelling has been conducted on Colorado participants in the ILTS, again genetic, shared environmental, and unique environmental influences on growth were of a comparable size to each other, but only genes and the unique environment were significant (Christopher et al., 2013b). From these few studies, genes and the unique environment were generally important contributors to growth in reading while the shared environment was important in some samples (i.e. WRRMP, FTPR, ILTS Scandinavia).

The current study will first build on previous work by assessing the genetic contribution to stability of performance in reading, spelling, grammar and punctuation, writing, and numeracy in students from Grade 3 through to Grade 9. Thus far, genetic influences on stability in reading and spelling have been assessed in Australian children only from preschool to Grade 2 (Byrne et al., 2009; Byrne et al., 2007; Byrne et al., 2005), but based on findings from the USA and the UK in older students and the substantial contribution of genes to stability in performance in the younger Australian sample we expect that genes will largely mediate the longitudinal phenotypic correlations in reading and spelling from Grade 3 to Grade 9. To date, there are no studies on the longitudinal influence of genes on stability in grammar and punctuation or writing performance, although genes are substantial contributors to variation in both of these domains, with .58 for grammar and punctuation and .45 for writing (when averaged across all four grades; Grasby et al., 2015). Longitudinal genetic influence on mathematical performance has only been reported on students aged 7 through to 10 in the UK (Kovas et al., 2007); this age-span is similar to our first two grades of testing and we expect to replicate strong genetic contribution to stability in numeracy performance.

Secondly, the current study will expand research into the etiology of variation in growth by using biometric growth curve models to assess growth in several achievement domains not previously studied, namely grammar and punctuation, writing, as well as numeracy. Moreover, our model will assess if sex moderates the genetic and environmental influence on growth. Phenotypic growth models have found differences between reading and numeracy in growth (Baumert et al., 2012; Morgan et al., 2011), so it is feasible that the influence of genes and the environment might be different between the two. In contrast with the studies from the ILTS, WRRMP and FTRP where growth was modelled from the first year or two of schooling, Grade 3

is our initial year of testing, by which grade the rapid initial growth in reading performance has typically slowed (Bloom, Hill, Black, & Lipsey, 2008). Previous biometric growth studies that show substantial influence from the shared environment might do so due to variation in the early instruction in learning to read, which will not be captured in the current study. However, the tests in the current study are designed to assess achievement in line with educational curricula (Senate Standing Committee on Education and Employment, 2014); as such, we expect to see continued growth and variation in growth over the grades and it is possible that differences in teaching methods and school factors, which would mostly emerge as shared environment effects in a twin study, will contribute to growth beyond the initial rapid development in reading.

Method

Participants

Twins and triplets born from 1993–2006 were recruited through the voluntary Australian Twin Registry. For the 34 sets of triplets, a random pair from each set was selected for the analyses, and from hereon all multiple births are referred to as *twins*. Twins were invited to participate if they had sat (or would sit by 2014) a NAPLAN test. Of the 6853 families contacted, 2272 (33%) consented to participate. Of those who consented to participate, state departments provided NAPLAN results for 1949 families and of these zygosity information was reported from 1940 families. For the current study, twins who either skipped or repeated grades were not included, resulting in 1927 twins who sat NAPLAN tests in the same year and with 2-year intervals between tests.

Zygosity was determined with a short questionnaire (Lykken, Bouchard, McGue, & Tellegen, 1990), which classified a sub-sample of twins in this study with 95% accuracy when compared to parent report on DNA results. Our final sample of twin pairs included 865 monozygotic (MZ; 454 female, 411 male) and 1062 dizygotic (DZ; 301 female, 286 male, 475 opposite-sex). The number of twin pairs at Grades 3, 5, 7, and 9 respectively, with percentage providing data on multiple grades indicated in parentheses, was 1178 (66%), 1101 (98%), 990 (98%), and 809 (81%). Due to the introduction of the tests in 2008, data are not available for all participants at each grade, with some participants being too old to have sat NAPLAN in their early grades of school and others not yet old enough to have taken the later years. Of our sample there were 664 twin pairs with data on at least three grades, and 149 twin pairs with data on all four grades. The average age in Grade 3 at the time of testing was 8.6 years.

Chapter 4

Materials

National Assessment Program in Literacy and Numeracy

The NAPLAN is a nationwide, standardised assessment introduced in 2008 in Australia. Each year, students in Grades 3, 5, 7, and 9 sit tests in reading, writing, language conventions, and numeracy. The test content is based on the "Statements of Learning for English" and the "Statements of Learning for Mathematics," which inform state and territory curricula. For each achievement domain, raw scores are equated vertically (within a year and across grades) and horizontally (within a grade and across years), scaled against the historic distribution and transformed into a score on a common scale from 0–1000. This scaled score spans all years of the test and was designed to measure growth within cohorts and to compare across cohorts. Technical information and test administration details were obtained from the Australian Curriculum Assessment and Reporting Authority (2015a; 2015b; R. Randall, personal communication, July 10, 2013). Example test papers and writing prompts are available at www.nap.edu.au.

Reading

The reading comprehension test was composed of 7–8 passages. The passages were extracts or adaptations from books, newspaper articles, posters or poems. Passage length varies from brief single paragraphs of about 100 words, to several paragraphs of about 450 words in total. There were 5–8 items relating to any given passage. Most items are multiple-choice format, with a few short answer questions in each test. For Grades 3 and 5 there were 35–38 items to be completed in 45–50 minutes, and for Grades 7 and 9 there were 45–50 items to be completed in 65 minutes. A Cronbach's alpha of .85 (or above) for each test in each year indicates a high internal reliability.

Spelling

The spelling test presents misspelt words in simple sentences and requires students to identify and correct the spelling errors. For Grades 3 and 5 there were 23–25 items, and for Grades 7 and 9 there were 25–30 items. The spelling test is administered in the same paper as the grammar and punctuation test, and students are given 40–45 minutes to complete both of these question sets. For the spelling test, a Cronbach's alpha of .90 (or above) for each test in each year indicates a high internal reliability.

Grammar and punctuation

The grammar questions ask students to choose the correct word(s) to complete a sentence. This form of question is used in early grades to identify correct tense, pronouns, conjunctions, and verb forms. In later grades relative pronouns, clauses, and comparative adjectives are also

assessed. The punctuation questions ask students to insert or identify punctuation marks at the correct location in a sentence. For all grades there were 23–28 items. A Cronbach's alpha of .71 to .87 for each test in each year indicates an acceptable internal reliability (average .80).

Writing

The writing test is composed of a writing stimulus, which provides an idea or topic, and students are asked to write a response in a specified writing style (i.e. narrative, informative, or persuasive). For example, "It is cruel to keep animals in cages. What do you think? Do you agree or disagree? Perhaps you can think of ideas for both sides of this topic." The same prompt and style is used for all grade levels in a given year. Students have 40 minutes of writing time. Marks are awarded on 10 criteria: audience, text structure, ideas, vocabulary, cohesion, paragraphing, sentence structure, punctuation, spelling, and the final criterion depended on the writing style specified. For persuasive writing the criterion was persuasive devices (2011–2014), and for narrative writing the criterion was character and setting (2008–2010). From 2008–2010 the maximum score was 47, from 2011–2014 the maximum score was 48. A Cronbach's alpha, calculated using pooled data from all grades, of .93 (or above) for each test year indicates a high internal reliability; unfortunately we were unable to obtain inter-rater reliability information.

Numeracy

The numeracy test assesses five aspects of mathematics. *Working mathematically* includes problem solving, reasoning and interpretation. *Number* includes counting and computation. *Algebra, function and pattern* includes working with functions and relationships, graphs, equations, and rules. *Measurement, chance and data* includes working with units, likelihood and inference. *Space* includes shape and location. Most items are multiple-choice format, with a few short answer questions in each test. For Grade 7 and Grade 9, students sit a calculator-allowed and a non-calculator numeracy test. For Grades 3 and 5 there were 35–40 items to be completed in 45–50 minutes. For Grades 7 and 9 there were 62–64 items from the combined calculator and non-calculator papers, with each paper to be completed in 40 minutes. A Cronbach's alpha of .84 (or above) for each test in each year indicates a high internal reliability.

Procedure

After receiving parental consent, the state and territory departments of education provided NAPLAN test results. The NAPLAN tests are administered in the morning over three consecutive days each year in the second full week of May (approximately 3.5 months into the school year). On the first day the language conventions test (comprising of the spelling and grammar and punctuation domains) is administered and, after a minimum 20 min break, is followed by the

writing test. On the second day the reading test is administered. On the third day the numeracy tests are administered; for Grades 7 and 9 the first test permits use of a calculator and the second test does not. Support within specific constraints can be provided for students with disability, such as scribing or reading questions in the numeracy test. Across the nation 96% of students participate in the tests.

Analyses

Models were estimated using the scaled scores and full information maximum likelihood estimation in OpenMx, which uses all available data (Boker et al., 2011). Sex, age, age-squared, age-by-sex, and cohort effects have been found covary with mean performance, and sex has been shown to moderate the heritability in some of the domains and grades (details in Grasby et al., 2015). Thus, age, age-squared, age-by-sex, and cohort effects were regressed out of the scaled scores. Sex effects were tested within the latent growth curve model for each domain (detailed below).

Before obtaining genetic, shared environmental and unique environmental correlations (r_A, r_C, r_E respectively), sex-effects were tested within a sex-limited correlated factors model (Neale, Røysamb, & Jacobson, 2006). This model estimated the path loadings and correlations across time separately for each sex. Sex effects were first tested by equating all parameters to be equal across sex and significance was determined using the likelihood ratio test, which compares the difference in log likelihood from nested models to a $\chi 2$ distribution with degrees of freedom equal to the difference in estimated parameters from the nested models (Neale & Maes, 2004). If sex-effects were present, the correlational structure alone was tested for sex-effects by equating the correlations but allowing the path estimates to vary by sex; this effectively allowed for quantitative sex differences.

The biometric growth curve models used on the WWRP, FTPR, and ILTS data have some differences. In particular, the errors of the growth curve model of the ILTS data were allowed to correlate between twins in a pair (Christopher et al., 2013a; 2013b). Moreover, Christopher et al. (2013b) showed that models with correlated errors resulted in a better fit than models with uncorrelated errors. The errors in a longitudinal growth model capture the deviation in a score from the value predicted by the model; given non-independence within twin pairs it is likely that there is some interdependence between twin pairs in these error variances, such as stressful family-level events. Christopher et al. (2013b) provided a contrast of results from both models, and showed that the influence of the shared environment on growth in word reading was significant when estimated in the model with uncorrelated errors but was negligible when errors

were allowed to correlate. Thus, in the ILTS data, the biometric latent growth curves with uncorrelated errors inflated the effect of the shared environment on variation in growth. To reduce this possible source of bias in measuring variance in growth, in our model we will allow the errors between twin pairs to correlate.

The biometric latent quadratic growth curve model is depicted in Figure 4.1. The model estimates variation in performance at an intercept (I) and rate of growth through linear (S) and quadratic (Q) slopes (McArdle, Prescott, Hamagami, & Horn, 1998 discusses in detail). Variation in the intercept and slopes and their covariances are decomposed into additive (A), shared environmental (C), and unique environmental (E) sources. In line with classic twin design methodology (see Plomin, DeFries, Knopik, & Neiderhiser, 2013), genetic correlations between MZ twins are fixed to 1, while those of DZ twins are fixed to .5; shared environmental correlations are fixed to 1 for all twins. The slopes were centred at Grade 3, so the mean of the intercept and the mean of the linear slope respectively measure mean performance at and rate of growth from Grade 3. Deviation in scores from their growth curve are modelled as time-point specific error variance (u); these errors were allowed to correlate between members of a twin pair and this error correlation (r_U) was estimated separately for MZ and DZ twins to allow for possible genetic influences in the errors. Sex was modelled as a covariate on the latent intercept and slopes and as a moderator on the A, C, and E variance-covariance structure. For each domain, sex was dropped from the covariance structure and significance was determined using the likelihood ratio test.

Results

Descriptive statistics and longitudinal correlations

Descriptive statistics, intraclass correlations, heritability, shared and unique environmental estimates for each grade and subject are reported in Table 4.1. The pattern of change in means over time show scores increase at a decreasing rate in each domain. The heritability estimates were substantial and significant in all domains and grades, most shared environment estimates were small, and the unique environment estimates (including measurement error) were greatest in the writing domain. Phenotypic correlations were high among most grades for each domain, though the correlations across grades for writing were more moderate (see Table 4.2). Correlations were equally high for grades further apart in time as for adjacent grades.
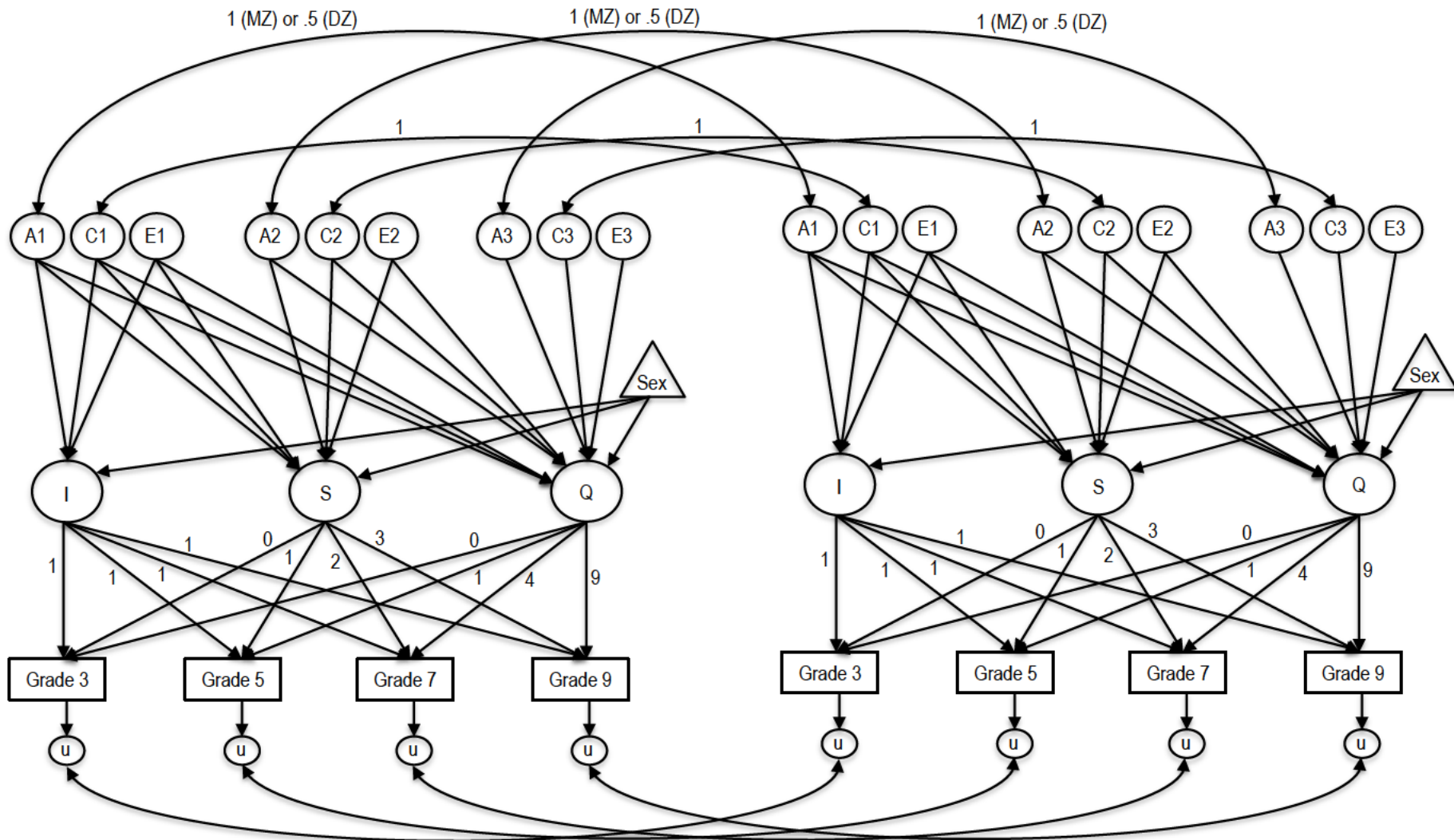
Figure 4.1 Biometric latent growth model. The left side of the figure represents Twin 1 and the right side represents Twin 2. Only fixed paths are labeled. Each path from latent additive (A1, A2, A3), shared environment (C1, C2, C3), and unique environment (E1, E2, E3) were moderated by sex, but is not depicted for lack of space. Correlations between the time-specific error variances (u) were equated across grades, but were allowed to vary between monozygotic (MZ) and dizygotic (DZ) twins. I = intercept; S = linear slope; Q = quadratic slope.

Table 4.1 Descriptive statistics, intraclass correlations by sex and zygosity, and heritability estimates for each domain and grade

																Intraclass correlations.							
	Descriptives			MZF		MZM		DZF		DZM		DZOS		Standardised variance estimates[a]									
Variable	M	SD	n	r	n	r	n	r	n	r	n	r	n	A		C		E					
Reading																							
Grade 3	447	85	1175	.74	261	.74	244	.41	181	.43	162	.42	299	.67*		.08*		.24*					
Grade 5	526	76	1098	.67	227	.67	220	.54	165	.49	170	.33	279	.58*		.11*		.30*					
Grade 7	582	66	988	.70	209	.71	201	.51	162	.47	152	.42	223	.55*		.18*		.27*					
Grade 9	634	64	809	.72	194	.72	169	.45	122	.45	119	.46	163	.56*		.19*		.25*					
Spelling																							
Grade 3	426	74	1177	.77	261	.76	244	.43	180	.47	163	.34	301	.73*		.05		.22*					
Grade 5	507	68	1099	.78	229	.75	221	.44	166	.48	171	.41	279	.77*		.03		.21*					
Grade 7	565	62	988	.78	212	.73	197	.43	161	.29	154	.35	216	.76*		.02		.21*					
Grade 9	610	63	808	.79	194	.72	165	.40	123	.27	120	.26	162	.76*		.01		.22*					
G & P																							
Grade 3	457	88	1177	.70	259	.64	241	.44	179	.49	161	.42	295	.51*		.19*		.30*					
Grade 5	532	83	1099	.71	230	.67	221	.46	165	.46	168	.33	278	.66*		.05		.29*					
Grade 7	579	74	989	.63	212	.69	199	.41	160	.37	152	.41	220	.57*		.11*		.32*					
Grade 9	615	71	806	.69	193	.62	165	.35	124	.45	122	.40	166	.52*		.16*		.32*					
Writing																							
Grade 3	436	58	1177	.49	261	.47	240	.28	177	.36	161	.30	297	.42*		.10*		.48*					
Grade 5	508	62	1098	.60	229	.53	219	.38	164	.37	167	.27	272	.47*		.11*		.43*					
Grade 7	546	69	989	.47	210	.50	197	.34	162	.25	151	.19	221	.41*		.08		.52*					
Grade 9	609	76	806	.52	193	.44	164	.17	123	.40	122	.20	166	.39*		.10*		.52*					
Numeracy															F	M	F	M	F	M			
Grade 3	424	70	1174	.74	258	.72	238	.52	182	.45	162	.39	297	.63*	.64*	.11	.13*	.26*	.23*			
Grade 5	517	70	1097	.79	228	.72	218	.53	166	.48	170	.43	279	.54*	.71*	.21*	.06	.25*	.23*			
Grade 7	574	71	987	.84	210	.79	199	.47	162	.44	151	.37	222	.79*	.73*	.04	.09	.17*	.17*			
Grade 9	632	69	802	.79	195	.77	165	.43	120	.43	114	.46	163	.59*	.61*	.20	.18*	.21*	.21*			

Note. MZ= monozygotic; DZ = dizygotic; OS = opposite-sex; F = female; M = male; *n* = number of twin pairs, for the correlations only the complete pairs are

counted; A = additive genetic; C = shared environment; E = unique environment.

[a] Estimates were obtained from the correlated factors model, estimates for males and females were significantly different for numeracy.

* significance calculated from 95% confidence intervals not including zero.

Table 4.2 Pearson correlations among the grades in each domain

Variable	Grade 3	Grade 5	Grade 7	Grade 9
Reading				
Grade 3	-			
Grade 5	.75	-		
Grade 7	.73	.73	-	
Grade 9	.70	.70	.77	-
Spelling				
Grade 3	-			
Grade 5	.83	-		
Grade 7	.81	.84	-	
Grade 9	.77	.79	.84	-
Grammar and Punctuation				
Grade 3	-			
Grade 5	.66	-		
Grade 7	.65	.67	-	
Grade 9	.65	.64	.68	-
Writing				
Grade 3	-			
Grade 5	.55	-		
Grade 7	.54	.53	-	
Grade 9	.44	.48	.53	-
Numeracy				
Grade 3	-			
Grade 5	.73	-		
Grade 7	.75	.81	-	
Grade 9	.75	.76	.83	-

Equating sex in the correlated factors model resulted in no significant loss of fit for reading, $\chi^2(62) = 53.3$, $p = .778$, spelling, $\chi^2(62) = 75.6$, $p = .115$, grammar and punctuation, $\chi^2(62) = 40.3$, $p = .985$, or writing, $\chi^2(62) = 57.9$, $p = .626$. By contrast, numeracy correlations could be equated across sex, $\chi^2(50) = 34.99$, $p = .947$, but not path loadings as well, $\chi^2(62) = 87.84$, $p = .017$. This indicated that the correlational structure was similar for females and males but the relative influence of genes and the environment on performance in numeracy was different. Accordingly, for all domains females and males were combined for analyses of correlations across time.

The genetic correlations within domain and across time were high; 95% confidence intervals typically included 1, and the estimates indicate that mostly the same genes were influencing performance at each grade level (see Table 4.3). Similarly, shared environmental correlations for reading were large and mostly significant, but for the other domains most estimates had wide confidence intervals that indicated non-significant estimates. The unique environmental correlations for reading, grammar and punctuation, and numeracy were modest but most were significant. For numeracy, they tended to be more substantial in between the later grades. Spelling had larger unique environmental correlations than the other domains (average .48). This indicates that generally some unique environmental factors were present and influential at multiple grades for each domain. However, for writing the unique environmental correlations were small and most

were non-significant. The mostly modest unique environmental correlations across grades suggest that much of the unique environment estimates at each grade level may have been caused by measurement error.

Table 4.3 Genetic, shared environmental, and unique environmental correlations between grades for each domain

Variable	Genetic			Shared environment			Unique environment		
	Grade 3	Grade 5	Grade 7	Grade 3	Grade 5	Grade 7	Grade 3	Grade 5	Grade 7
Reading									
Grade 3									
Grade 5	.93 [.88, 1]			.95 [.02, 1]			.24 [.14, .33]		
Grade 7	.92 [.85, 1]	.93 [.87, 1]		1 [.60, 1]	.99 [.58, 1]		.17 [.04, .30]	.24 [.13, .33]	
Grade 9	.85 [.74, 1]	.91 [.82, 1]	.95 [.91, 1]	1 [-.02, 1]	.97 [.38, 1]	1 [.81, 1]	.18 [-.04, .39]	.20 [.05, .33]	.31 [.21, .40]
Spelling									
Grade 3									
Grade 5	.96 [.92, 1]			.68 [-1, 1]			.43 [.34, .51]		
Grade 7	.88 [.84, .94]	.93 [.90, .96]		1 [-1, 1]	1 [-1, 1]		.54 [.44, .63]	.51 [.43, .58]	
Grade 9	.96 [.84, 1]	.94 [.86, 1]	.95 [.93, .99]	-1 [-1, 1]	-.75 [-1, 1]	1 [-1, 1]	.38 [.20, .54]	.48 [.35, .59]	.53 [.45, .60]
GP									
Grade 3									
Grade 5	.85 [.75, 1]			.94 [-.08, 1]			.21 [.11, .31]		
Grade 7	.86 [.76, 1]	.86 [.80, .97]		1 [.30, 1]	1 [-.58, 1]		.11 [-.02, .24]	.20 [.10, .29]	
Grade 9	1 [.67, 1]	.85 [.75, 1]	.93 [.87, 1]	.34 [-.82, 1]	1 [-.79, 1]	1 [.45, 1]	.24 [.06, .44]	.19 [.06, .31]	.22 [.11, .31]
Writing									
Grade 3									
Grade 5	.87 [.75, 1]			.95 [-.27, 1]			.09 [-.01, .19]		
Grade 7	.87 [.70, 1]	.83 [.70, 1]		1 [-1, 1]	1 [-1, 1]		.11 [-.02, .22]	.11 [0, .21]	
Grade 9	1 [.65, 1]	.65 [.44, .84]	1 [.83, 1]	-.50 [-1, .74]	1 [.26, 1]	.73 [-1, 1]	.06 [-.11, .22]	.18 [.05, .31]	.12 [.04, .21]
Numeracy									
Grade 3									
Grade 5	.95 [.86, 1]			.74 [-.03, 1]			.16 [.06, .26]		
Grade 7	.97 [.85, 1]	.93 [.89, 1]		.34 [-1, 1]	1 [-.83, 1]		.18 [.05, .32]	.37 [.28, .45]	
Grade 9	.95 [.80, 1]	.95 [.85, 1]	.99 [.92, 1]	.79 [-.24, 1]	.76 [-.73, 1]	.81 [-1, 1]	.24 [.02, .46]	.25 [.13, .36]	.40 [.31, .48]

Note. 95% confidence intervals are in brackets

Biometric quadratic growth curve

Sex could be dropped as a moderator on the covariance structure without significant loss of fit for reading, $\chi^2(18) = 24.06$, $p = .153$, and grammar and punctuation, $\chi^2(18) = 18.35$, $p = .433$. However, sex could not be dropped for spelling, $\chi^2(18) = 40.26$, $p = .002$, writing, $\chi^2(18) = 35.40$, $p = .008$, or numeracy, $\chi^2(18) = 61.54$, $p < .001$, indicating sex differences in the quantitative contribution of genes and the environment to variation in growth of performance in spelling, writing, and numeracy. Significant sex differences were found on the intercept in all domains; sex was coded with females as 0 and males as 1 so the estimates in Table 4.4 indicate that girls scored higher on the literacy domains and boys scored higher on numeracy. Sex effects on the linear slope indicated that growth in performance for boys was significantly steeper than for girls in reading, grammar and punctuation, and numeracy. The only significant sex-effect on the quadratic slope was for boys in numeracy, it indicated that the rate of deceleration in growth was greater for boys. In combination these sex-effects indicate that for reading and grammar and punctuation, girls scored higher than boys in Grade 3 but this sex-effect diminished over time as the boys had a steeper rate of growth. For writing and spelling girls performed higher than boys by a relatively

stable margin over the grades. For numeracy, boys scored higher than girls in Grade 3 and this effect increased in size as the grades progressed before beginning to taper off in the later grades.

Variation in the intercept and slopes are reported in Table 4.4. All domains are scored on a scale of 0–1000 in each grade. As such, writing has a noticeably constricted variation in performance on the intercept when compared to the other domains, and grammar and punctuation has a markedly higher variance on the linear slope. For spelling, writing, and numeracy, where sex could not be dropped from the covariance structure, boys have greater variance than girls on all intercepts and slopes.

Table 4.4 Estimated mean, variance and sex-effect on intercept, linear and quadratic slopes, and error correlations for each domain

Variable	Intercept				Linear slope				Quadratic slope				r_U	
	Mean	Sex	Variance		Mean	Sex	Variance		Mean	Sex	Variance		MZ	DZ
Reading	456	-13.9*	5848		76*	8.0*	610		-5.3*	-1.6	25		.17*	.05
Spelling	435	-16.6*	420 F	5051 M	84*	4.1	174 F	468 M	-8.2*	-0.6	15 F	34 M	.26*	.10*
GP	467	-20.6*	5699		77*	8.2*	1238		-8.5*	-2.0	108		.28*	.11*
Writing	447	-20.9*	1542 F	2114 M	67*	0.8	161 F	404 M	-3.6*	-0.7	41 F	62 M	.14*	.06
Numeracy	422	11.3*	3091 F	4166 M	87*	9.3*	476 F	987 M	-6.8*	-2.5*	82 F	121 M	.35*	.21*

Note. Sex is coded with females = 0 and males = 1, to obtain male estimates adjust the main effect by the sex-effect. r_U = error correlation between twins

F = female, M = males

* significance calculated from 95% confidence intervals not including zero.

Correlations between errors of twin pairs were significant in each domain. Most were small in size, indicating that most of the time-point-specific error variances are uncorrelated. However, correlations were larger between MZ twins than DZ twins, and could only be equated in writing without a significant loss of model fit. This indicated some genetic factors contribute to the deviation of scores from that predicted by the growth curve.

Genetic and environmental influences on growth

Standardised A, C, and E variance components of the intercept, linear and quadratic slopes are reported on the diagonal in Table 4.5 for each domain. On the off diagonal are reported the standardised A, C, and E covariances among the intercept and slopes. On the intercept, genetic influences were significant for each domain, estimates ranged from .50 for numeracy performance among girls to .88 for spelling performance among girls. Shared environmental influences were modest and significant for reading (.10), grammar and punctuation (.23), and writing among boys (.19), and were substantial for numeracy among girls (.42). The unique environmental influence on the intercept for each domain was small and most were significant, with estimates ranging from .08 for numeracy among girls to .20 for writing among boys.

Chapter 4

Table 4.5 Estimates of standardised A, C, E variance and covariance components of the intercept, linear and quadratic slopes for each domain.

Domain	Variable	A1	A2	A3	C1	C2	C3	E1	E2	E3
Reading										
	Intercept	.79 [.66, .91]*			.10 [.02, .22]*			.10 [.04, .18]*		
	Linear	-.59 [-.92, -.31]*	.55 [.15, 1]*		.03 [-.22, .14]	.01 [0, .61]		-.12 [-.30, .09]	.44 [0, .81]	
	Quadratic	.48 [.12, .95]*	-.40 [-.99, -.05]*	.30 [.02, 1]*	-.01 [-.21, .39]	0 [-.69, .10]	0 [0, .92]	.09 [-.21, .34]	-.48 [-.85, .26]	.69 [0, .96]
Spelling										
Female	Intercept	.88 [.60, .93]*			.01 [0, .28]			.11 [.07, .16]*		
	Linear	-.57 [-.96, .03]	.53 [0, 1]		-.02 [-.50, .12]	.37 [0, .99]		-.06 [-.28, .12]	.10 [0, .88]	
	Quadratic	.35 [-.18, .96]	-.30 [-.95, .20]	.17 [0, 1]	.04 [-.18, .50]	-.24 [-.91, .59]	.41 [0, 1]	.03 [-.21, .36]	-.18 [-.88, .31]	.42 [0, .99]
Male	Intercept	.68 [.44, .89]*			.17 [0, .41]			.14 [.10, .20]*		
	Linear	-.01 [-.48, .29]	.24 [0, .97]		-.33 [-.59, .07]	.63 [0, 1]		-.09 [-.25, .03]	.14 [0, .76]	
	Quadratic	-.19 [-.47, .35]	-.03 [-.89, .25]	.06 [0, .99]	.31 [-.19, .53]	-.61 [-.94, .09]	.62 [0, 1]	.11 [-.04, .36]	-.20 [-.83, .05]	.32 [0, 1]
Grammar and Punctuation										
	Intercept	.63 [.46, .83]*			.23 [.07, .38]*			.14 [.04, .24]*		
	Linear	-.05 [-.38, .18]	0 [0, .75]		-.20 [-.53, .02]	.24 [0, 1]		-.23 [-.38, .08]	.76 [0, .96]	
	Quadratic	-.04 [-.30, .22]	0 [-.63, .05]	0 [0, .64]	.14 [-.06, .47]	-.13 [-.89, .03]	.09 [0, .92]	.22 [-.08, .39]	-.80 [-.95, .05]	.91 [.05, 1]*
Writing										
Female	Intercept	.62 [.24, .99]*			.26 [0, .60]			.13 [0, .35]		
	Linear	.11 [-.42, .96]	.02 [0, 1]		.32 [-.28, .73]	.64 [0, 1]		-.11 [-.52, .30]	.34 [0, .98]	
	Quadratic	.09 [-.61, .51]	.01 [-.85, .35]	.07 [0, .98]	-.21 [-.61, .25]	-.42 [-.91, .20]	.27 [0, 1]	-.08 [-.43, .40]	-.32 [-.96, .88]	.66 [0, 1]
Male	Intercept	.61 [.24, .91]*			.19 [.01, .54]*			.20 [.03, .37]*		
	Linear	.23 [-.47, .93]	.29 [0, 1]		-.12 [-.51, .60]	.22 [0, 1]		-.25 [-.53, .38]	.49 [0, 1]	
	Quadratic	-.16 [-.56, .43]	-.21 [-.93, .30]	.15 [0, .98]	.05 [-.49, .42]	-.33 [-.96, .92]	.63 [0, 1]	.21 [-.09, .50]	-.25 [-.87, .63]	.22 [0, .88]
Numeracy										
Female	Intercept	.50 [.30, .79]*			.42 [.14, .60]*			.08 [.03, .16]*		
	Linear	.27 [-.04, .49]	.17 [0, .82]		-.25 [-.50, .16]	.41 [0, .88]		-.10 [-.27, .04]	.42 [.03, .97]*	
	Quadratic	-.12 [-.33, .14]	-.09 [-.72, .03]	.06 [0, .68]	.15 [-.17, .43]	-.37 [-.79, .02]	.36 [0, .75]	.10 [-.03, .25]	-.50 [-.94, -.07]*	.59 [.20, .97]*
Male	Intercept	.83 [.64, .94]*			.07 [0, .25]			.10 [.04, .17]*		
	Linear	.04 [-.20, .32]	.24 [0, .76]		-.03 [-.28, .10]	.02 [0, .63]		-.08 [-.23, .10]	.74 [.23, 1]*	
	Quadratic	-.06 [-.29, .16]	-.15 [-.61, .03]	.09 [0, .54]	.03 [-.10, .23]	-.04 [-.54, .03]	.09 [0, .51]	.07 [-.08, .20]	-.75 [-.97, -.30]*	.82 [.44, 1]*

Note. 95% confidence intervals are in brackets. A1, A2, A3 = genetic; C1, C2, C3 = shared environment; E1, E2, E3 = unique environment
* significance calculated from 95% confidence intervals not including zero.

Genetic influences were significant on both linear and quadratic slopes for reading; 55% of the variation in linear growth and 30% of the variation in quadratic deceleration was accounted for by genetic influences. Almost all of the genetic influence on the linear slope was shared with that of the intercept. The genetic correlation between the intercept and linear slope was -.92 (calculated from the genetic covariance divided by the square root of the product of the genetic variance of the intercept and the linear slope [$-.59/\text{sqrt}(.79 \times .55) = -.90$]. Likewise, essentially the same genes were influencing variation between the intercept and quadratic slope (genetic correlation = .98), and between linear and quadratic slopes (genetic correlation = -.97). Negative covariation between intercept and linear slope and positive covariation between intercept and quadratic slope indicated that (on average) individuals who scored higher in Grade 3 had a slower rate of linear growth and faster deceleration than individuals who scored lower in Grade 3. The high genetic correlations indicate that essentially the same genetic factors that influence individuals to score higher in Grade 3 also contribute to less growth. No genetic influences on either linear or quadratic slope reached significance in any other domain.

Shared environmental influences on slopes did not reach significance in any domain. The unique environment accounted for 42% of the variation in linear slope and for 59% of the variation in quadratic slope among girls in numeracy. Boys were more strongly influenced, with the unique environment contributing to 74% of the variation in linear slope and for 82% of variation in quadratic slope. Essentially the unique environmental influences on the slopes were shared (high genetic correlations of -1.0 for girls and -.97 for boys), but there was no significant overlap in the unique environmental influences between the intercept and either slope. This indicated that the unique environmental factors that contributed to variation in numeracy in Grade 3 were different from those that influenced growth, but those factors that contributed to steeper growth also contributed to faster deceleration. Unique environmental influences on the slopes did not reach significance in any other domain, except for the quadratic slope in grammar and punctuation, which indicated that most of the deceleration in growth was due to the unique environment.

Discussion

The purpose of this paper was to examine the relative influence of genes and the environment on longitudinal stability and growth in performance on various measures of literacy and numeracy in Australian school students through the middle years of formal education. Consistent with research into the development of reading skills, genetic factors were the strongest contributor to stability in performance over time, not only in reading, but in all domains. Genetic

factors were also the strongest influence on variation in growth of reading, with the same genes that influenced variation in performance at Grade 3 also influencing growth. Numeracy results contrasted with reading, such that unique environmental factors influenced growth, and these factors were different to those that influenced variation in numeracy performance at Grade 3. Unfortunately, the results for growth in spelling, grammar and punctuation, and writing were ambiguous, with estimates from neither genes nor the environment reaching significance.

Longitudinal stability

Strong phenotypic correlations across the four grades for reading, spelling, grammar and punctuation, and numeracy indicated a high level of stability in relative performance over time. Reading, numeracy, and particularly spelling correlations were very high (respectively .73, .77, and .81 averaged across all grade comparisons), and genes mediated most of this stability in performance (74% for reading, 81% for numeracy, and 87% for spelling). Grammar and punctuation phenotypic correlations were a little lower (.66 on average), but were also predominantly mediated by genes (76%). The reported internal reliabilities for the grammar and punctuation test were more variable from year-to-year than the other domains, and greater measurement error in the test might account for slightly lower phenotypic correlations between grades.

Performance in writing was somewhat less stable than the other domains, with phenotypic correlations among grades of .51 (on average). In any given grade, the unique environment accounted for as much of the variation in performance as genes. However, small unique environmental correlations between grades and high genetic correlations resulted in genes mediating most of the phenotypic correlations in writing (72% on average). This strong influence from the unique environment with low unique environmental correlations over time indicates a lack of continuity in the substantial influence of the unique environment. This is consistent with measurement error, but it could also result from genuine time-specific unique influences in the writing tests. The tests require students to develop a coherent argument or narrative based on a prompt; as such, it is reasonable to expect individuals to produce more creative or inspired work in some years compared to other years depending on their personal experiences and interests. The strength of genetic mediation on the stability in writing performance might stem from some of the foundation skills of writing, such as vocabulary, spelling, grammar, and punctuation. Stability of performance in these foundational skills has been shown in previous work (Olson et al., 2011; Samuelsson et al., 2008) and in this current study to be strongly mediated by genes.

Genetic correlations near to unity indicate that, for each domain, essentially the same genes were influencing performance across the different grades. For reading, this finding is consistent with results from simplex models conducted on reading assessments in the FTPR (Hart et al., 2013), the ILTS and WRRMP (Soden et al., 2015). These studies all found genetic factors at Grade 1 continued to influence performance through all the grades assessed (Grades 4, 5, or 6 depending on the study duration), but none found significant novel genetic influences after Grade 3. Our findings extend this evidence of genetic stability in performance to other literacy domains and to numeracy, although, in contrast to our results and those from the USA, novel genetic influences on later ages have been found for both reading and mathematics in children assessed as ages 7, 9 and 10 in the UK (Kovas et al., 2007). The differences in these findings might be linked to the different forms of assessment employed across these studies, teacher ratings in the UK study, oral reading fluency in FTPR, and reading comprehension in the ILTS, WRRMP, and the current study. Fundamentally, for these NAPLAN data the genetic variation among students in Grade 3 continues to influence performance and contribute to relative stability in performance through to Grade 9.

Growth

In each domain, genetic variation was the strongest contributor to the intercept of the growth function. The heritability of these intercepts was slightly higher than general estimates of heritability of performance in Grade 3, because some of the total variance in Grade 3 is modelled as error variance in the latent growth curve. For reading, girls scored higher than boys in Grade 3 but boys had a faster rate of growth than girls resulting in a reduced sex-effect on mean scores over time. Despite this, there were no significant sex differences in the relative contribution of genes and the environment to this difference in growth. Genes were the only significant contributor to individual differences in growth in reading, and these were the same genes that influenced performance at Grade 3. While a substantial portion of variation in the slopes was attributed to the unique environment, wide confidence intervals around the parameters meant this did not translate into statistical significance.

The negative covariation between the intercept and linear slope and positive covariation between the intercept and quadratic slope are consistent with a compensatory model of reading development, such that students with poorer initial performance catch up a little to students who are better readers. Moreover, our results show that this gain is primarily due to genetic factors. In Pfost et al.'s (2014) review, a compensatory developmental pattern—such as this—was associated more strongly with constrained reading skills. Constrained skills are skills universally mastered,

such as letter knowledge or phonics, and individuals primarily differ in their age of acquisition and duration until mastery (Paris, 2005). Interestingly, reading comprehension, which is the skill principally assessed in these NAPLAN data, is not a constrained reading skill. However, in Pfost et al.'s review reading comprehension was also associated more with compensatory growth than a stable or an increasing achievement gap. It is possible that the compensatory growth observed in reading comprehension is due to the compensatory growth pattern in the constrained reading skills that are precursors to reading comprehension. These precursors, or codependent reading skills, to reading comprehension are substantially influenced by genetic variability in the early years of school (Byrne et al., 2005; Petrill et al., 2007). Given the genetic nature of the compensatory growth in NAPLAN reading, it seems that this relative improvement in poorer readers is not due to educational intervention but due to an inherent developmental delay in the mastery of necessary reading skills.

In contrast to reading, boys scored higher than girls in Grade 3 numeracy and had a steeper linear growth and greater deceleration than girls. This resulted in a sharp increase in sex differences from Grade 3 to 5, which levelled off from Grade 5 onwards. Although the relative influence of genes and the environment on growth could not be equated between girls and boys, the difference was predominantly on performance at initial testing, with the shared environment having a significant influence on girls but not on boys. The influence of the shared environment on growth was higher in girls than boys, but did not reach statistical significance. It was the unique environment that significantly influenced variation in growth of numeracy performance in both girls and boys. Moreover, these unique environmental influences on growth were independent from unique environmental influences on performance in Grade 3. This might represent the impact of different teachers or changing classes over time. Unlike reading there was no evidence of compensatory growth. Instead, the absence of significant covariation between initial performance and growth indicated a relatively stable achievement gap between the highest and lowest achieving students over time.

Results for grammar and punctuation had some similarity to reading; girls scored higher than boys in Grade 3 and boys had a faster rate of growth than girls, resulting in a lessening in achievement gap between girls and boys over time. Furthermore, there were no significant sex differences in the relative contribution of genes and the environment to growth. Like reading, genes, the shared environment and the unique environment all made significant contributions to variation in performance at initial testing. However, the results differed from reading regarding the influences on growth. For grammar and punctuation the unique environment contributed the most to variation in growth, but this was only significant for the quadratic, or deceleration, effect.

The lack of significance for the estimate of the unique environment on the linear slope seems unusual. Grammar and punctuation had the greatest variation in both linear and quadratic slopes, primarily due to shared and unique environment factors. However, there appears to be considerable error around the estimates in this model. These estimates are derived from the covariation in growth of MZ twins relative to DZ twins. The large margin of error around these estimates suggests there is considerable disparity in the covariation of growth for MZ and DZ twins, at least for grammar and punctuation.

Of all the domains, spelling and writing had the most restricted variation in growth, with considerably less variation in girls than boys. As with the other literacy skills, girls scored higher than boys in Grade 3, but there were no significant sex differences in slopes, indicating a relatively stable achievement gap in spelling and writing over time. In each domain girls and boys significantly differed in the relative contribution of genes and the environment to their growth curves. For spelling, the only significant estimates were on the intercept, where genes predominantly influenced both girls and boys with a smaller effect from the unique environment. The pattern of significant estimates was not different between the sexes, but in the size of the estimates there was greater influence of genetic factors on growth in girls and of shared environment factors on growth in boys. Similarly for writing, the only significant estimates were on the intercept. Genetic influences were comparable for girls and boys on variation in initial performance, and both the shared and unique environments also significantly influenced boys. While the size of the relative contribution of the shared environment was similar for girls it did not reach significance, possibly due to more restricted variation in girls and smaller unstandardised estimates. As with spelling, in neither sex were any estimates on linear or quadratic slope significant. For these two domains, the lack of significant estimates of genes and the environment on growth might result from less total variation in growth and the concomitant small size of any specific estimate, which we lack power to detect.

Implications

These results suggest a tempering of the claim that growth in NAPLAN performance reflects the value added by the school. The timing of the NAPLAN tests in the school year and biennial administration means that growth in performance cannot be considered an accurate measure of teacher or class effect, but growth has been suggested to reflect a school effect. Most of our twins attended the same school (95%, 98%, 95%, and 92% in Grades 3, 5, 7, and 9 respectively); as such, the effect of variation in school would predominantly be a shared environment effect. The low estimate of the shared environment on variation in growth in reading

and numeracy does not support use of NAPLAN growth measures as an index of the value added to performance by the school. Shared environmental estimates for reading and numeracy in boys were particularly small. This is not to assert that schools are unimportant. Quite the contrary, instruction is necessary for students to learn to read, write, and enumerate, and this is evident in the mean growth in these skills. Instead, these results imply that attendance of a "good" or a "poor" school is not going to be the most substantial influence on whether improvement in performance on these NAPLAN tests is faster, slower, or reaches a plateau sooner instead of later.

Limitations

The amount of error around the linear and quadratic slope estimates limits the conclusions we can draw. Although several key biometric growth curve studies have previously reported wide confidence intervals, our sample size was comparable to the largest study (i.e. FTPR), which reported relatively tight confidence intervals. Hart et al. (2013) detected small environmental effects on linear and quadratic slopes of 5% and 2% (respectively) as significant. As such, the amount of error around our estimates was unexpected. In the FTPR study no total measure of variance on these slopes was reported, so it is possible that there were effects of greater size to be detected in the FTPR than in our NAPLAN data. It is possible that small effects in our data contribute to the wide confidence intervals in some of the domains; however, lack of variation in growth does not explain the lack of significance in our grammar and punctuation as it also had the greatest variation in growth of all the domains. Only 34% of our sample contributed to three or more time-points, which might contribute to some of the error in the growth curve estimates; although, latent growth structural equation models have been shown to obtain accurate estimates with a similar proportion of randomly missing data (McArdle & Hamagami, 1992). It is possible that the two-year gap between tests contributed to less systematic covariation of growth curves between twins and therefore greater error in the biometric estimates. Growth curve models have been reported on reading and mathematics achievement in past research with intervals between tests of two years (Grimm, Steele, Mashburn, Burchinal, & Pianta, 2010; Morgan et al., 2011), but these have not then decomposed the variance on growth into genetic and environmental effects. Christopher et al. (2013b) reported on a two-year gap in the ILTS Colorado data from Grade 2 to Grade 4, but the first three time-points measured were only a year apart with only the final gap at two years, perhaps resulting in greater covariation in curves.

Conclusions

There were two main goals of this paper: to assess the relative genetic and environmental influences on (a) stability and (b) growth in literacy and numeracy in Australian school students.

Genes were the predominant influence on stability in performance in reading, spelling, grammar and punctuation, writing, and numeracy. Phenotypic correlations were high among all grades, and genes were the principal mediator (78% on average) of this stability in performance. Genes were also the main influence on growth in reading. Essentially the same genes that contributed to variation in performance at initial testing also influenced growth in reading, and in such a way that those who performed poorer at Grade 3 closed the achievement gap a little in subsequent grades. Sex effects on growth were also evident in reading, such that boys who scored more poorly at initial testing had a faster rate of growth and closed the achievement gap with girls as the grades progressed. Quite different results were found in numeracy, with boys outperforming girls at initial testing and the achievement gap first increasing before starting to reduce in the later years. In contrast to reading, the environment contributed predominantly to individual differences in growth in numeracy. The results were largely inconclusive on the relative contribution of genes and the environment on variation in growth for the domains of spelling, grammar and punctuation, and writing.

Acknowledgements

References

Australian Curriculum Assessment and Reporting Authority. (2015a). *National Assessment Program—Literacy and Numeracy 2014: Technical report*

Australian Curriculum Assessment and Reporting Authority. (2015b). *National protocols for test administration*. Retrieved from http://www.nap.edu.au/naplan/school-support/national-protocols.html

Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools? *Child Development, 83*(4), 1347-1367. doi:10.1111/j.1467-8624.2012.01779.x

Betjemann, R., Willcutt, E., Olson, R., Keenan, J., DeFries, J., & Wadsworth, S. (2008). Word reading and reading comprehension: Stability, overlap and independence. *Reading and Writing, 21*(5), 539-558. doi:10.1007/s11145-007-9076-8

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 289-328. doi:10.1080/19345740802400072

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306-317. doi:10.1007/s11336-010-9200-6

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., . . . DeFries, J. C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to grade 2. *Journal of Neurolinguistics, 22*(3), 219-236. doi:10.1016/j.jneuroling.2008.09.003

Byrne, B., Coventry, W. L., Olson, R. K., Wadsworth, S. J., Samuelsson, S., Petrill, S. A., . . . Corley, R. (2010). "Teacher effects" in early literacy development: Evidence from a study of twins. *Journal of Educational Psychology, 102*(1), 32-42. doi:10.1037/a0017288

Byrne, B., Samuelsson, S., Wadsworth, S., Hulslander, J., Corley, R., DeFries, J., . . . Olson, R. (2007). Longitudinal twin study of early literacy development: Preschool through Grade 1. *Reading and Writing, 20*(1), 77-102. doi:10.1007/s11145-006-9019-9

Byrne, B., Wadsworth, S., Corley, R., Samuelsson, S., Quain, P., DeFries, J. C., . . . Olson, R. K. (2005). Longitudinal twin study of early literacy development: Preschool and kindergarten phases. *Scientific Studies of Reading, 9*(3), 219-235. doi:10.1207/s1532799xssr0903_3

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., . . . Olson, R. K. (2013a). The genetic and environmental etiologies of individual differences in early reading growth in Australia, the United States, and Scandinavia. *Journal of Experimental Child Psychology, 115*(3), 453-467. doi:10.1016/j.jecp.2013.03.008

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., . . . Olson, R. K. (2013b). Modeling the etiology of individual differences in early reading development: Evidence for strong genetic influences. *Scientific Studies of Reading*, 1-19. doi:10.1080/10888438.2012.729119

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research, 76*(1), 1-62. doi:10.3102/003465430760010001

de Zeeuw, E. L., de Geus, E. J. C., & Boomsma, D. I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education, 4*(3), 69-76. doi:10.1016/j.tine.2015.06.001

Grasby, K. L., Coventry, W. L., Byrne, B., Olson, R. K., & Medland, S. E. (2015). Genetic and environmental influences on literacy and numeracy performance in Australian school children in Grades 3, 5, 7, and 9. *Manuscript submitted for publication*.

Grimm, K. J., Steele, J. S., Mashburn, A. J., Burchinal, M., & Pianta, R. C. (2010). Early behavioral associations of achievement trajectories. *Developmental Psychology, 46*(5), 976-983. doi:10.1037/a0018878

Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development, 78*(1), 116-131. doi:10.1111/j.1467-8624.2007.00988.x

Hart, S. A., Logan, J. A. R., Soden-Hensler, B., Kershaw, S., Taylor, J., & Schatschneider, C. (2013). Exploring how nature and nurture affect the development of reading: An analysis of the Florida Twin Project on Reading. *Developmental Psychology*. doi:10.1037/a0031348

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

Kovas, Y., Haworth, C. M. A., Dale, P. S., Plomin, R., Weinberg, R. A., Thomson, J. M., & Fischer, K. W. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development, 72*(3), i-156. doi:10.2307/30163178

Logan, J. A. R., Hart, S. A., Cutting, L., Deater-Deckard, K., Schatschneider, C., & Petrill, S. (2013). Reading development in young children: Genetic and environmental influences. *Child Development, 84*(6), 2131-2144. doi:10.1111/cdev.12104

Lykken, D. T., Bouchard, T. J., McGue, M., & Tellegen, A. (1990). The Minnesota Twin Family Registry: Some initial findings. *Acta Geneticae Medicae et Gemellologiae, 39*(1), 35-70. doi:10.1017/S0001566000005572

Masters, G., Rowley, G., Ainley, J., & Khoo, S. (2008). *Reporting and comparing school performances*. Retrieved from http://research.acer.edu.au/ar_misc/8

McArdle, J. J., & Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental Aging Research, 18*(3), 145-166. doi:10.1080/03610739208253917

McArdle, J. J., Prescott, C. A., Hamagami, F., & Horn, J. L. (1998). A contemporary method for developmental‐genetic analyses of age changes in intellectual abilities. *Developmental Neuropsychology, 14*(1), 69-114. doi:10.1080/87565649809540701

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44*(5), 472-488. doi:10.1177/0022219411414010

Neale, M. C., & Maes, H. H. (2004). *Methodology for the genetic studies of twins and families* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Neale, M. C., Røysamb, E., & Jacobson, K. (2006). Multivariate genetic analysis of sex limitation and G x E interaction. *Twin Research and Human Genetics, 9*(4), 481-489. doi:10.1375/183242706778024937

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257. doi:10.2307/3699577

Olson, R. K., Keenan, J. M., Byrne, B., & Samuelsson, S. (2014). Why do children differ in their development of reading and related skills? *Scientific Studies of Reading, 18*(1), 38-54. doi:10.1080/10888438.2013.800521

Olson, R. K., Keenan, J. M., Byrne, B., Samuelsson, S., Coventry, W. L., Corley, R., . . . Hulslander, J. (2011). Genetic and environmental influences on vocabulary and reading development. *Scientific Studies of Reading, 15*(1), 26-46. doi:10.1080/10888438.2011.536128

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184-202. doi:10.2307/4151679

Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., DeThorne, L., & Vandenbergh, D. J. (2007). Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. *Reading and Writing, 20*(1-2), 127-146. doi:10.1007/s11145-006-9021-2

Petrill, S. A., Hart, S. A., Harlaar, N., Logan, J., Justice, L. M., Schatschneider, C., . . . Cutting, L. (2010). Genetic and environmental influences on the growth of early reading skills. *Journal of Child Psychology and Psychiatry, 51*(6), 660-667. doi:10.1111/j.1469-7610.2009.02204.x

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on matthew effects in reading. *Review of Educational Research, 84*(2), 203-244. doi:10.3102/0034654313509492

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral genetics*. New York: Worth.

Samuelsson, S., Byrne, B., Olson, R. K., Hulslander, J., Wadsworth, S., Corley, R., . . . DeFries, J. C. (2008). Response to early literacy instruction in the United States, Australia, and Scandinavia: A behavioral-genetic analysis. *Learning and Individual Differences, 18*(3), 289-295. doi:10.1016/j.lindif.2008.03.004

Senate Standing Committee on Education and Employment. (2014). *Effectiveness of the National Assessment Program – Literacy and Numeracy*. Canberra: Senate Printing Unit Retrieved from http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Empl oyment/Naplan13/Report/index.

Shin, T., Davison, M. L., Long, J. D., Chan, C.-K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences, 23*, 92-100. doi:10.1016/j.lindif.2012.10.002

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453. doi:10.2307/3515987

Smyth, E., Whelan, C., McCoy, S., Quail, A., & Doyle, E. (2010). Understanding parental influence on educational outcomes among 9 year olds in Ireland: The mediating role of resources, attitudes and children's own perspectives. *Child Indicators Research, 3*(1), 85-104. doi:10.1007/s12187-009-9051-9

Soden, B., Christopher, M. E., Hulslander, J., Olson, R. K., Cutting, L., Keenan, J. M., . . . Petrill, S. A. (2015). Longitudinal stability in reading comprehension is largely heritable from Grades 1 to 6. *PloS One, 10*(1), e0113807. doi:10.1371/journal.pone.0113807

Stankov, L., Morony, S., & Lee, Y. P. (2014). Confidence: the best non-cognitive predictor of academic achievement? *Educational Psychology, 34*(1), 9-28. doi:10.1080/01443410.2013.814194

Wadsworth, S. J., Corley, R. P., Hewitt, J. K., & DeFries, J. C. (2001). Stability of genetic and environmental influences on reading performance at 7, 12, and 16 years of age in the Colorado Adoption Project. *Behavior Genetics, 31*(4), 353-359. doi:10.1023/a:1012218301437

Higher Degree Research Thesis by Publication
University of New England

STATEMENT OF AUTHORS' CONTRIBUTION

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research Master/PhD candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated in the *Statement of Originality*.

	Author's Name (please print clearly)	**Contribution**
Candidate	Katrina Grasby	Responsible for study design; data cleaning and analysis; writing of paper
Other Authors	William Coventry	Provided editorial comments; author of grant, which provided data for paper

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry

28/1/1
Date

28/1/16
Date

Higher Degree Research Thesis by Publication
University of New England

STATEMENT OF ORIGINALITY

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research Master/PhD candidate and the candidate's Principal Supervisor, certify that the following text, figures and diagrams are the candidate's original work.

Type of work	Page number/s
All text, tables, and figures	Chapter 4, pp 86–112

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry

28/1/16
Date

29/1/16
Date

105
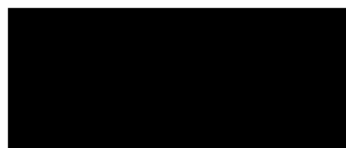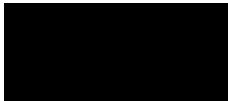
Chapter 5
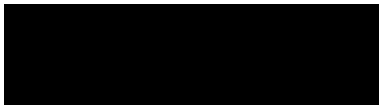
Socioeconomic Status Does Not Modify Heritability of Literacy and Numeracy in Australia

Katrina L. Grasby and William L. Coventry

Candidate Signature

Principal Supervisor Signature

Chapter 5

Once the province of the wealthy, the 19th century heralded movements toward systemic education in European countries and the United States of America (USA; Meyer, Tyack, Nagel, & Gordon, 1979; Ramirez & Boli, 1987). In the Australian colonies, from the mid-1800s there was a push for national, secular and free education (Crane, 1951). Compulsory primary education was introduced into New South Wales with the Public Institution Act of 1880, but high school education remained academically selective and largely restricted to families who could afford the fees until reforms in 1911-12 (Proctor, 2007). With education available to children from a broad range of economic backgrounds, debate on the influence of nature and nurture on the mental ability of children moved from theoretical to empirical. To put to rest arguments based on specific individuals who exhibited genius despite their origins, in 1918 Kornhauser published a study demonstrating a quantitative relationship between greater wealth with greater school achievement. Interest in understanding the relationship between socioeconomic status (SES) and academic ability continued and meta-analyses on research conducted throughout the twentieth century found higher SES was related to better academic outcomes, although the size of this effect was modest ($r = .27$, Sirin, 2005; $r = .22$, White, 1982). Almost all of the studies included in these meta-analyses have been conducted on children in the USA, however the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) has found an association between SES and academic achievement in most OECD countries (OECD, 2013). Although the strength of the relationship between SES and academic achievement differs across OECD countries, it was similar in strength between Australia (12.3% of the variance) and the United Kingdom (UK; 12.5%), and slightly stronger in the USA (14.8%; OECD, 2013). Note that these estimates from the PISA study are higher than the effect sizes in the meta-analyses; this might be evidence of SES having a stronger influence in more recent years, or the greater effect in PISA might be due to other study-specific factors such as more representative sampling of the population. Although greater equality was reported for Australia than the average of OECD countries, the difference between the lowest and highest SES quartile on mathematics, science and reading for Australian students was equivalent to 2.5 years of schooling (Thomson, De Bortoli, & Buckley, 2013). This indicates substantial room for improvement in the provision of equitable education opportunities among Australian school children.

One of the major limitations in most studies of the relationship between SES and academic achievement is the inability to distinguish the impact of SES on academic ability independent of innate influences on both. Genetic variation is a substantial contributor to individual differences in academic achievement; in developed countries genes typically explain more than half of the

variation in performance of these academic outcomes (de Zeeuw, de Geus, & Boomsma, 2015). For SES, recent studies in the UK using genome-wide complex trait analysis have estimated the heritability at approximately 20%, which is underestimated by the extent that relevant sources of genetic variation, such as copy number polymorphisms and markers of methylation are absent from the genotyping chips employed in the analyses (Marioni et al., 2014; Trzaskowski et al., 2014). Furthermore, genes mediate about half of the correlation between academic achievement and SES (Krapohl & Plomin, 2015). For the purpose of this paper the correlation between academic achievement and SES, whether due to genes or the environment, will be considered a main effect of the measure of SES in predicting academic achievement.

In addition to a main effect, SES has also been shown to moderate the influence of genes and the environment on academic achievement and cognitive ability, resulting in a gene-environment interaction (Fischbein, 1980; Harden, Turkheimer, & Loehlin, 2007; Kremen et al., 2005; Rhemtulla & Tucker-Drob, 2012; Rowe, Jacobson, & Van den Oord, 1999; Tucker-Drob, Rhemtulla, Harden, Turkheimer, & Fask, 2011; Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). A gene-environment interaction occurs when the influence of genes is different for people of different environments, or vice versa (Plomin, DeFries, & Loehlin, 1977). Thus far, few studies have explored whether family SES moderates the influence of genes and the environment on performance in literacy or numeracy, and these studies vastly differ in age of participants and measures of achievement (Harden et al., 2007; Kremen et al., 2005; Rhemtulla & Tucker-Drob, 2012; van den Oord & Rowe, 1997). Rhemtulla and Tucker-Drob (2012) showed heritability of early mathematical skills to increased in 4-year-olds with increasing family SES, and Tucker-Drob and Harden (2012b) demonstrated that this interaction was mediated by genetic variation in motivation to learn. At the other end of the school years, Harden et al. (2007) showed heritability of a latent construct of academic aptitude in 17 year-olds increased with increasing family income, and Tucker-Drob and Harden (2012a) demonstrated that this interaction was mediated by intellectual interest. This actualizing of genetic potential is consistent with that predicted by proximal processes of the bioecological model of development detailed by Bronfenbrenner and Ceci (1994). The bioecological model predicts heritability to be greater in more advantageous environments, and constrained in impoverished environments. While the environmental-disadvantage hypothesis also predicts constrained heritability in impoverished environments (Scarr-Salapatek, 1971), the bioecological model extended this prediction across the environmental range. The families in Harden et al.'s (2007) study reported higher SES than the population average, thus demonstrating an interaction between genes and the environment across relatively advantageous levels of the measured environment. Although there are few studies with

literacy or numeracy as outcomes, several others using measures of cognitive ability have also found increasing heritability with increasing family SES (Fischbein, 1980; Rowe et al., 1999; Tucker-Drob et al., 2011; Turkheimer et al., 2003). More recently, this effect has been termed the Scarr-Rowe hypothesis (Turkheimer, Harden, D'Onofrio, & Gottesman, 2009).

In most studies on the moderation of the heritability of general cognitive ability by SES an environment-environment interaction is also reported, with the influence of the shared environment decreasing with increasing SES (Rowe et al., 1999; Tucker-Drob et al., 2011; Turkheimer et al., 2003). In the most powerful gene-environment SES moderation study to-date, Hanscombe et al. (2012) assessed cognitive ability in a longitudinal study in the UK at eight different ages, ranging from 2–14 years old. Surprisingly, the most consistent result was decreasing shared environment and no significant change in heritability with increasing SES. This suggests a differential effect of the shared environment depending on SES, and not specific fostering of genetic potential in more advantaged environments. Similarly, among studies of literacy measures, Kremen et al. (2005) found the heritability of word reading in middle-aged adults did not change with increasing childhood family SES. Instead, an environment-environment interaction best explained their data, and again the influence of the environment decreased with increasing childhood family SES. This also suggests family SES can have an enduring environment influence on reading throughout life.

However, some studies have found no significant gene-environment interaction or environment-environment interactions. For early reading skills in 4 year-olds, Rhemtulla and Tucker-Drob (2012) found no interaction, and in children approximately 9.5 years old, van den Oord and Rowe (1997) examined a range of home SES factors and found little evidence for moderation of the heritability of reading or mathematics. It is possible this latter study was underpowered; they used a kinship algorithm to determine relatedness, categorizing children into full or half siblings or cousins, and tested gene-environment interaction using multilevel modelling. Using a similar method, Rowe et al. (1999) showed that without twins or unrelated siblings there was considerably less power to detect gene-environment interactions. In adult men (average age 19.6 ±1.5), Grant et al. (2010) found home SES did not moderate heritability of general cognitive ability. Interestingly, this was a much larger sample of which Kremen et al.'s 2005 study was a subset. These two studies used the same measure for SES, and while Grant et al. found no moderation on general cognitive ability at age 19, Kremen et al. did find an environment-environment interaction between SES and reading ability assessed approximately 30 years later. The different findings from these two studies might result from tests administered at different ages, or perhaps reflect different interactions with SES for cognitive ability and reading.

That heritability of performance on different measures might interact variously with SES is also apparent in a study of 4-year-olds, where there was no moderation of mental ability or early reading skills but there was for early mathematical skills (Rhemtulla & Tucker-Drob, 2012; Tucker-Drob et al., 2011). That age might also be a reason for differing results is apparent in van der Sluis et al.'s (2008) study in the Netherlands, where an environment-environment interaction on the heritability of IQ was found for men in an older cohort (average age 49.39), but not for men in a younger cohort (average age 26.56) or for women in either cohort.

In the only study to-date using school SES as a moderator, Hart et al. (2013) found heritability was higher in reading performance for twins attending lower SES schools, which is in the opposite direction to the effect more frequently reported. But consistent with the broader findings, they noted that the shared environment and total variance were greater in lower SES schools. The authors suggest schools with more resources might be more similar on various characteristics that facilitate reading proficiency in all students. Although it is important to not over-interpret this singular finding, in a study from the same project (the Florida Twin Project on Reading) the heritability of oral reading fluency in slightly younger twins was found to significantly increase with better quality teachers (Taylor, Roehrig, Hensler, Connor, & Schatschneider, 2010). It is not easy to explain the opposite direction of the interactions from these two studies that assess aspects of school environments as moderators of heritability, and replication in future studies might help to clarify these contradictory findings.

Taken together, research into the moderating effect of SES on the heritability of literacy, numeracy or general cognitive ability indicates a trend of increasing heritability with increasing family SES in school-aged children, although there are instances where this effect was not found. For cognitive ability, the shared environment has a decreasing influence with increasing SES. Gene-environment interactions with SES might not be as influential in adults, and the degree of influence might differ somewhat across measures of academic achievement and cognitive ability, and across countries. Furthermore, the type of interaction may differ between family and school measures of SES.

In order to examine if SES moderates genetic and environmental influence on academic achievement, we will employ the classic twin design. This capitalises on the genetic relatedness of monozygotic (MZ) and dizygotic (DZ) twins. Monozygotic twins share all of their segregating genes, while DZ twins share, on average, half of their segregating genes. Members of a twin pair will share some environments and not others. For twins who are children or adolescents and brought up in the same home then family factors (including SES) would contribute to shared environmental influences. Unique environmental influences would include environmental events

that are not shared, perhaps different classrooms, friends, or medical histories. From these known relationships, the covariation between MZ twins in a phenotype is compared to the covariation between DZ twins and can be used to estimate the influence of genes, the shared environment, and the unique environment. Specifically, twice the extent that MZ twin correlations are greater than that of DZ twins indicates additive genetic effects (A), twice the extent that DZ twin intraclass correlations are greater than half that of MZ twins indicates shared environmental effects (C), and the extent MZ twin intraclass correlations is less than unity indicates unique environment effects (E; for more detailed discussion see Plomin, DeFries, Knopik, & Neiderhiser, 2013). One assumption of the classic twin study is that genes and the environment independently influence the phenotype of interest. When some of the variation in the phenotype is due to gene-environment correlation, such as the genetic influence on the covariation between academic achievement and SES (Krapohl & Plomin, 2015), then estimates of A, C, and E will be biased (Purcell, 2002). In this paper we will control for the extent of any gene-environment correlation by removing the main effect of SES on academic achievement, and then explore if SES moderates the influence of genes and the environment on the remaining variance.

Specifically, we will assess if either family or school SES moderates the heritability of literacy and numeracy in school-aged children in Australia. Each year the National Assessment Program in Literacy and Numeracy (NAPLAN) tests are administered to students in Grades 3, 5, 7, and 9 in five academic domains: reading, spelling, grammar and punctuation, writing, and numeracy. In a recent study, we found genetic variation to substantially contribute to variation in performance on the NAPLAN, while the influence of the shared environment was low to modest (Grasby, Coventry, Byrne, Olson, & Medland, 2015). However, if a gene-SES interaction is present then estimates from the standard twin model will be biased; specifically, estimates of genetic effects will be inflated (Purcell, 2002). To further tease apart the possibility that the moderators of family and school SES have a differential impact on heritability, the current study will assess both of these SES indicators and in the same sample with the same achievement outcomes. We will assess the five domains of reading, spelling, grammar and punctuation, writing, and numeracy at several different grades, namely Grade 3, 5, 7, and 9. This will extend current research on these important school outcomes across the middle years of schooling.

Method

Participants

Twins and triplets born from 1993–2006 were recruited through the voluntary Australian Twin Registry. For the 34 sets of triplets, a random pair from each set was selected for the

analyses, and from hereon all multiple births are referred to as *twins*. Twins were invited to participate if they had sat (or would sit by 2014) a NAPLAN test. Of the 6853 families contacted, 2272 (33%) consented to participate. Of those, state departments provided NAPLAN results for 1949 families and of these zygosity information was reported from 1940 families. Most of the twins attended the same school: 99% in Grade 3, 98% in Grade 5, 95% in Grade 7, and 92% in Grade 9. Dizygotic twins were more likely to attend different schools, so to minimise potential confounds and adhere as closely as possible to the equal environments assumption only twins who were in the same school and year were included. This resulted in 1889 twin pairs who sat NAPLAN tests in the same year and school.

Zygosity was determined with a short questionnaire (Lykken, Bouchard, McGue, & Tellegen, 1990), which classified a sub-sample of twins in this study with 95% accuracy when compared to parent report of DNA-assigned zygosity. Our final sample of twin pairs included 860 monozygotic (MZ; 452 female, 408 male) and 1029 dizygotic (DZ; 296 female, 280 male, 453 opposite-sex. Table 1 details the number of pairs in each grade for each domain. There was substantial longitudinal overlap: 66% of participants in Grade 3, 99% in Grade 5, 98% in Grade 7, and 82% in Grade 9 provided results from more than one grade. The average age in Grade 3 at the time of testing was 8.6 years.

Materials

NAPLAN

The NAPLAN is a nationwide, standardised assessment introduced in 2008 in Australia. Each year, students in Grades 3, 5, 7, and 9 sit tests in reading, writing, language conventions, and numeracy. The test content is based on the "Statements of Learning for English" and the "Statements of Learning for Mathematics," which inform state and territory curricula. For each achievement domain, raw scores are transformed into a score on a common scale from 0–1000. This scaled score spans all years of the test and was designed to measure growth within cohorts and to compare across cohorts. Technical information and test administration details were obtained from the Australian Curriculum Assessment and Reporting Authority (ACARA; ACARA, 2015b; 2015c; R. Randall, personal communication, July 10, 2013). Example test papers and writing prompts are available at www.nap.edu.au.

Reading

The reading comprehension test was composed of 7–8 passages. The passages were extracts or adaptations from books, newspaper articles, posters or poems. Passage length varies from brief single paragraphs of about 100 words, to several paragraphs of about 450 words in

total. There were 5–8 items relating to any given passage. Most items are multiple-choice format, with a few short answer questions in each test. For Grades 3 and 5 there were 35–38 items to be completed in 45–50 minutes, and for Grades 7 and 9 there were 45–50 items to be completed in 65 minutes. A Cronbach's alpha of .85 (or above) for each test in each year indicates a high internal reliability.

Spelling

The spelling test presents misspelt words in simple sentences and requires students to identify and correct the spelling errors. For Grades 3 and 5 there were 23–25 items, and for Grades 7 and 9 there were 25–30 items. The spelling test is administered in the same paper as the grammar and punctuation test, and students are given 40–45 minutes to complete both of these question sets. For the spelling test, a Cronbach's alpha of .90 (or above) for each test in each year indicates a high internal reliability.

Grammar and punctuation

The grammar questions ask students to choose the correct word(s) to complete a sentence. This form of question is used in early grades to identify correct tense, pronouns, conjunctions, and verb forms. In later grades relative pronouns, clauses, and comparative adjectives are also assessed. The punctuation questions ask students to insert or identify punctuation marks at the correct location in a sentence. For all grades there were 23–28 items. A Cronbach's alpha of .71 to .87 for each test in each year indicates an acceptable internal reliability (average .80).

Writing

The writing test is composed of a writing stimulus, which provides an idea or topic, and students are asked to write a response in a specified writing style (i.e. narrative, informative, or persuasive). For example, "It is cruel to keep animals in cages. What do you think? Do you agree or disagree? Perhaps you can think of ideas for both sides of this topic." The same prompt and style is used for all grade levels in a given year. Students have 40 minutes of writing time. Marks are awarded on 10 criteria: audience, text structure, ideas, vocabulary, cohesion, paragraphing, sentence structure, punctuation, spelling, and the final criterion depended on the writing style specified. For persuasive writing the criterion was persuasive devices (2011–2014), and for narrative writing the criterion was character and setting (2008–2010). From 2008–2010 the maximum score was 47, from 2011–2014 the maximum score was 48. A Cronbach's alpha, calculated using pooled data from all grades, of .93 (or above) for each test year indicates a high internal reliability; unfortunately we were unable to obtain inter-rater reliability information.

Chapter 5

Numeracy

The numeracy test assesses five aspects of mathematics. *Working mathematically* includes problem solving, reasoning and interpretation. *Number* includes counting and computation. *Algebra, function and pattern* includes working with functions and relationships, graphs, equations, and rules. *Measurement, chance and data* includes working with units, likelihood and inference. *Space* includes shape and location. Most items are multiple-choice format, with a few short answer questions in each test. For Grade 7 and Grade 9, students sit a calculator-allowed and a non-calculator numeracy test. For Grades 3 and 5 there were 35–40 items to be completed in 45–50 minutes. For Grades 7 and 9 there were 62–64 items from the combined calculator and non-calculator papers, with each paper to be completed in 40 minutes. A Cronbach's alpha of .84 (or above) for each test in each year indicates a high internal reliability.

Family socioeconomic status

Level of parent education achieved was used as a measure of family SES. Parents were asked to select a level of education from: (a) some high school but did not finish; (b) school certificate[1]; (c) higher school certificate[2]; (d) TAFE or trade (including certificate/diploma)[3]; (e) 3-year university degree; (f) 4-year university degree; (g) some postgraduate study; (h) master's degree, and (i) doctoral degree. Note, within the Australian education system at the time that the parents were educated both school certificate in Grade 10 and higher school certificate in Grade 12 were legitimate exit points to graduate from secondary schooling. These responses were scored from 1 to 9. Mother and father education correlated .44, and an average parent education level was obtained and used in analyses.

School socioeconomic status

The Index of Community Socio-Educational Advantage (ICSEA) was used as a measure of the SES level of each school. ICSEA values are reported each year for each school by ACARA, and predict average school performance on the NAPLAN. The value incorporates family and community variables, including parent occupation and education, school location (metropolitan, regional, or remote), proportion of Indigenous students, and proportion of students with a disadvantaged language background other than English (LBOTE). Not all LBOTE students are disadvantaged; on average they outperform non-LBOTE students, but there are some particularly disadvantaged groups within the broader LBOTE group and they were identified by combining

---

[1] Exit school at Grade 10
[2] Exit school at Grade 12
[3] TAFE stands for Technical And Further Education institutes in Australia, which typically provide vocational training at the certificate and diploma level.

LBOTE with parents who have an education level of Grade 9 or below. Initially, census data was for parent measures, but direct family data has been used from 2010 (71% of schools in 2010 to 87% in 2014). The weight of each variable that contributes to the ICSEA value is calculated using stepwise regression (for more details see ACARA, 2015a; Barnes, 2010). The Australian median is 1000 with a standard deviation of 100. A higher score indicates a school with higher educational advantage.

Procedure

After receiving parental consent, the state and territory departments of education provided NAPLAN test results. The NAPLAN tests are administered in the morning over three consecutive days each year in the second full week of May (approximately 3.5 months into the school year). On the first day the language conventions test (comprising of the spelling and grammar and punctuation domains) is administered and, after a minimum 20 min break, is followed by the writing test. On the second day the reading test is administered. On the third day the numeracy tests are administered; for Grades 7 and 9 the first test permits use of a calculator and the second test does not. Support within specific constraints can be provided for students with disability, such as scribing or reading questions in the numeracy test. Across the nation 96% of students participate in the tests.

Analyses

Sex, age, age-squared, age-by-sex, and cohort effects have been found to influence mean performance, and sex has been shown to moderate the heritability in some of the domains and grades (details in Grasby et al., 2015). Thus, sex, age, age-squared, age-by-sex, and cohort effects were regressed out of the scaled scores. The impact of sex on the covariances was tested by adding it as a second moderator within the gene-environment interaction models for each test. Z-scored data were used in the gene-environment interaction models.

A continuous univariate gene-environment interaction model was used to estimate both school and family moderation of SES on genetic and environment variation in NAPLAN performance (Purcell, 2002). The path diagram of the model is depicted in Figure 5.1. Unmodelled gene-environment correlation can resemble gene-environment interaction, and there is evidence that covariation between family SES and academic achievement is substantially due to shared genes (Krapohl et al., 2015). In our model we are unable to test for gene-environment correlation, as twins share each of the moderators used in this study. Thus we controlled for gene-environment correlation by modelling the effect of the moderator on the means. Modelling the moderator on the means removes from the total variance the covariation between the moderator

and the outcome; as such the gene-environment interaction model estimates the influence of A, C, and E as a function of the moderator on the variance in the outcome that is independent of the moderator (Purcell 2002). Although we cannot estimate any gene-environment correlation in these data, by comparing unstandardised A, C, E variances before and after covariance with the moderator is removed we estimate the extent of genetic and shared environmental variance contributing to the SES and academic achievement correlation. This will also estimate how much of the total variation in academic achievement is influenced by genes and the environment independent of SES.
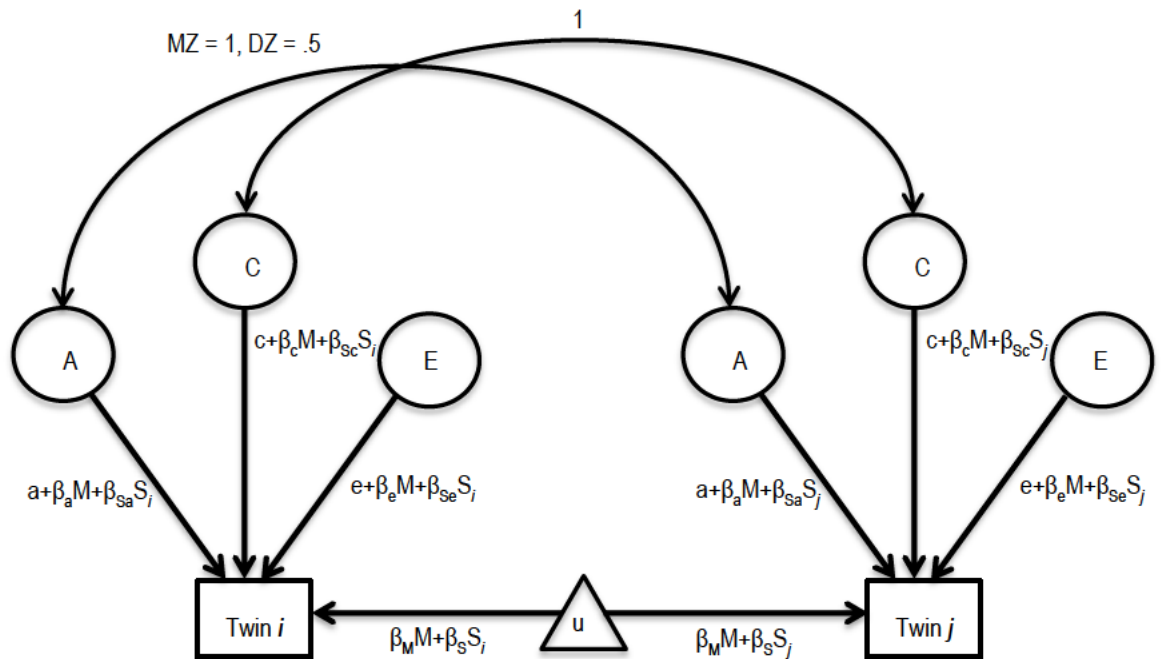
Figure 5.1 Path diagram of the moderation model. The value of the moderator (M; either family or school SES) is the same for each twin within a family, while sex (S) is allowed to differ within a family. The influence of the moderator on the mean (u), genetic variance (A), shared environmental variance (C), and unique environmental variance (E) are estimated respectively by the parameters β_M, β_a, β_c, and β_e, while the influence of sex is estimated by the parameters β_S, β_{Sa}, β_{Sc}, and β_{Se}. The genetic correlation for monozygotic (MZ) twins is fixed at 1 and for dizygotic (DZ) twins is fixed at .5, while the shared environment correlation is fixed at 1 for all twins.

Parental education contributes to the ICSEA calculation. To assess the influence of ICSEA after controlling for family SES, we removed the covariance of the outcome with parent education by modelling parent education onto the means. This was done to assess if school SES moderated the influence of genes and the environment after the main effects of both family and school SES were removed.

Sex has been found to the moderate the influence of genes and the environment in these NAPLAN tests (Grasby et al., 2015), so sex was modelled as a second moderator and the interaction of sex and each SES moderator was initially included. First, the interaction terms were dropped and significance was tested using the likelihood ratio test (LRT), which compares the

change in log likelihood to a χ^2 distribution with degrees of freedom equal to the difference in estimated parameters (Neale & Maes, 2004). Second, the interaction was dropped from the means. Third, sex as a moderator on A, C, and E was dropped. Finally, the SES moderator was dropped. When the moderator interacts in opposite directions with variance components, the LRT has reduced power to detect significant moderation when dropping all moderating parameters in a single step (Purcell, 2002). Therefore, each moderating parameter was dropped individually and tested. The best-fitting model was determined as the model with the lowest Akaike's Information Criterion (AIC; Akaike, 1987) that included significant moderating parameters from the individual tests. Models were estimated using raw data and full information maximum likelihood estimation in OpenMx, which uses all available data (Boker et al., 2011).

Results

Means, standard deviations and number of twin pairs for each moderator and domain by grade are reported in Table 5.1. NAPLAN scores were residualised for age, age-squared, age-by-sex, and cohort effects. The Australian ICSEA median is 1000 and standard deviation is 100. We have reported the ICSEA mean for each grade instead of the median; the two values differed by only up to 5 points in our data, as such the means in Table 5.1 directly relate to zero on the x-axis in Figures 5.2 and 5.3. On average, our sample is about 0.5 of a standard deviation above the nation in school educational advantage, suggesting a more advantaged sample than the population for school SES. Our sample also reported a higher level of education than that reported by the Australian Bureau of Statistics for people aged 15–74; however, people over 55 years of age are less likely than younger cohorts to have post-school qualifications (Australian Bureau of Statistics, 2014). In our study, the educational level of parents with post-school qualifications is in-fact proportional to those of the population who are of 25–54 years old (the equivalent age the parents). On the whole, our sample represents a higher than average school SES but is similar to the Australian population in family SES.

Phenotypic correlation between family SES and school SES for each grade was .40–.41. The correlations between family SES and the NAPLAN test performance were modest, .25 averaged over all domains and grades (see Table 5.1). Partial correlations between school SES and NAPLAN test performance, controlling for family SES, averaged .17 over all domains and grades. These indicate a small, positive relationship between school advantage and performance that is independent of family SES.

Chapter 5

Table 5.1 Descriptive statistics and phenotypic correlations between
NAPLAN domains and SES measures

| Variable | Descriptive statistics | | | Correlations | |
	M	*SD*	*n*	Family SES	School SES[a]
Grade 3					
Reading	442	85	1160	.31	.15
Spelling	425	74	1162	.23	.12
GP	450	88	1162	.30	.17
Writing	437	58	1163	.21	.13
Numeracy	424	70	1159	.28	.17
Family SES	4.7	1.6	1143	–	–
School SES	1050	78	1155	.41	–
Grade 5					
Reading	526	76	1076	.28	.16
Spelling	507	67	1076	.19	.14
GP	536	81	1077	.25	.15
Writing	520	61	1076	.22	.18
Numeracy	513	70	1076	.27	.20
Family SES	4.6	1.6	1055	–	–
School SES	1049	78	1067	.40	–
Grade 7					
Reading	578	65	933	.31	.17
Spelling	564	62	933	.19	.17
GP	567	68	935	.27	.15
Writing	570	68	934	.28	.17
Numeracy	574	71	932	.28	.20
Family SES	4.6	1.6	919	–	–
School SES	1042	73	928	.41	–
Grade 9					
Reading	631	63	741	.28	.15
Spelling	609	64	740	.14	.13
GP	614	71	739	.23	.17
Writing	607	76	738	.22	.24
Numeracy	630	69	736	.27	.21
Family SES	4.5	1.7	730	–	–
School SES	1045	72	735	.40	–

Note. *n* = number of twin pairs; SES = socioeconomic status;

GP = Grammar and punctuation

[a] Partial correlations controlling for family SES

The percentage of variance in NAPLAN performance due to family SES ranged from a mere 2% for spelling in Grade 9 to 10% for reading in Grade 3 and 5 (see Table 5.2). Family SES accounted for 7% of the variance in NAPLAN test performance when averaged across all domains and grades. After removing the effect of family SES, school SES accounted for an extra 1–4% of the variation in NAPLAN test performance. Table 5.2 shows the percentage of total variance in each NAPLAN domain explained by genes and the environment before and after the covariation with family SES and school SES are separately partitioned. Lower estimates of genetic variance after partitioning indicated some of the correlation between SES and academic achievement is due to genes; this was particularly the case for reading in Grade 3, spelling and writing in Grades 7 and 9, and numeracy in Grade 7. Comparing the change in estimates of the

shared environment before and after partitioning SES influences indicated shared environment influences that are independent of SES were only relevant to a few domains. The shared environment influences independent of SES contributed 5% or more to only grammar and punctuation in Grade 3, numeracy in Grade 5, reading in Grade 7, and reading, grammar and punctuation, and numeracy in Grade 9.

Table 5.2 Percentage of variance explained by A, C, E, before and after the main effects of SES are partitioned

Variable	Before			After				
	A	C	E	A	C	E	Family SES	School SES
Grade 3								
Reading	.73	.03	.24	.64	0	.25	.10	.02
Spelling	.77	.01	.22	.71	0	.23	.05	.02
GP	.52	.17	.30	.51	.06	.31	.09	.03
Writing	.39	.12	.49	.42	.03	.48	.04	.02
Numeracy	.65	.10	.25	.63	0	.25	.09	.02
Grade 5								
Reading	.53	.16	.30	.55	.04	.30	.10	.02
Spelling	.76	.04	.20	.74	0	.20	.04	.02
GP	.62	.09	.29	.62	0	.29	.07	.02
Writing	.47	.10	.43	.48	.01	.43	.06	.03
Numeracy	.56	.21	.24	.57	.08	.24	.09	.04
Grade 7								
Reading	.52	.20	.28	.53	.07	.28	.09	.02
Spelling	.77	0	.23	.71	0	.23	.04	.02
GP	.61	.07	.32	.59	0	.32	.07	.02
Writing	.52	0	.48	.41	0	.50	.06	.03
Numeracy	.75	.07	.18	.70	0	.18	.09	.03
Grade 9								
Reading	.56	.18	.27	.56	.08	.27	.08	.01
Spelling	.76	0	.24	.73	0	.24	.02	.01
GP	.51	.15	.34	.51	.07	.34	.06	.02
Writing	.40	.06	.54	.36	0	.54	.05	.04
Numeracy	.55	.23	.22	.54	.11	.22	.09	.03

Note. A = additive genetic variance; C = shared environment variance; E = unique environment variance; SES = socioeconomic status

The best-fitting models for each domain and grade by moderator are indicated in Table 5.3, while the details of the model-fitting statistics are available in Tables A15–A22. Where sex was a significant moderator, then it was the case for both family and school SES, so Table 5.3 has a single column indicating if sex was included as a moderator. For example, the asterisks in Table 5.3 indicate that for Grade 3 spelling when family SES was the moderator the best-fitting model allowed sex to moderate A, C, and E and family SES to moderate both C and E. For Grade 3 numeracy, when school SES was the moderator the best-fitting model allowed sex to moderate A, C, and E but no moderation by school SES.

Table 5.3 Summary of best-fitting moderated models for each domain and grade

Variable	Sex ace	Family SES a	c	e	ac	ae	ce	ace	null	School SES a	c	e	ac	ae	ce	ace	null
Grade 3																	
Reading							*										*
Spelling	*						*								*		
GP							*										*
Writing	*		*								*						
Numeracy	*							*									*
Grade 5																	
Reading	*							*									*
Spelling						*										*	
GP								*									*
Writing								*									*
Numeracy	*							*									*
Grade 7																	
Reading								*									*
Spelling	*							*									*
GP	I																*
Writing								*									*
Numeracy	*							*					*				
Grade 9																	
Reading							*									*	
Spelling	*							*									*
GP								*									*
Writing								*									*
Numeracy	*							*									*

Note. a = additive genetic; c = shared environment; e = unique environment; null = model without any environmental moderators; GP = Grammar and punctuation. I = significant interaction between sex and moderator.

*best-fitting model based on likelihood ratio tests of individual moderation terms and Akaike's Information Criterion for model fit

Overall, sex was a significant moderator on A, C, and E variance components for both family and school SES for numeracy in each grade; spelling in Grades 3, 7, and 9; writing in Grade 3, and reading in Grade 5. Parameter estimates of the moderated models for family SES are in Table 5.4 and school SES are in Table 5.5. Females were coded as 1 and males as 0. Generally, the influence of genes was greater in males and the environment greater in females, except for numeracy where both genes and the environment were greater in males. Only for family SES was there a significant interaction between the moderator and sex, which was for Grade 7 grammar and punctuation. Family SES interacted with genetic variation such that the influence of genes

increased for girls with increasing family SES while the influence of genes decreased for boys with increasing family SES.

Very similar results were found for family and school SES, and the most common result was no significant moderation of A, C, and E variance components. Moderation by family SES was present in the Grade 3 literacy tests, such that C and E decreased with increasing family SES in reading, spelling, and grammar and punctuation. Of these, only for spelling was the interaction also present with school SES as a moderator. For writing, C decreased with increasing family SES and this was particularly the case for girls. While for school SES the effect of sex indicated a reversal of direction of shared environmental effect on performance, when combined with school SES as a moderator this resulted in the influence of the shared environment decreasing in girls with increasing school SES, while the influence increased in boys with increasing school SES. In Grade 5, there was an unusual reversal from the Grade 3 spelling results such that C and E increased with increasing SES (both family and school) and A decreased. In Grade 7 numeracy, E decreased with increasing school SES. In Grade 9 reading, C and E decreased with increasing SES (both family and school level). Where interactions were significant, there was a general pattern of decreasing variance with increasing levels of either family or school SES.

Figure 5.2 displays the moderation of each domain for each grade with family SES as the moderator, and Figure 5.3 displays the results with school SES as the moderator. Where sex-effects were present, females and males are displayed separately. In each graph the y-axis is fixed to the same height, allowing visual comparisons of total variance across domains and grades.

Discussion

The most consistent finding in this study, across both family and school SES, was one of no moderation; 28 of 40 analyses showed no significant moderation. Where interactions were significant they typically involved a decrease in environmental variation with increasing SES. The only exception to this were for Grade 5 spelling, which showed decreasing genetic variance and increasing environmental variance with both family and school SES, and Grade 7 grammar and punctuation, which showed increasing genetic and environmental variance for girls with increasing family SES. As such, there was no evidence in our study of either environmental-disadvantage constricting genetic variation or of enriched environments potentiating genetic expression.
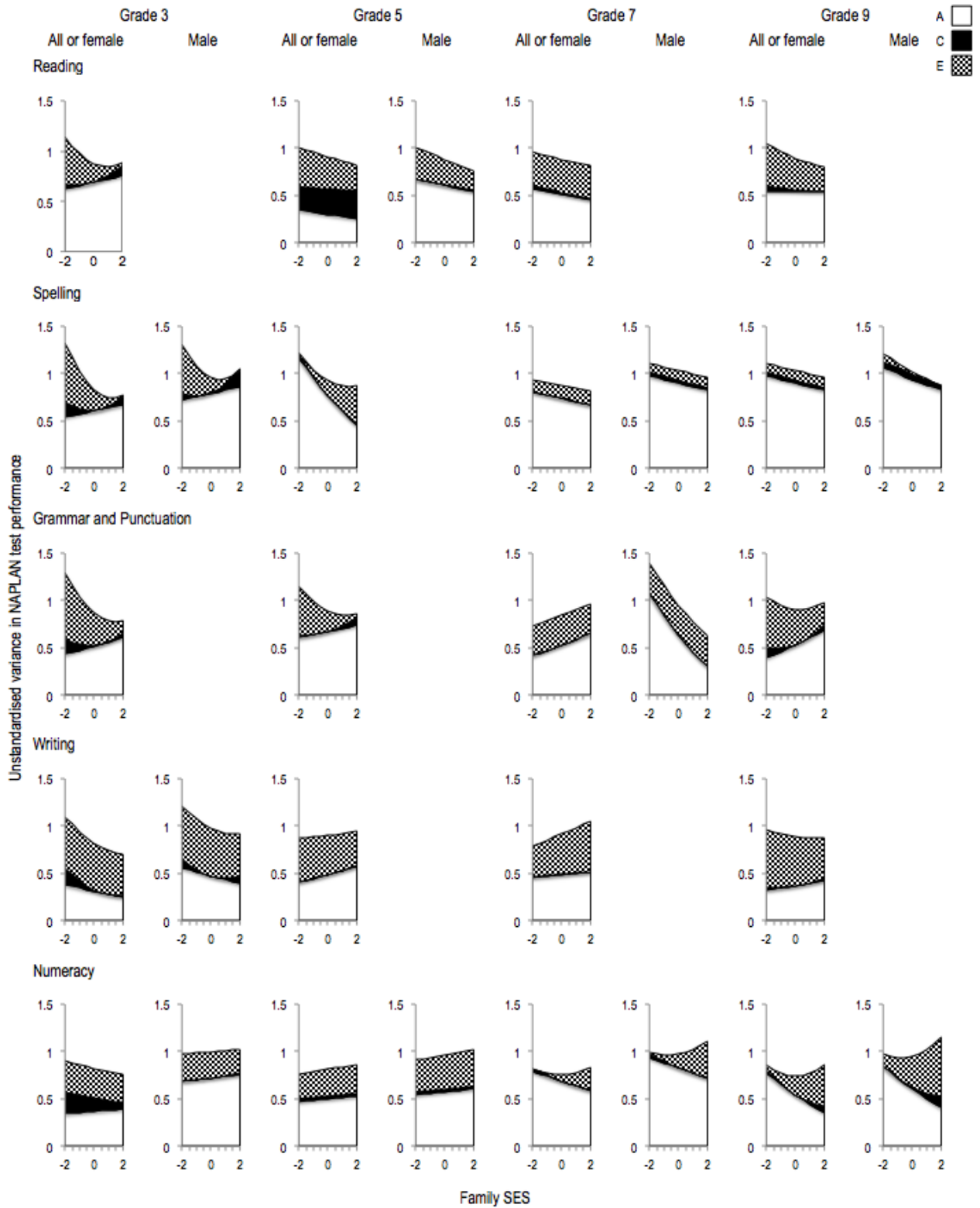
Figure 5.2 Figures of the family SES moderating additive genetic (A), shared environmental (C), and unique environmental (E) unstandardised variance components of the NAPLAN domains in each grade. For each grade, when sex could be dropped as a moderator, the left column of figures represents all twins when sex could be dropped as a moderator. When sex could not be dropped, the left column in a grade represents females and the right column represents males. Z-scored data was used, so comparison of total variation over grades is not possible but the relative contribution of A, C, and E across domain or time can be compared.
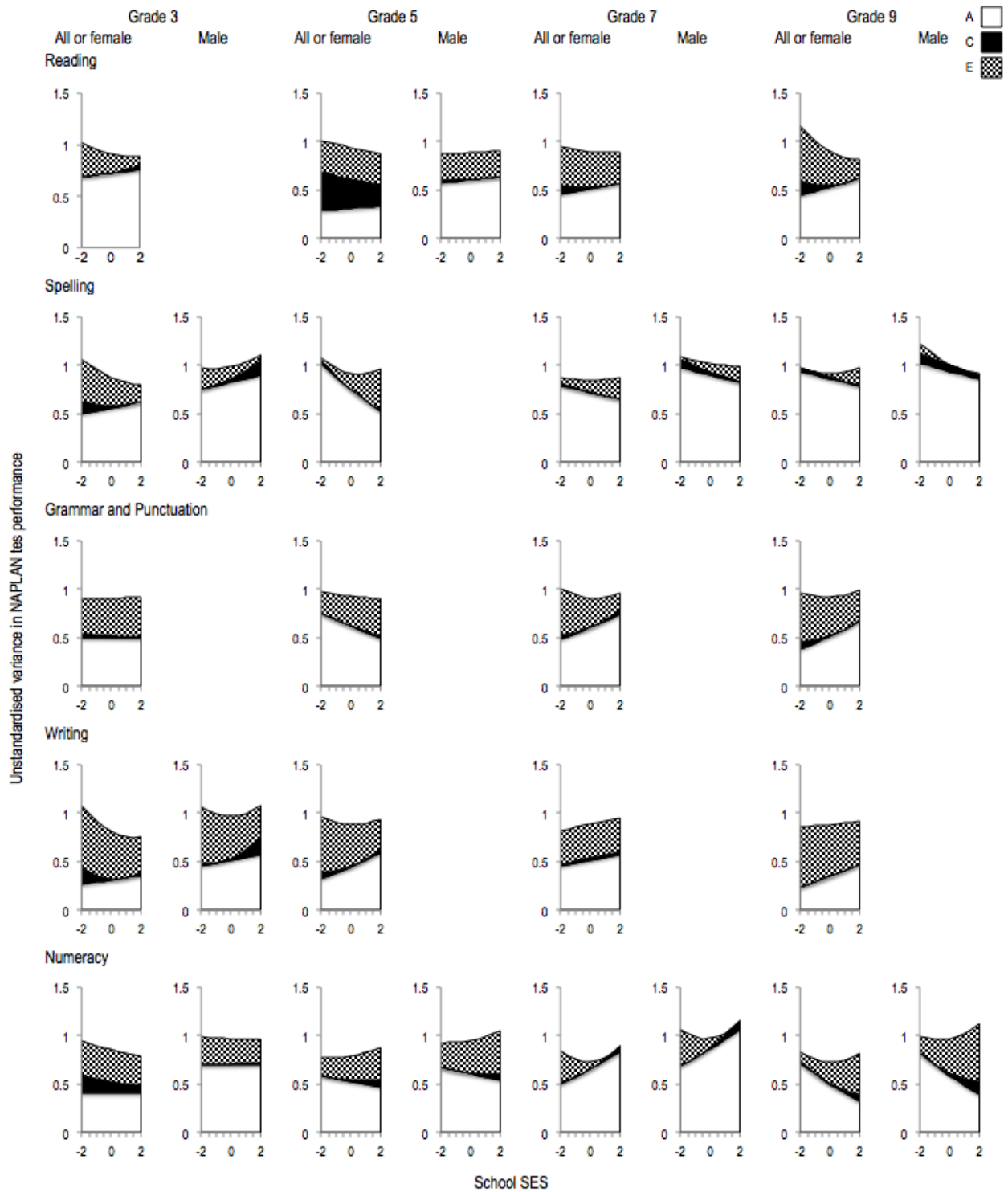
Figure 5.3 Figures of the school SES moderating additive genetic (A), shared environmental (C), and unique environmental (E) unstandardised variance components of the NAPLAN domains in each grade. For each grade, when sex could be dropped as a moderator, the left column of figures represents all twins when sex could be dropped as a moderator. When sex could not be dropped, the left column in a grade represents females and the right column represents males. Z-scored data was used, so comparison of total variation over grades is not possible but the relative contribution of A, C, and E across domain or time can be compared.

Table 5.4 Unstandardised parameter estimates of ACE model moderated by family SES and sex for each NAPLAN domain and grade

	Parameter estimates														
	Unmoderated			Moderated									Means		
				Family SES			Sex			Family SES*sex					
Variable	*a*	*c*	*e*	*a*	*c*	*e*	*a*	*c*	*e*	*a*	*c*	*e*	Sex	SES	SES*sex
Grade 3															
Reading	0.83	-0.05	0.43	0.02	-0.13	-0.13	-	-	-	-	-	-	-0.16	0.33	-
Spelling	0.78	0.05	0.47	0.02	-0.17	-0.16	0.11	-0.13	0.07	-	-	-	-0.20	0.22	-
GP	0.72	0.09	0.59	0.03	-0.16	-0.12	-	-	-	-	-	-	-0.25	0.30	-
Writing	0.56	0.13	0.70	-0.03	-0.14	-0.02	0.13	-0.11	0.01	-	-	-	-0.36	0.22	-
Numeracy	0.61	0.37	0.56	0.01	-0.05	-0.01	0.24	-0.41	-0.04	-	-	-	0.14	0.31	-
Grade 5															
Reading	0.55	0.52	0.58	-0.02	0.01	-0.03	0.23	-0.45	-0.06	-	-	-	-0.12	0.31	-
Spelling	0.88	-0.03	0.39	-0.10	0.07	0.12	-	-	-	-	-	-	-0.23	0.21	-
GP	0.82	-0.06	0.46	0.02	-0.11	-0.13	-	-	-	-	-	-	-0.22	0.27	-
Writing	0.70	0.01	0.64	0.03	-0.04	-0.02	-	-	-	-	-	-	-0.31	0.23	-
Numeracy	0.71	0.17	0.53	0.01	0.00	0.01	0.05	0.00	0.07	-	-	-	0.26	0.28	-
Grade 7															
Reading	0.72	0.12	0.59	-0.02	-0.03	0.00	-	-	-	-	-	-	-0.11	0.31	-
Spelling	0.85	-0.05	0.38	-0.02	0.01	0.01	0.10	-0.14	-0.07	-	-	-	-0.20	0.15	0.12
GP	0.74	0.00	0.56	0.04	0.01	0.01	0.07	0.00	0.00	-0.14	-0.07	-0.02	-0.12	0.24	0.04
Writing	0.70	-0.07	0.65	0.01	0.01	0.04	-	-	-	-	-	-	-0.38	0.25	-
Numeracy	0.82	-0.05	0.27	-0.03	0.07	0.12	0.08	-0.01	0.12	-	-	-	0.27	0.24	0.13
Grade 9															
Reading	0.74	0.11	0.58	0.00	-0.07	-0.04	-	-	-	-	-	-	-0.02	0.28	-
Spelling	0.93	-0.12	0.26	-0.03	0.01	0.07	0.04	-0.12	-0.41	-	-	-	-0.11	0.15	-
GP	0.73	0.05	0.61	0.05	-0.14	-0.06	-	-	-	-	-	-	-0.15	0.25	-
Writing	0.61	0.01	0.72	0.02	-0.08	-0.03	-	-	-	-	-	-	-0.32	0.22	-
Numeracy	0.74	0.03	0.44	-0.07	0.11	0.11	0.04	0.10	0.13	-	-	-	0.34	0.28	-

Note. SES = family socioeconomic status; a = additive genetic; c = shared environment; e = unique environment; GP = Grammar and punctuation.

Table 5.5 Unstandardised parameter estimates of ACE model moderated by school SES and sex for each NAPLAN domain and grade

| | Parameter estimates | | | | | | | | | Moderated | | | |
| | Unmoderated | | | School SES | | | Sex | | | | | Means | |
Variable	a	c	e	a	c	e	a	c	e	Sex	School SES	School SES*Sex	Family SES
Grade 3													
Reading	0.85	-0.06	0.43	0.01	-0.07	-0.07	-	-	-	-0.17	0.22	-	0.06
Spelling	0.75	0.16	0.54	0.02	-0.10	-0.06	0.16	-0.35	-0.19	-0.22	0.13	0.12	0.03
GP	0.71	0.14	0.62	0.00	-0.03	0.01	-	-	-	-0.26	0.18	0.13	0.05
Writing	0.56	0.11	0.70	0.02	-0.15	-0.05	0.16	-0.24	-0.04	-0.38	0.19	-	0.04
Numeracy	0.64	0.35	0.57	0.00	-0.04	-0.01	0.20	-0.40	-0.06	0.12	0.17	0.12	0.06
Grade 5													
Reading	0.56	0.55	0.57	0.01	-0.04	0.00	0.22	-0.46	-0.05	-0.12	0.23	-	0.05
Spelling	0.87	-0.03	0.40	-0.07	0.08	0.12	-	-	-	-0.22	0.14	0.10	0.03
GP	0.79	0.03	0.55	-0.04	0.03	0.04	-	-	-	-0.22	0.22	-	0.04
Writing	0.67	0.04	0.66	0.05	-0.11	-0.05	-	-	-	-0.32	0.23	-	0.03
Numeracy	0.73	0.10	0.50	-0.02	0.08	0.04	0.05	-0.01	0.08	0.25	0.25	-	0.04
Grade 7													
Reading	0.72	0.13	0.60	0.02	-0.08	-0.02	-	-	-	-0.11	0.24	-	0.05
Spelling	0.85	-0.07	0.35	-0.02	0.05	0.06	0.10	-0.12	-0.07	-0.20	0.15	0.12	0.03
GP	0.78	-0.01	0.55	0.04	-0.11	-0.07	-	-	-	-0.12	0.23	-	0.04
Writing	0.72	-0.14	0.59	0.02	-0.01	0.00	-	-	-	-0.37	0.25	-	0.04
Numeracy	0.81	-0.05	0.27	0.05	-0.10	-0.15	0.12	-0.05	0.04	0.26	0.25	-	0.05
Grade 9													
Reading	0.73	0.15	0.59	0.03	-0.12	-0.08	-	-	-	-0.01	0.21	-	0.05
Spelling	0.93	-0.14	0.19	-0.02	0.03	0.12	0.04	-0.11	-0.28	-0.11	0.15	-	0.03
GP	0.72	0.08	0.63	0.05	-0.10	-0.04	-	-	-	-0.13	0.22	-	0.04
Writing	0.59	0.03	0.73	0.05	-0.01	-0.03	-	-	-	-0.30	0.27	-	0.02
Numeracy	0.71	0.06	0.47	-0.07	0.10	0.09	0.06	0.10	0.12	0.36	0.24	-	0.06

Note. SES = school socioeconomic status; a = additive genetic; c = shared environment; e = unique environment; GP = Grammar and punctuation.

Chapter 5

Of the NAPLAN domains and grades assessed, most instances of significant SES moderation were in the younger students. In Grade 3, family SES significantly moderated each of the literacy variables; although, the moderation of reading, spelling, and grammar and punctuation differed slightly from that of writing. For Grade 3 writing, the influence of the shared environment on writing was modest when family SES was lower, about 15% for girls and 7% for boys when neither parent had completed a high school certificate. This effect reduced to a negligible level when both parents reported some post-school qualifications. Although there was no significant interaction between sex and family SES on Grade 3 writing, the direction of moderation was different for girls and boys. The influence of school SES mirrored that of family SES for girls, decreasing from 16 to 1% for girls attending schools with an ICSEA value of 900 through to 1050. In contrast, for boys the effect of moderation by the shared environment was greater in more advantaged schools. The shared environment increased from 2% to 17% for boys attending schools with an ICSEA value of 1050 through to 1200.

In contrast to writing, for reading, spelling, and grammar and punctuation both shared and unique environmental influences decreased in influence as family SES increased. For reading and spelling the effect of moderation on the shared environment was negligible, essentially because the main effect of the shared environment was so small. For grammar and punctuation, the influence of the shared environment was modest, at about 13%, for children with parents who had not completed a higher school certificate. As with writing, the influence of the shared environment became negligible for children with both parents completing some post-school qualifications. There were more substantial differences in the influence of the unique environment, which accounted for about 50% of the variance in children with parents who did not finish high school, compared to about 22% of the variance in children with parents with postgraduate qualifications. For reading and spelling the effect on the unique environment was even more pronounced, influencing 40–50% of the variance in low-SES families and reducing to a negligible amount in the highest family SES. Variation that is due to measurement error is included in the influence of the unique environment. Thus, it is improbable that there is no measurement error, so the negligible estimates of the unique environment found in the tail of our family SES data is probably a consequence of the model requiring the interaction to be linear. Although it is not reasonable to conclude there is no unique environmental variation when SES is high, there is a reduced influence of the unique environment with increasing SES, at least in these data. Our findings might support the presence of genuine environmental influences that are uniquely experienced by twins that impact on NAPLAN reading, spelling, and grammar and punctuation outcomes in Grade 3, and that these factors are particularly prevalent in children from

families with lower SES. Alternatively, children from families with lower SES might be exposed to more environmental factors that might introduce measurement error into NAPLAN test performance, such as incidence of illness. Only with spelling was the effect also present with school SES, where the size of this SES moderation effect was less marked, but the sex-effect was more pronounced. For girls the influence of the unique environment was at least twice that of boys. When converted to standardised variance estimates, 40% of the variance in spelling for girls attending schools with an ICSEA of 930 was due to the unique environment and this reduced to 25% for girls attending schools with an ICSEA of 1170.

In contrast to Grade 3, by Grade 5 the only domain to continue to show a significant interaction with SES was spelling. Quite inexplicably the nature of the interaction was the reverse of that at Grade 3, such that the influence of the environment increased with increasing family and school SES. The size of the moderation was comparable for family and school SES. Oddly, Grade 5 spelling was the only grade that showed no sex-effect for spelling. Grade 5 spelling was also one of the two analyses to show a significant gene-environment interaction, with the genetic influence being much stronger in lower-SES families. The uniqueness of this result, the reversal of environmental effect in the space of two grades, and a substantial overlap in participants taking the Grade 3 and Grade 5 NAPLAN tests reduces confidence in the robustness of this finding.

Another result that is unusual in our data is from Grade 7 grammar and punctuation. It was the only other significant gene-environment interaction in these data; moreover, there was an interaction between family SES and sex, such that genetic effects increased in girls and decreased in boys. As with Grade 5 spelling, there are several reasons we suspect this significant interaction to be a chance effect. (1) These results deviate from a more general pattern in that there are no sex-effects in grammar and punctuation at any other grade, nor were any noted in a slightly larger sample that included twins in different schools from which this sample was drawn (Grasby et al., 2015). (2) The interaction did not replicate in any other grade or with school SES as a moderator in Grade 7. There is considerable overlap in participants across years, so some replication or trends within our data is expected. (3) Higher heritability in less advantaged environments is rarely reported in the broader literature, and interactions with sex almost never. (4) Given the number of domains and grades assessed, it is likely that we will have some significant results due to chance. In a study similar to ours, Hanscombe et al. (2012) also reported an anomalous finding of higher heritability at lower SES that did not replicate across their data when using measures of SES taken at different ages. We presume this is a sampling effect.

Unlike the Grade 5 spelling and Grade 7 grammar and punctuation, Grade 9 reading was significantly moderated in a manner consistent with a general pattern of decreasing variance and

with increasing SES. The shared and unique environment decreased with increasing family and school SES. The size of the effect of the unique environment was similar for both family and school SES, but did not reduce as dramatically at Grade 9 as at Grade 3. The unique environment accounted for 42% of the variation in children with parents who did not finish high school, and this reduced to about 30% when both parents had postgraduate qualifications. The effect of the shared environment was more marked for school SES than family SES. For family SES the effect reduced from 6% of the variation in reading for twins with parents who had not completed high school, to a negligible effect for twins with parents who both completed high school. In contrast, the influence of the shared environment on reading for children in schools with an ICSEA of 900 was 13%, which reduced to a negligible amount for schools with an ICSEA of 1050. Although most literacy domains showed no moderation, where significant moderation was noted the trend is of greater influence from shared and unique environments in lower SES families or schools, particularly when parent education did not progress beyond high school or school ICSEA values were below 1040.

Decreasing environmental variance with increasing SES was also noted in Grade 7 numeracy. Although there was a sex-effect, the standardised variance estimates were very similar. The influence of unique environment decreased from 35% for boys and 38% for girls to 10% for boys and girls attending schools with an ICSEA value of 900 through to 1042, and continues to decrease to a negligible amount for children attending schools with an ICSEA value of 1200. As with the earlier grades, the very small estimate for children attending high-SES schools is partly due to the linearity of the model. Although this effect is in the same direction as most of the interactions that were significant in the literacy variables, the rest of the numeracy interactions for Grades 7 and 9 trended in the opposite direction. However, if the effect is real, then it might reflect a genuine environmental effect that is particularly relevant at the school level during this year that transitions students from primary to high school. Compared to the literacy tests, the numeracy test changes to incorporate a non-calculator and a calculator test. It is feasible that new demands on students requiring access and use of resources could result in greater environmental variation in numeracy performance in lower SES schools, while higher SES schools are able to effectively provide comparable and consistent access to useful resources.

Unlike studies conducted in the USA, our main finding across all domains and grades was that genetic variation did not vary as function of either family SES or school SES in these Australian NAPLAN tests. There is no evidence of constrained genetic potential in children attending less advantaged schools or who have parents with less education. Given that our data are from families who volunteer to participate in research, we acknowledge that our data are

unlikely to capture extreme environmental disadvantage. However, as the bioecological model is relevant to changes in heritability across the normal range of environmental advantage, it was surprising to find no evidence of advantageous family and school environments potentiating genetic expression. Our findings are more similar to Hanscombe et al.'s (2012) study from the UK, although Hanscombe et al. found more consistent evidence of greater shared environmental influences with lower SES than we did. This decrease in the influence of the shared environment is also found in the broader research on cognitive ability in the USA (Rowe et al., 1999; Tucker-Drob et al., 2011; Turkheimer et al., 2003). Our findings are also consistent with a very recent meta-analysis, which suggests the Scarr-Rowe effect is evident in the USA but not other Western countries (Tucker-Drob & Bates, 2015).

The general finding of no moderation in our Australian data, as compared to significant results in both the USA and UK, might result from cross-country differences. While we cannot test with these data what these factors are, we propose how differences in education and health care might contribute to these different findings. The UK has an established centralised education system, with a national curriculum introduced in 1989 (Whetton, 2009). Meanwhile Australia and the USA only moved toward a national curriculum in the late 2000's (Donnelly & Wiltshire, 2014; Kornhaber, Griffith, & Tyler, 2014). Australia implemented their initial phase in 2013 (Donnelly & Wiltshire, 2014). Prior to this the state and territory curricula were informed by "statements of learning" that were more limited in scope than a national curriculum, but these statements contribute directly to the development of the NAPLAN tests used in the current study (Donnelly & Wiltshire, 2014). In the USA, 45 states had adopted the Common Core State Standards Initiative by 2014 (Kornhaber et al., 2014). These Common Core State Standards differed considerably from existing standards in many states (Porter, McMaken, Hwang, & Yang, 2011). Considering the recency of adopting a common standard in the USA, increasing heritability with increasing SES noted in the USA might reflect considerable differences in educational standards across the population. Importantly, there is evidence that the USA has a greater disparity in provision of quality teaching and educational resources based on SES than other developed countries (Akiba, LeTendre, & Scribner, 2007; Darling-Hammond, 2014; OECD, 2009). Therefore, compared to the UK and Australia, the USA has both more variation in educational standards and access to higher standards has been greater for children with higher SES, potentially culminating in greater expression of genetic variation for those with greater resources and a constriction of genetic expression for those with less. Thus, we speculate that a more equitable provision of quality education in the UK and Australian results in the relative importance of genetic variation in performance regardless of SES.

Compounding differences in educational opportunity, access to health care is markedly different in the USA compared to the UK and Australia. The USA is not a poor country, but it has one of the highest rates of child poverty among OECD countries (OECD, 2009). Furthermore, poorer families in the USA report significantly poorer access, wait times, quality, and ability to follow up with treatment due to costs than higher SES families, while this is not the case in Australia and the UK (Schoen & Doty, 2004). Greater environmental variance in poorer families in the USA might result from greater exposure to environmental impediments to performance, such as poorer quality educational resources, greater risks to health, and a reduced capacity to obtain appropriate health care. Although both the UK and Australia have equitable access to health care, greater shared environmental effects for poorer families were more consistently observed in the UK study. This might result from Australia having, on average, better child wellbeing on measures of housing, local environment, and educational wellbeing than the UK (OECD, 2009), or it might reflect our sample not sufficiently representing the lowest range of SES in Australia while the UK sample is more representative.

Consistent with broader research into the relationship between SES and academic achievement, both family and school SES were related to NAPLAN performance. Family SES accounted for 2–10% (depending on the assessment domain) of the total variation in NAPLAN performance across the SES range in our data. Over and above this family effect, school SES contributed 1–4% of the variation. These estimates are not solely a measure of the shared environment; some genetic variation contributed to this SES and achievement correlation. The marked drop in the influence of the shared environment when the covariation with SES was partitioned indicates that much of the shared environmental influence on NAPLAN performance is due to aspects of the shared environment that systematically vary with parent education and/or ICSEA, meaning that these aspects differ between lower and higher SES, but do not differ within an SES level. After controlling for genetic and environmental influences that covary with SES, the remaining influence of the shared environment averaged across all domains and grades was only 3% of the variance. That is, the variation in shared environmental factors within a level of SES has little influence on NAPLAN performance. The moderation analyses indicate that, for some domains, this influence was stronger with lower SES. For most domains and much of the SES range, the shared environment did not influence variation in NAPLAN performance beyond the main effects of SES.

The NAPLAN tests have been subject to extensive public criticism; one concern is that "teaching to the test" will result in a "narrowing of the curriculum" (Donnelly & Wiltshire, 2014). Our purpose in raising this is not to engage in a discussion on the validity of these concerns,

instead the current study gives us an opportunity to examine the extent of influence on performance that teaching to the test might have. According to a study by the Whitlam Institute (Dulfer, Polesel, & Rice, 2012) there is considerable variability in the extent of test preparation. Within our model there are two portions of variance that might be influenced by teaching to the test. The effect might be included in the variance explained by the main effect of school SES (2% on average); this would capture preparation that correlates with ICSEA values. Alternatively, test preparation would be captured in the shared environmental variance component of each domain and grade after the variance explained by parent education and school advantage is removed, which averaged 3% of the variance. These shared environmental estimates include all school factors and environmental factors beyond parental education that are shared by twins that influence performance in these tests. As such, the degree that test preparation is likely to have on NAPLAN performance is minimal. These estimates give qualitative support to the recommendation by the test developers that teaching to the test is a largely futile endeavour.

Limitations

There are several limitations with the current study. Our sample is similar to that of the wider Australian population on level of parental education attained; however, it is drawn from families who voluntarily register to participate in research and is unlikely to represent the entire distribution of families in the population. For example, it is unlikely that many of our participants come from extremely disadvantaged homes. The distribution of our school advantage values indicate that our sample attended schools that are generally more advantaged than the wider population; for example, two standard deviations below the mean in our sample is equivalent to only one standard deviation below the national ICSEA value. As such our findings do not have the range, or the power in the lower range, to detect any gene-environment interactions that might address more disadvantaged schools in Australia. Our upper SES distribution more closely aligns with the population, as such our results speak most closely to the effects in the normal to more advantaged range of environments in Australia.

Although parent education has frequently been employed as a measure of family SES, we acknowledge that socioeconomic status is a complex construct, and perhaps aspects of SES that are unique from parent education may moderate the heritability of literacy and numeracy, as noted by Harden et al. (2007).

A broader limitation in assessing the influence of SES on literacy and numeracy, either as a direct relationship or as a moderator of genetic and environmental influences, is that SES is a distal construct of proximal process that are the direct interface with variability in literacy and

numeracy. Future studies could seek to measure those processes that mediate the relationship between SES and literacy and numeracy. These processes might be genetic or environmental in origin. They might be motivational factors, as explored by Tucker-Drob and Harden (2012a, 2012b), or they might be access to specific resources, like health care.

Another limitation is our inability to estimate the presence, or extent, of assortative mating on performance in literacy and numeracy. The parents in this study correlated on their level of education, which has been shown to account for the shared environmental variance in twin level of education (Baker, Treloar, Reynolds, Heath, & Martin, 1996). While literacy and numeracy ability might contribute to level of education, without access to parent performance on the literacy and numeracy measures we are unable to assess the degree of assortative mating in these data. Rather we acknowledge that the shared environmental estimate may be overestimated.

Conclusions

The main finding of the current study was that the influence of genes and the environment on performance in literacy and numeracy tests in Australia are largely the same across different levels of SES. In particular, genetic effects are substantial and stable regardless of whether parents did not complete high school or have postdoctoral degrees and regardless of whether children attended school with an ICSEA value of 900 or 1200. Although family SES contributes 2–10% and school SES contributes a further 1–4%, generally more than half of variation in NAPLAN performance is due to inherited child characteristics independent of both family and school SES. The absence of SES moderating heritability in Australia is encouraging for a society that aims to provide quality education regardless of the school attended.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*(3), 317-332. doi:10.1007/BF02294359

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 36*(7), 369-387. doi:10.3102/0013189x07308739

Australian Bureau of Statistics. (2014). *Education and work, Australia, May 2014*. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/6227.0Main+Features1May 2014?OpenDocument

Australian Curriculum Assessment and Reporting Authority. (2015a). *ICSEA 2014: Technical Report*

Australian Curriculum Assessment and Reporting Authority. (2015b). *National Assessment Program—Literacy and Numeracy 2014: Technical report*

Australian Curriculum Assessment and Reporting Authority. (2015c). *National protocols for test administration*. Retrieved from http://www.nap.edu.au/naplan/school-support/national-protocols.html

Baker, L., Treloar, S., Reynolds, C., Heath, A., & Martin, N. (1996). Genetics of educational attainment in Australian twins: Sex differences and secular changes. *Behavior Genetics, 26*(2), 89-102. doi:10.1007/bf02359887

Barnes, G. (2010). *Report on the generation of the 2010 index of community socio-educational advantage (ICSEA)*

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306-317. doi:10.1007/s11336-010-9200-6

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nuture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*(4), 568-586. doi:10.1037/0033-295x.101.4.568

Crane, A. R. (1951). One of our yesterdays: Sir Henry Parkes and the passing of the 1880 Public Instruction Act of N.S.W. *The Australian Quarterly, 23*(3), 25-30. doi:10.2307/20633367

Darling-Hammond, L. (2014). What can PISA tell us about U.S. education policy? *New England Journal of Public Policy, 26*(1).

de Zeeuw, E. L., de Geus, E. J. C., & Boomsma, D. I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education, 4*(3), 69-76. doi:10.1016/j.tine.2015.06.001

Donnelly, K., & Wiltshire, K. (2014). Review of the Australian curriculum. Final report. Retrieved from https://docs.education.gov.au/documents/review-australian-curriculum-final-report

Dulfer, N., Polesel, J., & Rice, S. (2012). The experience of education: The impacts of high stakes testing on school students and their families-an educator's perspective. The Whitlam Institute, Sydney. Retrieved April 30, 2013.

Fischbein, S. (1980). IQ and social class. *Intelligence, 4*(1), 51-63. doi:10.1016/0160-2896(80)90006-9

Grant, M., Kremen, W., Jacobson, K., Franz, C., Xian, H., Eisen, S., . . . Lyons, M. (2010). Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood? *Behavior Genetics, 40*(4), 438-446. doi:10.1007/s10519-010-9351-3

Grasby, K. L., Coventry, W. L., Byrne, B., Olson, R. K., & Medland, S. E. (2015). Genetic and environmental influences on literacy and numeracy performance in Australian school children in Grades 3, 5, 7, and 9. *Manuscript submitted for publication*.

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PloS One, 7*(2), e30320. doi:10.1371/journal.pone.0030320

Harden, K. P., Turkheimer, E., & Loehlin, J. (2007). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavior Genetics, 37*(2), 273-283. doi:10.1007/s10519-006-9113-4

Hart, S. A., Soden, B., Johnson, W., Schatschneider, C., & Taylor, J. (2013). Expanding the environment: Gene × school-level SES interaction on reading comprehension. *Journal of Child Psychology and Psychiatry, 54*(10), 1047-1055. doi:10.1111/jcpp.12083

Kornhaber, M., Griffith, K., & Tyler, A. (2014). It's not education by zip code anymore—but what is it? Conceptions of equity under the Common Core. *Education Policy Analysis Archives, 22*(4). doi:10.14507/epaa.v22n4.2014

Kornhauser, A. W. (1918). The economic standing of parents and the intelligence of their children. *Journal of Educational Psychology, 9*(3), 159-164. doi:10.1037/h0072624

Krapohl, E., & Plomin, R. (2015). Genetic link between family socioeconomic status and children/'s educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry*. doi:10.1038/mp.2015.2

Kremen, W. S., Jacobson, K. C., Xian, H., Eisen, S. A., Waterman, B., Toomey, R., . . . Lyons, M. J. (2005). Heritability of word recognition in middle-aged men varies as a function of parental education. *Behavior Genetics, 35*(4), 417-433. doi:10.1007/s10519-004-3876-2

Lykken, D. T., Bouchard, T. J., McGue, M., & Tellegen, A. (1990). The Minnesota Twin Family Registry: Some initial findings. *Acta Geneticae Medicae et Gemellologiae, 39*(1), 35-70. doi:10.1017/S0001566000005572

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., . . . Deary, I. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence, 44*, 26-32. doi:10.1016/j.intell.2014.02.006

Meyer, J. W., Tyack, D., Nagel, J., & Gordon, A. (1979). public education as nation-building in America: Enrollments and bureaucratization in the American states, 1870-1930. *American Journal of Sociology, 85*(3), 591-613. doi:10.1086/227051

Neale, M. C., & Maes, H. H. (2004). *Methodology for the genetic studies of twins and families* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Organisation for Economic Co-operation and Development. (2009). Comparative child well-being across the OECD *Doing better for children*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2013). *PISA 2013 Results: Excellence through equity: Giving every student the chance to suceed (Volumne II)*. Retrieved from http://dx.doi.org/10.1787/9789264201132-en

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral genetics*. New York: Worth.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84*(2), 309-322. doi:10.1037/0033-2909.84.2.309

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher, 40*(3), 103-116. doi:10.3102/0013189x11405038

Proctor, H. (2007). Gender and merit: Coeducation and the construction of a meritocratic educational ladder in New South Wales, 1880–1912. *Paedagogica Historica, 43*(1), 119-134. doi:10.1080/00309230601080618

Purcell, S. (2002). Variance components models for gene x environment interaction in twin analysis. *Twin Research, 5*(6), 554-571. doi:10.1375/136905202762342026

Ramirez, F. O., & Boli, J. (1987). The political construction of mass schooling: european origins and worldwide institutionalization. *Sociology of Education, 60*(1), 2-17. doi:10.2307/2112615

Rhemtulla, M., & Tucker-Drob, E. M. (2012). Gene-by-socioeconomic status interaction on school readiness. *Behavior Genetics, 42*(4), 549-558. doi:10.1007/s10519-012-9527-0

Rowe, D. C., Jacobson, K. C., & Van den Oord, E. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70*(5), 1151-1162. Retrieved from <Go to ISI>://WOS:000083197900008

Scarr-Salapatek, S. (1971). Race, Social Class, and IQ. *Science, 174*(4016), 1285-1295. doi:10.1126/science.174.4016.1285

Schoen, C., & Doty, M. M. (2004). Inequities in access to medical care in five countries: findings from the 2001 Commonwealth Fund International Health Policy Survey. *Health Policy, 67*(3), 309-322. doi:10.1016/j.healthpol.2003.09.006

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453. doi:10.2307/3515987

Taylor, J., Roehrig, A. D., Hensler, B. S., Connor, C. M., & Schatschneider, C. (2010). Teacher quality moderates the genetic effects on early reading. *Science, 328*(5977), 512-514. doi:10.1126/science.1186149

Thomson, S., De Bortoli, L., & Buckley, S. (2013). *PISA in brief: Highlights from the full Australian report: PISA 2012: How Australia measures up*. Retrieved from http://research.acer.edu.au/ozpisa/14

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., . . . Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence, 42*, 83-88. doi:10.1016/j.intell.2013.11.002

Tucker-Drob, E. M., & Bates, T. C. (2015). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science*. doi:10.1177/0956797615612727

Tucker-Drob, E. M., & Harden, K. P. (2012a). Intellectual interest mediates gene × socioeconomic status interaction on adolescent academic achievement. *Child Development, 83*(2), 743-757. doi:10.1111/j.1467-8624.2011.01721.x

Tucker-Drob, E. M., & Harden, K. P. (2012b). Learning motivation mediates gene-by-socioeconomic status interaction on mathematics achievement in early childhood. *Learning and Individual Differences, 22*(1), 37-45. doi:10.1016/j.lindif.2011.11.015

Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E., & Fask, D. (2011). Emergence of a gene x socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychological Science, 22*(1), 125-133. doi:10.1177/0956797610392926

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*(6), 623-628. doi:10.1046/j.0956-7976.2003.psci_1475.x

Turkheimer, E., Harden, K. P., D'Onofrio, B., & Gottesman, I. I. (2009). The Scarr–Rowe interaction between measured socioeconomic status and the heritability of cognitive ability. In K. McCartney & R. A. Weinberg (Eds.), *Experience and development: A festschrift in honor of Sandra Wood Scarr* (pp. 81–97). New York, NY: Psychology Press.

van den Oord, E. J. C. G., & Rowe, D. C. (1997). An examination of genotype- environment interactions for academic achievement in an U.S. national longitudinal survey. *Intelligence, 25*(3), 205-228. doi:10.1016/S0160-2896(97)90043-X

van der Sluis, S., Willemsen, G., de Geus, E. C., Boomsma, D., & Posthuma, D. (2008). Gene-environment interaction in adults' IQ scores: Measures of past and present environment. *Behavior Genetics, 38*(4), 348-360. doi:10.1007/s10519-008-9212-5

Whetton, C. (2009). A brief history of a testing time: national curriculum assessment in England 1989–2008. *Educational Research, 51*(2), 137-159. doi:10.1080/00131880902891222

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461-481. doi:10.1037/0033-2909.91.3.461

Higher Degree Research Thesis by Publication
University of New England

STATEMENT OF AUTHORS' CONTRIBUTION

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research Master/PhD candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated in the *Statement of Originality*.

	Author's Name (please print clearly)	Contribution
Candidate	Katrina Grasby	Responsible for study design; data cleaning and analysis; writing of paper
Other Authors	William Coventry	Provided editorial comments; author of grant, which provided data for paper

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry

28/1/16
Date

26/1/16
Date

Higher Degree Research Thesis by Publication
University of New England

STATEMENT OF ORIGINALITY

(To appear at the end of each thesis chapter submitted as an article/paper)

We, the Research Master/PhD candidate and the candidate's Principal Supervisor, certify that the following text, figures and diagrams are the candidate's original work.

Type of work	Page number/s
All text, tables, and figures	Chapter 5, pp 115–146

Name of Candidate: Katrina Grasby

Name/title of Principal Supervisor: Dr William Coventry
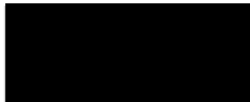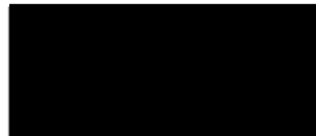
28/1/16
Date

28 · 1 · 16
Date

Chapter 6

General Discussion

Chapter 6

The papers in this thesis are the first to address the influence of genes and the environment on individual differences in literacy and numeracy in Australian children through the middle grades of school. Literacy and numeracy were assessed with the NAPLAN tests of reading, spelling, grammar and punctuation, writing, and numeracy. These tests have attracted contentious debate for a variety of reasons; where possible I seek to apply these behaviour genetic findings to contribute to a better understanding of the tests and the Australian context around them. Specifically, I addressed whether performance on large-scale reading tests converged with performance on individually administered tests of reading skills. I tested the structure of genetic and environmental influences on covariation among the five domains, and whether these influences were stable as children progressed through school. I assessed the genetic and environmental influence on variation in stability and growth in each of the five domains. Finally, I examined the extent of both family and school SES as a main effect on performance and as a moderator of genetic and environmental influences on performance.

The first paper focused on the validity of large-scale reading tests. This study differs somewhat from the other papers in this thesis, in that the sample of participants were drawn from the ILTS, a project that administered well-accepted, individually administered tests of reading skills to twins annually from their preschool year through to Grade 2. In Grade 3, participants provided their school literacy and numeracy results from either the NAPLAN or its antecedent the BST. Dyadic analyses found tests of word reading, vocabulary, and reading comprehension were preferentially related to performance on large-scale reading tests than large-scale numeracy tests. Higher genetic and unique environmental correlations between the individually administered tests and large-scale reading tests compared to large-scale numeracy tests further supported the validity. Large-scale, predominantly multiple-choice reading tests might lack finesse in assessing reading skills and the capacity to provide diagnostic information regarding the causes of poor reading ability, but they broadly capture reading ability.

Although validity was not the main aim of the second empirical study, evidence of genetic independence among the NAPLAN domains also suggests a degree of specificity to the assessment domains. Studies conducted in the UK and the USA with a variety of reading and mathematical measures have found both genetic overlap and independence between these domains (Hart, Petrill, & Thompson, 2010; Hart, Petrill, Thompson, & Plomin, 2009; Kovas, Harlaar, Petrill, & Plomin, 2005). Therefore, if these NAPLAN data had revealed no genetic independence between reading and numeracy, this would have implied a problem with test validity. The findings from these two papers are in no way conclusive on the validity of NAPLAN testing. Validity cannot be removed from the intended purpose of a test and how the results are

interpreted, but to thoroughly explore these aspects is beyond the scope of this thesis. Instead, these findings contribute incrementally to establishing the extent of validity of these large-scale tests.

The principal purpose of the second paper was to report on the genetic and environmental influences on individual performance in each domain and grade, and on those influences on the covariation and independence among the domains in each grade. This study is the foundation of the thesis. The sample is large, approximately 2000 twins pairs with close to 1000 pairs in each grade. Averaged across domains and grades, genetic factors explained 60%, shared environment 10%, and unique environment 30% of the variation in NAPLAN achievement. Writing notably deviated from this broad summary with genes and the unique environment each contributing close to 50% of the variation in performance. I propose that the source of the strong unique environmental influence might be due to the creative and generative demands of the task, or it might reflect greater measurement error in the test; however, in these data the source of this unique environmental influence unknown. Inter-rater reliability, criterion-based validity testing, and assessing genetic and environmental covariation with other creative tasks could all be avenues for future studies.

An unexpected finding in these initial univariate analyses was sex-effects on the variance components in some of the domains and/or grades. Occasionally quantitative sex-effects have been reported in studies with measures of reading or mathematics (e.g. Harlaar, Spinath, Dale, & Plomin, 2005; Petrill & Thompson, 1994), but most do not (e.g. Davis et al., 2008; Hart, Soden, Johnson, Schatschneider, & Taylor, 2013; Taylor & Schatschneider, 2010; Wadsworth & DeFries, 2005). The results from this study followed a general pattern of no sex-effects in reading, grammar and punctuation, and writing, but with scalar sex-effects, such that boys had greater variation than girls, in spelling and numeracy. Quantitative sex-effects were occasional exceptions and where present there was a stronger influence from the shared environment in girls and a stronger genetic influence in boys. For reasons discussed in the second paper, I suspect the large sex-effect evident in the NAPLAN Grade 5 reading domain is due to sampling fluctuation; no doubt incorporating data from future waves of NAPLAN participants will help to clarify if this is the case. However, the presence of substantial sex-effects in both Grade 3 numeracy and Grade 5 reading meant that sex-effects were tested in subsequent analyses. In the multivariate test of sex-effects there was no difference between girls and boys in the genetic and environmental influences on covariation among NAPLAN domains. When assessing variation in growth, there were some statistical differences between girls and boys in sources of variation in growth, but except for Grade 3 numeracy they were small in size. Similarly, sex-effects on the moderation of

genetic and environmental effects by SES were evident in some domains, but only significant in Grade 7 grammar and punctuation. Overall in the analyses conducted there is a broad pattern of no sex-effect or of small sex-effects. These effects have been discussed in each of the relevant papers; overall, caution in interpreting sex-effects that deviate from a more general pattern is warranted.

Unlike the unexpected sex-effects, the results from the second part of the second paper were consistent with wider research into genetic and environmental influences on academic performance. The generalist gene hypothesis was supported with high genetic correlations among the domains, and most of the genetic variation in any specific domain resulted from genes that influenced all domains. A general genetic influence on all five domains accounted for most of the genetic influences on each of the domains. Other than this general genetic influence, there was a second set of genetic influences in the earlier grades on some of the literacy domains: spelling, grammar and punctuation, and writing. There were also some independent genetic influences on numeracy. The genetic structure across the domains was largely unchanged throughout the grades. However, by Grade 9, this second set of genetic influences loaded specifically onto spelling instead of contributing to covariation in performance with grammar and punctuation and writing. Although this change in genetic covariance structure was significant by Grade 9, the size of the change resulted from a gradual reduction. Genetic mediation of correlations across domain remained stable, with genes accounting for three quarters of the phenotypic correlations among these achievement domains in each grade.

Stability was also a feature of the genetic mediation of correlations across time. In the third paper, genes mediated most of the phenotypic correlations over time. Although genes predominantly mediated relative stability in performance in all of the domains, it does not follow that genes mediate variation in growth of literacy and numeracy skills. It is well accepted that past performance is predictive of future performance (Hattie, 2008); however, students vary in the rate at which they improve in literacy and numeracy. Consistent with some studies, genes were the principal influence on variation in growth of reading (Christopher et al., 2013a; Christopher et al., 2013b). The genes that influenced the variation in reading at Grade 3 were the same that influenced growth of reading through to Grade 9. Unlike reading, genes did not significantly influence variation in growth of numeracy; instead the unique environment was the predominant influence on variation in growth. As this was the first reported study of its kind on numeracy, replication is important. Future studies might investigate specific environmental factors that possibly influence variation in numeracy growth.

Masters, Rowley, Ainley, & Khoo (2008) argue growth in performance is an indicator of the added value to a student's performance by the school. In these data, twins in a pair attended the same school; as such, school influence would be evident in the shared environmental effects. School-level effects were not a key influence on different rates of growth among school children in either reading or numeracy. Therefore, attributing greater or lesser growth to the school is, perhaps, mistaken. However, some school factors that are experienced uniquely by an individual (e.g. teachers when twins are in separate classes), might contribute to the variation in growth captured by the unique environment, which was particularly important to numeracy.

The final paper of this thesis explored the impact of school and family SES on the influence of genes and the environment on literacy and numeracy. Consistent with wider research, there was a positive correlation of SES with literacy and numeracy (OECD, 2013; Sirin, 2005; White, 1982). Family SES ranged from a small (2% of the variance) to modest (10%) effect size, depending on the subject; and school SES consistently contributed an additional small amount (1−4%). As twins share SES, it was not possible to test the genetic overlap between literacy and numeracy with SES. When SES was partitioned within the model, reduced genetic influence indicated the relationship between domain and SES was partially mediated by genes, at least in some domains. More notable was the substantial reduction in the influence of the shared environment in all domains. This indicates that much of the impact of the shared environment results from factors within the shared environment that systematically vary with SES, and these factors were predominantly at the family level and were captured by parent education. The lack of substantial influence of school SES after controlling for family factors and the small residual influence of the shared environment are reasons to be cautious before attributing performance to school-level factors.

Aside from the correlations between SES and each of the five domains, the final paper focused on whether SES moderated the influence of genes and the environment on literacy and numeracy. Principally there was no significant moderation of genetic influences. As such, there was no evidence of constrained genetic potential in children attending less advantaged schools or who have parents with less education. Nor was there evidence of advantageous family and school environments potentiating genetic expression. Where interactions were significant they typically involved a decrease in environmental variation, both shared and unique, with increasing SES. These environmental effects were most evident in Grade 3 literacy domains. The correlations between SES and the test outcomes indicate a source of inequality in opportunity for children in Australia. However, family SES measured with parental education captured much of this relationship. Little additional variance was explained by school advantage, and the moderation

analyses showed genetic and environmental influences to be largely stable across both family and school SES. These findings indicate that at a systematic level, education in Australia is largely equitable.

An overarching finding from these papers is the small influence of shared environmental factors on individual differences, after controlling for genetic differences. This means that differences in factors such as place of residence, family resources, and school attended has little influence on differences in literacy and numeracy outcomes in Australia. For the wider Australian community, the small school-level effect present in these papers discourages use of NAPLAN results to compile league tables to compare schools. From the perspective of policy-making, these findings support funding decisions that specifically target struggling individuals rather than struggling schools. Such funding could support individualised intervention programs for those struggling students. Given the foremost sources of differences in literacy and numeracy were those that were unique (either genetic or environmental), this argues against a one-size-fits-all approach to assisting students reach acceptable levels of literacy and numeracy.

Limitations

Limitations are discussed in each of the empirical papers; here I recap the limitations influencing the body of work in this thesis. A principal limitation in drawing inferences about individual differences in literacy and numeracy is in assessing performance using only a few large-scale tests. Although there is evidence and rationale to support using these tests, they are a point-in-time estimate of performance, obtained through a specific test medium. Multiple tests on a measure might provide for greater precision in assessment of component skills, or can be modelled to represent a latent variable of the construct of interest. A latent variable would have the advantage of not including test-specific measurement error, thus clarifying the unique environmental influences. While acknowledging this limitation, teachers, principals, and politicians use the results from NAPLAN tests; as such, having an understanding of the influence of genes and the environment on these specific outcomes is important.

Another key question was how representative our sample was of the Australian population. Certainly the dataset represented a wide range of abilities, and the families came from a representative range of educational backgrounds. However, the current participants performed, on average, higher than the national average and attended schools that were more advantaged than the national average. This limits the generalisablity of these findings, particularly to understanding genetic and environmental influences on individual differences among the more disadvantaged students in Australia.

That data was available on twins, and not on their extended family with diverse genetic relationships, limited these analyses to models based on the classic twin design. Non-additive genetic effects were not estimated, nor could assortative mating and gene-environment correlations be estimated. To ameliorate any further source of bias, estimates of the shared environment were retained, even when it was not a statistically significant contributor to the model.

Conclusions

There are several key findings from these studies. First, well-accepted, individually administered tests of reading were preferentially related, both genetically and environmentally, to large-scale reading tests compared to large-scale numeracy tests, imparting some confidence in the validity of such large-scale tests. Second, individual differences on literacy and numeracy through the middle years of school in Australia were strongly influences by genes, with more moderate influences from the unique environment, and a small influence from the shared environment. Although sex-effects on the variance components were evident in some domains and/or grades, these effects were mostly small in size. Third, a single set of genes predominantly mediated the covariation among reading, spelling, grammar and punctuation, writing, and numeracy. Fourth, genes mediated the stability in performance over time in each domain. Fifth, genes also mediated variation in growth of reading, with the same genes at initial performance influencing growth, possibly indicating that variation in growth of reading is reflecting a genetically mediated developmental process in acquiring the component skills of reading. Unlike reading, variation in growth of numeracy was influenced by the unique environment. There was no evidence that school-level factors influenced variation in growth in any domain. Sixth, most of the influence of the shared environment on NAPLAN performance is due family SES, and genetic and environmental influences were reasonably stable across the SES range. On the whole, this provides support for substantial equality in the Australian education system.

References

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., . . . Olson, R. K. (2013a). The genetic and environmental etiologies of individual differences in early reading growth in Australia, the United States, and Scandinavia. *Journal of Experimental Child Psychology, 115*(3), 453-467. doi:10.1016/j.jecp.2013.03.008

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., . . . Olson, R. K. (2013b). Modeling the etiology of individual differences in early reading development: Evidence for strong genetic influences. *Scientific Studies of Reading*, 1-19. doi:10.1080/10888438.2012.729119

Davis, O. S. P., Kovas, Y., Harlaar, N., Busfield, P., McMillan, A., Frances, J., . . . Plomin, R. (2008). Generalist genes and the Internet generation: Etiology of learning abilities by web testing at age 10. *Genes, Brain and Behavior, 7*(4), 455-462. doi:10.1111/j.1601-183X.2007.00370.x

Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: A study of 7-year-old twins. *Journal of Child Psychology and Psychiatry, 46*(4), 373-384. doi:10.1111/j.1469-7610.2004.00358.x

Hart, S. A., Petrill, S. A., & Thompson, L. A. (2010). A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learning and Individual Differences, 20*(2), 63-69. doi:10.1016/j.lindif.2009.10.004

Hart, S. A., Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology, 101*(2), 388. doi:10.1037/a0015115

Hart, S. A., Soden, B., Johnson, W., Schatschneider, C., & Taylor, J. (2013). Expanding the environment: Gene × school-level SES interaction on reading comprehension. *Journal of Child Psychology and Psychiatry, 54*(10), 1047-1055. doi:10.1111/jcpp.12083

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

Kovas, Y., Harlaar, N., Petrill, S. A., & Plomin, R. (2005). "Generalist genes" and mathematics in 7-year-old twins. *Intelligence, 33*(5), 473-489.

Masters, G., Rowley, G., Ainley, J., & Khoo, S. (2008). *Reporting and comparing school performances*. Retrieved from http://research.acer.edu.au/ar_misc/8

Organisation for Economic Co-operation and Development. (2013). *PISA 2013 Results: Excellence through equity: Giving every student the chance to suceed (Volumne II)*. Retrieved from http://dx.doi.org/10.1787/9789264201132-en

Petrill, S. A., & Thompson, L. A. (1994). The effect of gender upon heritability and common environmental estimates in measures of scholastic achievement. *Personality and Individual Differences, 16*(4), 631-640. doi:10.1016/0191-8869(94)90190-2

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453. doi:10.2307/3515987

Taylor, J., & Schatschneider, C. (2010). Genetic influence on literacy constructs in kindergarten and first grade: Evidence from a diverse twin sample. *Behavior Genetics, 40*(5), 591-602. doi:10.1007/s10519-010-9368-7

Wadsworth, S. J., & DeFries, J. C. (2005). Genetic etiology of reading difficulties in boys and girls. *Twin Research and Human Genetics, 8*(06), 594-601.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461-481. doi:10.1037/0033-2909.91.3.461

Appendix

Supplementary Tables

Table A1 Assumption testing model fit comparisons for reading in each grade

Model	Grade 3			Grade 5			Grade 7			Grade 9		
	Δ-2LL	Δdf	p	Δ-2LL	Δdf	p	Δ-2LL	Δdf	p	Δ-2LL	Δdf	p
Saturated	-	-	-	-	-	-	-	-	-	-	-	-
Means 1	1.93	4	0.749	1.15	4	0.886	3.78	4	0.437	1.75	4	0.782
Means 2	0.38	2	0.827	0.48	2	0.787	0.53	2	0.767	1.44	2	0.487
Means 3	4.56	2	0.102	2.79	2	0.248	4.35	2	0.114	9.14	2	0.010
Means 4	6.47	6	0.373	1.61	6	0.952	6.89	6	0.331	9.69	6	0.138
Means 5	0.98	1	0.322	0.19	1	0.663	0.1	1	0.752	0.16	1	0.689
Means 6	1.44	1	0.230	1.41	1	0.235	2.78	1	0.095	2.74	1	0.098
Means 7	2.41	1	0.121	3.11	1	0.078	2.39	1	0.122	0.02	1	0.888
Means 8	27.87	1	0.000	7.91	1	0.005	4.82	1	0.028	1.44	1	0.230
Means 9	10.94	1	0.001	5.11	1	0.024	2.31	1	0.129	0.3	1	0.584
Variances 1	3.83	4	0.430	2.88	4	0.578	1.56	4	0.816	4.14	4	0.387
Variances 2	0.82	2	0.664	2.94	2	0.230	3.17	2	0.205	0.49	2	0.783
Variances 3	4.72	2	0.094	8.83	2	0.012	2.52	2	0.284	1.68	2	0.432
Variances 4	3.54	1	0.060	0.13	1	0.718	0.43	1	0.512	0.69	1	0.406
Covariances 1	0.47	2	0.791	4.97	2	0.083	2.55	2	0.279	0.27	2	0.874
Covariances 2	0	1	1.000	8.93	1	0.003	2.65	1	0.104	0	1	1.000
Covariances 3	119.9	1	0.000	67.2	1	0.000	61.66	1	0.000	58.74	1	0.000

Note. Saturated = all means, variances, and covariances are allowed to differ by birth order, sex, and zygosity; Means 1 = equates the means of twins across birth order; Means 2 = equates the means of same-sex and opposite-sex dizygotic (DZ) twins within sex; Means 3 = equates the means of DZ and monozygotic (MZ) twins within sex; Means 4 = drops cohort effects from the means; Means 5 = drops age-squared-by-sex effects from the means; Means 6 = drops age-by-sex effects from the means; Means 7 = drops age-squared effects from the means; Means 8 = drops age effects from the means; Means 9 = drops sex effects from the means; Variances 1 = equates the variances of twins across birth order; Variances 2 = equates the variances of same-sex and opposite-sex DZ twins within sex; Variances 3 = equates the variances of DZ and MZ twins within sex; Variances 4 = equates variances across females and males; Covariances 1 = equates covariances of females with males; Covariances 2 = equates covariances of same-sex and opposite-sex DZ; Covariances 3 = equates covariances of DZ and MZ twins.

Appendix

Table A2   Assumption testing model fit comparisons for spelling in each grade

Model	Grade 3 Δ-2LL	Δdf	p	Grade 5 Δ-2LL	Δdf	p	Grade 7 Δ-2LL	Δdf	p	Grade 9 Δ-2LL	Δdf	p
Saturated	-	-	-	-	-	-	-	-	-	-	-	-
Means 1	1.65	4	0.800	2.49	4	0.646	0.88	4	0.927	0.33	4	0.988
Means 2	3.39	2	0.184	0.29	2	0.865	0.41	2	0.815	0.53	2	0.767
Means 3	1.38	2	0.502	1.37	2	0.504	1.21	2	0.546	0.45	2	0.799
Means 4	11.74	6	0.068	3.14	6	0.791	5.71	6	0.456	7.16	6	0.306
Means 5	0.02	1	0.888	0.02	1	0.888	1.41	1	0.235	0.15	1	0.699
Means 6	0.01	1	0.920	0.02	1	0.888	2	1	0.157	7.55	1	0.006
Means 7	8.12	1	0.004	9.05	1	0.003	5.47	1	0.019	2.34	1	0.126
Means 8	20.64	1	0.000	6.21	1	0.013	0.39	1	0.532	0.12	1	0.729
Means 9	20.15	1	0.000	19.28	1	0.000	15.6	1	0.000	5.1	1	0.024
Variances 1	9.92	4	0.042	2.66	4	0.616	0.86	4	0.930	5.43	4	0.246
Variances 2	1.85	2	0.397	6.09	2	0.048	1.78	2	0.411	3.05	2	0.218
Variances 3	1.83	2	0.401	12.54	2	0.002	5.5	2	0.064	1.3	2	0.522
Variances 4	4.37	1	0.037	5.41	1	0.020	9.17	1	0.002	0.9	1	0.343
Covariances 1	3.09	2	0.213	1.96	2	0.375	8.68	2	0.013	5.47	2	0.065
Covariances 2	4.98	1	0.026	3.79	1	0.052	0.28	1	0.597	2.57	1	0.109
Covariances 3	141.95	1	0.000	148.11	1	0.000	130.93	1	0.000	125.46	1	0.000

Note. Saturated = all means, variances, and covariances are allowed to differ by birth order, sex, and zygosity; Means 1 = equates the means of twins across birth order; Means 2 = equates the means of same-sex and opposite-sex dizygotic (DZ) twins within sex; Means 3 = equates the means of DZ and monozygotic (MZ) twins within sex; Means 4 = drops cohort effects from the means; Means 5 = drops age-squared-by-sex effects from the means; Means 6 = drops age-by-sex effects from the means; Means 7 = drops age-squared effects from the means; Means 8 = drops age effects from the means; Means 9 = drops sex effects from the means; Variances 1 = equates the variances of twins across birth order; Variances 2 = equates the variances of same-sex and opposite-sex DZ twins within sex; Variances 3 = equates the variances of DZ and MZ twins within sex; Variances 4 = equates variances across females and males; Covariances 1 = equates covariances of females with males; Covariances 2 = equates covariances of same-sex and opposite-sex DZ; Covariances 3 = equates covariances of DZ and MZ twins.

Appendix

Table A3   Assumption testing model fit comparisons for grammar and punctuation in each grade

Model	Grade 3			Grade 5			Grade 7			Grade 9		
	Δ-2LL	Δ*df*	*p*	Δ-2LL	Δ*df*	*p*	Δ-2LL	Δ*df*	*p*	Δ-2LL	Δ*df*	*p*
Saturated	-	-	-	-	-	-	-	-	-	-	-	-
Means 1	1.06	4	0.901	2.03	4	0.730	5.11	4	0.276	4.3	4	0.367
Means 2	1.87	2	0.393	0.13	2	0.937	0.85	2	0.654	0.1	2	0.951
Means 3	1.91	2	0.385	3.67	2	0.160	1.86	2	0.395	5	2	0.082
Means 4	7.33	6	0.291	6.5	6	0.370	8.76	6	0.188	4.52	6	0.607
Means 5	0.5	1	0.480	0.27	1	0.603	0.19	1	0.663	0.12	1	0.729
Means 6	0.65	1	0.420	0	1	1.000	0.46	1	0.498	0.02	1	0.888
Means 7	4.2	1	0.040	3.67	1	0.055	1.25	1	0.264	0.15	1	0.699
Means 8	19.81	1	0.000	3.93	1	0.047	0.98	1	0.322	0.12	1	0.729
Means 9	28.83	1	0.000	18.21	1	0.000	4.79	1	0.029	8.73	1	0.003
Variances 1	7.29	4	0.121	6.03	4	0.197	4.18	4	0.382	7.06	4	0.133
Variances 2	1.18	2	0.554	0.28	2	0.869	0.02	2	0.990	2.77	2	0.250
Variances 3	3.42	2	0.181	5.61	2	0.061	4.13	2	0.127	0.11	2	0.946
Variances 4	0.21	1	0.647	0.02	1	0.888	3.47	1	0.062	0.22	1	0.639
Covariances 1	3.68	2	0.159	0.38	2	0.827	0.58	2	0.748	1.91	2	0.385
Covariances 2	2.37	1	0.124	4.14	1	0.042	0.3	1	0.584	0.04	1	0.841
Covariances 3	60.89	1	0.000	81.97	1	0.000	56.69	1	0.000	35.5	1	0.000

Note. Saturated = all means, variances, and covariances are allowed to differ by birth order, sex, and zygosity; Means 1 = equates the means of twins across birth order; Means 2 = equates the means of same-sex and opposite-sex dizygotic (DZ) twins within sex; Means 3 = equates the means of DZ and monozygotic (MZ) twins within sex; Means 4 = drops cohort effects from the means; Means 5 = drops age-squared-by-sex effects from the means; Means 6 = drops age-by-sex effects from the means; Means 7 = drops age-squared effects from the means; Means 8 = drops age effects from the means; Means 9 = drops sex effects from the means; Variances 1 = equates the variances of twins across birth order; Variances 2 = equates the variances of same-sex and opposite-sex DZ twins within sex; Variances 3 = equates the variances of DZ and MZ twins within sex; Variances 4 = equates variances across females and males; Covariances 1 = equates covariances of females with males; Covariances 2 = equates covariances of same-sex and opposite-sex DZ; Covariances 3 = equates covariances of DZ and MZ twins.

Appendix

Table A4   Assumption testing model fit comparisons for writing in each grade

Model	Grade 3			Grade 5			Grade 7			Grade 9		
	Δ-2LL	Δ*df*	*p*	Δ-2LL	Δ*df*	*p*	Δ-2LL	Δ*df*	*p*	Δ-2LL	Δ*df*	*p*
Saturated	-	-	-	-	-	-	-	-	-	-	-	-
Means 1	5.49	4	0.241	1.64	4	0.802	5.74	4	0.219	1.85	4	0.763
Means 2	1.65	2	0.438	1.37	2	0.504	1.49	2	0.475	0.97	2	0.616
Means 3	0.55	2	0.760	1.76	2	0.415	2.11	2	0.348	0.16	2	0.923
Means 4	10.75	6	0.096	36.71	6	0.000	30.05	6	0.000	23.67	6	0.001
Means 5	0.27	1	0.603	3.14	1	0.076	0.02	1	0.888	0.33	1	0.566
Means 6	0.03	1	0.862	0	1	1.000	1.02	1	0.313	0.1	1	0.752
Means 7	1.39	1	0.238	2.11	1	0.146	6.02	1	0.014	7.08	1	0.008
Means 8	36.16	1	0.000	6.92	1	0.009	0.5	1	0.480	0.69	1	0.406
Means 9	68.12	1	0.000	41.56	1	0.000	54.63	1	0.000	37.57	1	0.000
Variances 1	2.14	4	0.710	1.82	4	0.769	2.51	4	0.643	1.88	4	0.758
Variances 2	0.04	2	0.980	0.45	2	0.799	0.71	2	0.701	1.19	2	0.552
Variances 3	4.58	2	0.101	3.59	2	0.166	1.84	2	0.399	1.41	2	0.494
Variances 4	7.27	1	0.007	0.19	1	0.663	3.53	1	0.060	1.26	1	0.262
Covariances 1	0.62	2	0.733	3.36	2	0.186	1.14	2	0.566	4.37	2	0.112
Covariances 2	0.4	1	0.527	4.23	1	0.040	3.4	1	0.065	2.35	1	0.125
Covariances 3	22	1	0.000	34.59	1	0.000	27.68	1	0.000	17.07	1	0.000

Note. Saturated = all means, variances, and covariances are allowed to differ by birth order, sex, and zygosity; Means 1 = equates the means of twins across birth order; Means 2 = equates the means of same-sex and opposite-sex dizygotic (DZ) twins within sex; Means 3 = equates the means of DZ and monozygotic (MZ) twins within sex; Means 4 = drops cohort effects from the means; Means 5 = drops age-squared-by-sex effects from the means; Means 6 = drops age-by-sex effects from the means; Means 7 = drops age-squared effects from the means; Means 8 = drops age effects from the means; Means 9 = drops sex effects from the means; Variances 1 = equates the variances of twins across birth order; Variances 2 = equates the variances of same-sex and opposite-sex DZ twins within sex; Variances 3 = equates the variances of DZ and MZ twins within sex; Variances 4 = equates variances across females and males; Covariances 1 = equates covariances of females with males; Covariances 2 = equates covariances of same-sex and opposite-sex DZ; Covariances 3 = equates covariances of DZ and MZ twins.

Appendix

Table A5   Assumption testing model fit comparisons for numeracy in each grade

Model	Grade 3			Grade 5			Grade 7			Grade 9		
	Δ-2LL	Δdf	p	Δ-2LL	Δdf	p	Δ-2LL	Δdf	p	Δ-2LL	Δdf	p
Saturated	-	-	-	-	-	-	-	-	-	-	-	-
Means 1	3.07	4	0.546	3.59	4	0.464	1.54	4	0.820	0.72	4	0.949
Means 2	1.61	2	0.447	0.95	2	0.622	0.62	2	0.733	1.63	2	0.443
Means 3	1.03	2	0.598	0.52	2	0.771	2.55	2	0.279	5.32	2	0.070
Means 4	9.74	6	0.136	5.62	6	0.467	4.33	6	0.632	2.99	6	0.810
Means 5	0.2	1	0.655	0.04	1	0.841	2.03	1	0.154	1.26	1	0.262
Means 6	2.2	1	0.138	0.87	1	0.351	0.26	1	0.610	1.51	1	0.219
Means 7	3.62	1	0.057	2.93	1	0.087	1.04	1	0.308	0.27	1	0.603
Means 8	28.6	1	0.000	7.3	1	0.007	0	1	1.000	1.36	1	0.244
Means 9	10.37	1	0.001	35.22	1	0.000	33.03	1	0.000	30.92	1	0.000
Variances 1	3.38	4	0.496	0.32	4	0.988	2.8	4	0.592	6.85	4	0.144
Variances 2	1.28	2	0.527	1.46	2	0.482	1.29	2	0.525	2.78	2	0.249
Variances 3	2.33	2	0.312	2.27	2	0.321	0.28	2	0.869	1.4	2	0.497
Variances 4	10.75	1	0.001	9.87	1	0.002	19.83	2	0.000	4.34	1	0.037
Covariances 1	8.49	2	0.014	11.79	2	0.003	13.82	2	0.001	3.18	2	0.204
Covariances 2	6.62	1	0.010	1.59	1	0.207	1.67	1	0.196	0.08	1	0.777
Covariances 3	101.46	1	0.000	94.64	1	0.000	158.08	1	0.000	82.49	1	0.000

Note. Saturated = all means, variances, and covariances are allowed to differ by birth order, sex, and zygosity; Means 1 = equates the means of twins across birth order; Means 2 = equates the means of same-sex and opposite-sex dizygotic (DZ) twins within sex; Means 3 = equates the means of DZ and monozygotic (MZ) twins within sex; Means 4 = drops cohort effects from the means; Means 5 = drops age-squared-by-sex effects from the means; Means 6 = drops age-by-sex effects from the means; Means 7 = drops age-squared effects from the means; Means 8 = drops age effects from the means; Means 9 = drops sex effects from the means; Variances 1 = equates the variances of twins across birth order; Variances 2 = equates the variances of same-sex and opposite-sex DZ twins within sex; Variances 3 = equates the variances of DZ and MZ twins within sex; Variances 4 = equates variances across females and males; Covariances 1 = equates covariances of females with males; Covariances 2 = equates covariances of same-sex and opposite-sex DZ; Covariances 3 = equates covariances of DZ and MZ twins.

Appendix

Table A6   Univariate sex-limitation model fit comparisons for reading

Model	Model Fit			Change Statistics		
	-2LL	df	AIC	Δ-2LL	Δdf	p
Grade 3						
Full r_C	26844.59	2324	22196.59			
Full r_A	26844.32	2324	22196.32	-	-	-
Common	26844.59	2325	22194.59	0.27	1	.605
Scalar	26844.80	2327	22190.80	0.47	3	.925
Null	**26848.39**	**2328**	**22192.39**	**4.07**	**4**	**.397**
Grade 5						
Full r_C	24663.42	2169	20325.42			
Full r_A	24663.42	2169	20325.42	-	-	-
Common	**24663.83**	**2170**	**20323.83**	**0.41**	**1**	**.522**
Scalar	24675.89	2172	20331.89	12.46	3	.006
Null	24676.18	2173	20330.18	12.76	4	.012
Grade 7						
Full r_C	21559.65	1949	17661.65			
Full r_A	21559.65	1949	17661.65	-	-	-
Common	21560.02	1950	17660.02	0.37	1	.541
Scalar	21564.79	1952	17660.79	5.14	3	.162
Null	**21565.27**	**1953**	**17659.27**	**5.62**	**4**	**.229**
Grade 9						
Full r_C	17443.50	1586	14271.50			
Full r_A	17443.50	1586	14271.50	-	-	-
Common	17443.50	1587	14269.50	0.00	1	.970
Scalar	17443.85	1589	14265.85	0.35	3	.951
Null	**17444.47**	**1590**	**14264.47**	**0.97**	**4**	**.915**

Note. Full = full sex-limitation model, r_C = allows the dizygotic opposite-sex shared environment correlation to be estimated, r_A = allows the dizygotic opposite-sex genetic correlation to be estimated. Common = common effects sex-limitation model. Scalar = scalar effects sex-limitation model. Null model allows no sex effects. In bold is the most parsimonious model with no significant loss of fit from the full sex-limitation model.

Appendix

Table A7   Univariate sex-limitation model fit comparisons for spelling

Model	Model Fit			Change Statistics		
	-2LL	df	AIC	Δ-2LL	Δdf	p
Grade 3						
Full r_C	26195.66	2327	21541.66			
Full r_A	26195.66	2327	21541.66	-	-	-
Common	26196.48	2328	21540.48	0.82	1	.366
Scalar	**26199.47**	**2330**	**21539.47**	**3.81**	**3**	**.283**
Null	26207.31	2331	21545.31	11.65	4	.020
Grade 5						
Full r_C	23995.74	2171	19653.74			
Full r_A	23995.74	2171	19653.74	-	-	-
Common	23995.86	2172	19651.86	0.12	1	.728
Scalar	**23998.66**	**2174**	**19650.66**	**2.92**	**3**	**.405**
Null	24005.88	2175	19655.88	10.14	4	.038
Grade 7						
Full r_C	21181.53	1941	17299.53			
Full r_A	21181.47	1941	17299.47	-	-	-
Common	21181.50	1942	17297.50	0.03	1	.865
Scalar	**21182.75**	**1944**	**17294.75**	**1.28**	**3**	**.734**
Null	21199.19	1945	17309.19	17.72	4	.001
Grade 9						
Full r_C	17409.37	1583	14243.37			
Full r_A	17404.73	1583	14238.73	-	-	-
Common	**17406.33**	**1584**	**14238.33**	**1.60**	**1**	**.206**
Scalar	17414.27	1586	14242.27	9.53	3	.023
Null	17418.07	1587	14244.07	13.33	4	.010

Note. Full = full sex-limitation model, r_C = allows the dizygotic opposite-sex shared environment correlation to be estimated, r_A = allows the dizygotic opposite-sex genetic correlation to be estimated. Common = common effects sex-limitation model. Scalar = scalar effects sex-limitation model. Null model allows no sex effects. In bold is the most parsimonious model with no significant loss of fit from the full sex-limitation model.

Appendix

Table A8   Univariate sex-limitation model fit comparisons for grammar and punctuation

Model	Model Fit			Change Statistics		
	-2LL	df	AIC	Δ-2LL	Δdf	p
Grade 3						
Full r_C	26966.88	2315	22336.88			
Full r_A	26966.88	2315	22336.88	-	-	-
Common	26967.40	2316	22335.40	0.52	1	.472
Scalar	26970.30	2318	22334.30	3.42	3	.331
Null	**26971.75**	**2319**	**22333.75**	**4.87**	**4**	**.301**
Grade 5						
Full r_C	24915.39	2167	20581.39			
Full r_A	24915.39	2167	20581.39	-	-	-
Common	24916.31	2168	20580.31	0.92	1	.337
Scalar	24918.07	2170	20578.07	2.68	3	.444
Null	**24918.16**	2171	**20576.16**	**2.77**	**4**	**.598**
Grade 7						
Full r_C	22137.45	1950	18237.45			
Full r_A	22137.45	1950	18237.45	-	-	-
Common	22137.48	1951	18235.48	0.03	1	.854
Scalar	22138.48	1953	18232.48	1.03	3	.793
Null	**22141.79**	1954	**18233.79**	**4.35**	**4**	**.361**
Grade 9						
Full r_C	17816.65	1584	14648.65			
Full r_A	17816.27	1584	14648.27	-	-	-
Common	17816.65	1585	14646.65	0.38	1	.535
Scalar	17817.85	1587	14643.85	1.58	3	.664
Null	**17818.69**	1588	**14642.69**	**2.42**	**4**	**.659**

Note. Full = full sex-limitation model, r_C = allows the dizygotic opposite-sex shared environment correlation to be estimated, r_A = allows the dizygotic opposite-sex genetic correlation to be estimated. Common = common effects sex-limitation model. Scalar = scalar effects sex-limitation model. Null model allows no sex effects. In bold is the most parsimonious model with no significant loss of fit from the full sex-limitation model.

Appendix

Table A9   Univariate sex-limitation model fit comparisons for writing

Model	Model Fit			Change Statistics		
	-2LL	df	AIC	Δ-2LL	Δdf	p
Grade 3						
Full r_C	25213.04	2316	20581.04			
Full r_A	25212.98	2316	20580.98	-	-	-
Common	25213.04	2317	20579.04	0.06	1	.802
Scalar	25213.38	2319	20575.38	0.41	3	.939
Null	**25221.90**	**2320**	**20581.90**	**8.92**	**4**	**.063**
Grade 5						
Full r_C	23703.21	2151	19401.21			
Full r_A	23703.21	2151	19401.21	-	-	-
Common	23704.70	2152	19400.70	1.48	1	.223
Scalar	23706.82	2154	19398.82	3.60	3	.308
Null	**23709.22**	**2155**	**19399.22**	**6.00**	**4**	**.199**
Grade 7						
Full r_C	21830.43	1936	17958.43			
Full r_A	21830.43	1936	17958.43	-	-	-
Common	21830.44	1937	17956.44	0.01	1	.927
Scalar	21831.52	1939	17953.52	1.09	3	.779
Null	**21835.29**	**1940**	**17955.29**	**4.86**	**4**	**.302**
Grade 9						
Full r_C	18155.48	1576	15003.48			
Full r_A	18155.23	1576	15003.23	-	-	-
Common	18155.48	1577	15001.48	0.25	1	.618
Scalar	18159.33	1579	15001.33	4.10	3	.251
Null	**18161.92**	**1580**	**15001.92**	**6.68**	**4**	**.154**

Note. Full = full sex-limitation model, r_C = allows the dizygotic opposite-sex shared environment correlation to be estimated, r_A = allows the dizygotic opposite-sex genetic correlation to be estimated. Common = common effects sex-limitation model. Scalar = scalar effects sex-limitation model. Null model allows no sex effects. In bold is the most parsimonious model with no significant loss of fit from the full sex-limitation model.

Appendix

Table A10 Univariate sex-limitation model fit comparisons for numeracy

Model	Model Fit			Change Statistics		
	-2LL	*df*	AIC	Δ-2LL	Δ*df*	*p*
Grade 3						
Full r_C	25815.68	2313	21189.68			
Full r_A	25815.68	2313	21189.68	-	-	-
Common	**25816.53**	**2314**	**21188.53**	**0.85**	**1**	**0.355**
Scalar	25823.88	2316	21191.88	8.21	3	0.042
Null	25839.47	2317	21205.47	23.79	4	0.000
Grade 5						
Full r_C	24124.80	2169	19786.80			
Full r_A	24124.80	2169	19786.80	-	-	-
Common	24125.04	2170	19785.04	0.24	1	0.622
Scalar	**24128.47**	**2172**	**19784.47**	**3.67**	**3**	**0.299**
Null	24147.04	2173	19801.04	22.24	4	0.000
Grade 7						
Full r_C	21585.03	1944	17697.03			
Full r_A	21585.03	1944	17697.03	-	-	-
Common	21585.14	1945	17695.14	0.10	1	0.746
Scalar	**21587.18**	**1947**	**17693.18**	**2.15**	**3**	**0.541**
Null	21619.19	1948	17723.19	34.16	4	0.000
Grade 9						
Full r_C	17353.72	1569	14215.72			
Full r_A	17353.72	1569	14215.72	-	-	-
Common	17353.72	1570	14213.72	0.00	1	0.966
Scalar	17354.08	1572	14210.08	0.36	3	0.949
Null	**17361.06**	**1573**	**14215.06**	**7.34**	**4**	**0.119**

Note. Full = full sex-limitation model, r_C = allows the dizygotic opposite-sex shared environment correlation to be estimated, r_A = allows the dizygotic opposite-sex genetic correlation to be estimated. Common = common effects sex-limitation model. Scalar = scalar effects sex-limitation model. Null model allows no sex effects. In bold is the most parsimonious model with no significant loss of fit from the full sex-limitation model.

Appendix

Table A11 Unstandardized parameter estimates from the best-fitting model for each NAPLAN domain and grade

	Grade 3		Grade 5		Grade 7	Grade 9	
Reading							
a	72.41		38.97 F	63.96 M	50.88	50.65	
c	19.70		50.50 F	10.93 M	24.89	23.14	
e	41.75		43.45 F	39.52 M	34.93	33.02	
Spelling							
a	62.50		57.39		51.77	52.58 F	54.53 M
c	8.19		11.16		0.00	21.62 F	-15.88 M
e	33.72		28.93		27.55	28.55 F	33.76 M
GP							
a	63.60		66.38		59.56	51.81	
c	39.23		20.65		18.94	24.65	
e	48.40		43.38		42.25	41.41	
Write							
a	35.94		43.76		48.31	49.12	
c	20.19		15.85		8.31	17.09	
e	40.01		39.37		47.19	54.65	
Numeracy							
a	42.03 F	63.56 M	51.00		57.58	55.08	
c	40.86 F	12.63 M	26.72		12.12	24.77	
e	33.28 F	36.77 M	31.34		26.99	31.62	

Note: *a* = additive genetic path estimate; *c* = shared environmental path estimate; *e* = unique environmental path estimate; F = female; M = male.

Appendix

Table A12 Phenotypic correlations among NAPLAN domains in Grades 3,
5, 7, and 9

Grade 3	Reading	Spelling	GP	Writing	Numeracy
Reading	-				
Spelling	.65	-			
GP	.72	.70	-		
Writing	.55	.60	.58	-	
Numeracy	.68	.61	.69	.51	-
Grade 5	Reading	Spelling	GP	Writing	Numeracy
Reading	-				
Spelling	.61	-			
GP	.69	.68	-		
Writing	.54	.60	.59	-	
Numeracy	.66	.58	.68	.53	-
Grade 7	Reading	Spelling	GP	Writing	Numeracy
Reading	-				
Spelling	.62	-			
GP	.72	.67	-		
Writing	.56	.56	.58	-	
Numeracy	.67	.60	.68	.54	-
Grade 9	Reading	Spelling	GP	Writing	Numeracy
Reading	-				
Spelling	.59	-			
GP	.71	.65	-		
Writing	.55	.56	.60	-	
Numeracy	.66	.57	.68	.53	-

Note. GP = Grammar and Punctuation.

Appendix

Table A13 Multivariate sex-limitated nonscalar correlated factors model fit comparisons

Model	Model Fit			Change Statistics		
	-2LL	df	AIC	Δ-2LL	Δdf	p
Grade 3						
F ≠ M	125361.77	11485	102391.77	-	-	-
F = M	125480.31	11580	102320.31	118.54	95	0.051
Grade 5						
F ≠ M	116374.46	10692	94990.46	-	-	-
F = M	116486.77	10787	94912.77	112.31	95	0.109
Grade 7						
F ≠ M	103674.46	9589	84496.46	-	-	-
F = M	103780.32	9684	84412.32	105.86	95	0.210
Grade 9						
F ≠ M	84488.47	7759	68970.47	-	-	-
F = M	84586.47	7854	68878.47	97.99	95	0.396

Note. F ≠ M female and male path estimates and correlations allowed to differ; F = M female and male path estimates and correlations constrained to be equal.

Appendix

Table A14 Multivariate model fit comparisons for the independent and common pathway models

Model	Model Fit			Change Statistics		
	-2LL	df	AIC	Δ-2LL	Δdf	p
Grade 3						
CorFac	125480.31	11580	102320.31	-	-	-
Ind2a2c2e	125484.62	11583	102318.62	4.31	3	0.230
Ind2a2ce	125489.25	11587	102315.25	8.94	7	0.257
Ind2ac2e	125490.32	11587	102316.32	10.01	7	0.188
Inda2c2e	125515.29	11587	102341.29	34.98	7	0.000
Ind2ace	**125494.24**	**11591**	**102312.24**	**13.93**	**11**	**0.237**
IndACE	125554.6	11595	102364.6	74.29	15	0.000
ComACE	125642.25	11603	102436.25	161.94	23	0.000
Grade 5						
CorFac	116486.77	10787	94912.77	-	-	-
Ind2a2c2e	116488.92	10790	94908.92	2.15	3	0.543
Ind2a2ce	116498.92	10794	94910.92	12.15	7	0.096
Ind2ac2e	116492.87	10794	94904.87	6.1	7	0.529
Inda2c2e	116497.15	10794	94909.15	10.38	7	0.168
Ind2ace	**116500.76**	**10798**	**94904.76**	**13.99**	**11**	**0.233**
IndACE	116522.61	10802	94918.61	35.84	15	0.002
ComACE	116608.2	10810	94988.2	121.43	23	0.000
Grade 7						
CorFac	103780.32	9684	84412.32	-	-	-
Ind2a2c2e	103781.81	9687	84407.81	1.49	3	0.685
Ind2a2ce	103786.1	9691	84404.1	5.78	7	0.566
Ind2ac2e	103783.46	9691	84401.46	3.13	7	0.872
Inda2c2e	103798.44	9691	84416.44	18.12	7	0.011
Ind2ace	**103787.26**	**9695**	**84397.26**	**6.93**	**11**	**0.805**
IndACE	103811.96	9699	84413.96	31.63	15	0.007
ComACE	290952.61	9707	271538.61	187172.29	23	0.000
Grade 9						
CorFac	84586.47	7854	68878.47	-	-	-
Ind2a2c2e	84589.54	7857	68875.54	3.08	3	0.380
Ind2a2ce	84593.57	7861	68871.57	7.1	7	0.418
Ind2ac2e	84590.42	7861	68868.42	3.95	7	0.785
Inda2c2e	84598.33	7861	68876.33	11.87	7	0.105
Ind2ace	84593.79	7865	68863.79	7.33	11	0.772
IndACE	**84605.32**	**7869**	**68867.32**	**18.86**	**15**	**0.220**
ComACE	84648.58	7877	68894.58	62.11	23	0.000

Note. CorFac = correlated factors model; Ind2a2c2e = base independent pathway model where the latent genetic, shared environmental, and unique environmental structure is modelled each with a common factor loading onto each of the five domains, a common factor loading onto the four literacy domains, and specific factors loading onto each domain; Ind2a2ce = from the base independent pathway model, the unique environmental structure common factor loading onto the four literacy domains is dropped; Ind2ac2e = from the base independent pathway model, the shared environment common factor loading onto the four literacy domains is dropped; Inda2c2e = from the base independent pathway model, the genetic common factor loading onto the four literacy domains is dropped; Ind2ace = from the base independent pathway model, the shared and unique environmental common factors loading onto the four literacy domains are dropped; IndACE = for genes, the shared and unique environment the common factors loading onto the four literacy domains are dropped; ComACE = common pathway model; In bold is the best-fitting model with no significant loss of fit from the correlated factors model.

Appendix

Table A15 Model fit statistics for family SES moderation of genetic and environmental influences on performance in Grade 3 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δdf	p
Reading	Moderated	1211.49	-	-	-
	βa = 0	**1210.45**	**0.96**	**1**	**.327**
	βc = 0	1215.16	5.66	1	.017*
	βe = 0	1214.86	5.37	1	.021*
	βa = βc = 0	1215.26	7.77	2	.021*
	βa = βe = 0	1215.22	7.73	2	.021*
	βc = βe = 0	1213.19	5.70	2	.058
	βa = βc = βe = 0	1213.5	8.00	3	.046*
	$β_{SES}$ family	1372.89	169.4	4	<.001*
Spelling	Moderated[a]	1258.32	-	-	-
	βa = 0	**1257.33**	**1.01**	**1**	**.315**
	βc = 0	1265.49	9.17	1	.002*
	βe = 0	1266.17	9.86	1	.002*
	βa = βc = 0	1266.89	12.57	2	.002*
	βa = βe = 0	1268.94	14.62	2	.001*
	βc = βe = 0	1264.22	9.90	2	.007*
	βa = βc = βe = 0	1267	14.68	3	.002*
	$β_{SES}$ family	1339.06	88.74	4	<.001*
GP	Moderated	1274.91	-	-	-
	βa = 0	**1274.11**	**1.20**	**1**	**.273**
	βc = 0	1277.88	4.97	1	.026*
	βe = 0	1278.76	5.85	1	.016*
	βa = βc = 0	1279.7	8.79	2	.012*
	βa = βe = 0	1285.42	14.51	2	.001*
	βc = βe = 0	1277.09	6.18	2	.046*
	βa = βc = βe = 0	1283.42	14.51	3	.002*
	$β_{SES}$ family	1416.99	150.08	4	<.001*
Writing	Moderated[a]	1530.18	-	-	-
	βa = 0	1528.64	0.45	1	.502
	βc = 0	1531.48	3.30	1	.069
	βe = 0	1528.5	0.31	1	.576
	βa = βc = 0	1540.94	14.76	2	.001*
	βa = βe = 0	**1527.54**	**1.36**	**2**	**.508**
	βc = βe = 0	1532.17	5.98	2	.005*
	βa = βc = βe = 0	1539.11	14.93	3	.002*
	$β_{SES}$ family	1614.18	91.99	4	<.001*
Numeracy	Moderated[a]	1214.01	-	-	-
	βa = 0	1212.19	0.18	1	.674
	βc = 0	1212.98	0.97	1	.325
	βe = 0	1212.12	0.11	1	.744
	βa = βc = 0	1211	0.99	2	.609
	βa = βe = 0	1210.2	0.18	2	.912
	βc = βe = 0	1211.09	1.08	2	.584
	βa = βc = βe = 0	**1209.09**	**1.08**	**3**	**.783**
	$β_{SES}$ family	1349.22	143.21	4	<.001

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δdf = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

[a]Included sex as a second moderator

* significant loss of fit when parameter(s) dropped

Appendix

Table A16 Model fit statistics for family SES moderation of genetic and environmental influences on performance in Grade 5 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δ*df*	*p*
Reading	Moderated[a]	1178.77	-	-	-
	βa = 0	1177.05	0.28	1	.600
	βc = 0	1176.79	0.02	1	.900
	βe = 0	1179.10	2.33	1	.127
	βa = βc = 0	1175.06	0.29	2	.864
	βa = βe = 0	1178.94	4.17	2	.125
	βc = βe = 0	1177.32	2.54	2	.280
	βa = βc = βe = 0	**1177.29**	**4.52**	**3**	**.211**
	β$_{SES}$ family	1314.05	143.28	4	<.001
Spelling	**Moderated**	**1118.61**	**-**	**-**	**-**
	βa = 0	1123.20	6.60	1	.010*
	βc = 0	1119.35	2.74	1	.098
	βe = 0	1120.40	3.79	1	.051
	βa = βc = 0	1126.23	11.62	2	.003*
	βa = βe = 0	1124.89	10.29	2	.006*
	βc = βe = 0	1120.97	6.36	2	.042*
	βa = βc = βe = 0	1125.15	12.55	3	.006*
	β$_{SES}$ family	1183.68	73.08	4	<.001*
GP	Moderated	1199.96	-	-	-
	βa = 0	1198.76	0.81	1	.369
	βc = 0	1200.85	2.89	1	.089
	βe = 0	1200.94	2.98	1	.084
	βa = βc = 0	1200.10	4.14	2	.126
	βa = βe = 0	1200.81	4.86	2	.088
	βc = βe = 0	1198.94	2.99	2	.225
	βa = βc = βe = 0	**1198.82**	**4.86**	**3**	**.182**
	β$_{SES}$ family	1302.77	110.82	4	<.001*
Writing	Moderated	1344.82	-	-	-
	βa = 0	1343.42	0.60	1	.439
	βc = 0	1343.25	0.43	1	.513
	βe = 0	1343.00	0.18	1	.673
	βa = βc = 0	1341.42	0.60	2	.740
	βa = βe = 0	1341.63	0.81	2	.666
	βc = βe = 0	1341.35	0.53	2	.768
	βa = βc = βe = 0	**1339.86**	**1.04**	**3**	**.792**
	β$_{SES}$ family	1422.18	85.36	4	<.001*
Numeracy	Moderated[a]	1029.81	-	-	-
	βa = 0	1027.87	0.06	1	.799
	βc = 0	1027.81	0	1	.997
	βe = 0	1027.86	0.05	1	.822
	βa = βc = 0	1026.18	0.37	2	.830
	βa = βe = 0	1027.26	1.45	2	.485
	βc = βe = 0	1026.48	0.67	2	.714
	βa = βc = βe = 0	**1025.67**	**1.86**	**3**	**.602**
	β$_{SES}$ family	1137.50	115.69	4	<.001*

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δ*df* = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

[a]Included sex as a second moderator

* significant loss of fit when parameter(s) dropped

Appendix

Table A17 Model fit statistics for family SES moderation of genetic and environmental influences on performance in Grade 7 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δ*df*	*p*
Reading	Moderated	978.03	-	-	-
	βa = 0	976.11	0.08	1	.775
	βc = 0	976.17	0.15	1	.700
	βe = 0	976.03	0.01	1	.941
	βa = βc = 0	976.17	2.15	2	.341
	βa = βe = 0	974.29	0.27	2	.875
	βc = βe = 0	976.23	2.21	2	.332
	βa = βc = βe = 0	**974.95**	**2.92**	**3**	**.404**
	β$_{SES}$ family	1090.13	120.10	4	<.001*
Spelling	Moderated[a]	1026.25	-	-	
	βa = 0	1024.91	0.65	1	.419
	βc = 0	1024.30	0.05	1	.824
	βe = 0	1024.28	0.03	1	.865
	βa = βc = 0	1022.95	0.70	2	.705
	βa = βe = 0	1023.03	0.78	2	.677
	βc = βe = 0	1022.31	0.06	2	.972
	βa = βc = βe = 0	**1021.09**	**0.84**	**3**	**.839**
	β$_{SES}$ family	1067.05	48.8	4	<.001*
GP	Moderated[b]	1087.53	-	-	-
	βa*sex = 0	1089.83	4.30	1	.038*
	βc*sex = 0	1085.93	0.40	1	.526
	βe*sex= 0	1085.60	0.07	1	.791
	βa*sex = βc*sex = 0	1094.30	10.77	2	.005*
	βa*sex = βe*sex = 0	1093.56	10.03	2	.007*
	βc*sex = βe*sex = 0	**1084.36**	**0.83**	**2**	**.660**
	βa*sex = βc*sex = βe*sex = 0	1092.69	11.17	3	.011*
	β$_{SES}$ family $_{sex}$	1091.27	11.74	4	.019*
Writing	Moderated	1221.60	-	-	-
	βa = 0	1219.64	0.04	1	.845
	βc = 0	1219.61	0	1	.962
	βe = 0	1219.79	0.18	1	.670
	βa = βc = 0	1217.93	0.33	2	.850
	βa = βe = 0	1220.67	3.07	2	.216
	βc = βe = 0	1219.68	2.08	2	.354
	βa = βc = βe = 0	**1220.80**	**5.20**	**3**	**.158**
	β$_{SES}$ family	1306.98	93.38	4	<.001*
Numeracy	Moderated[a]	782.05	-	-	-
	βa = 0	780.81	0.76	1	.384
	βc = 0	780.85	0.80	1	.371
	βe = 0	780.83	0.78	1	.377
	βa = βc = 0	778.90	0.85	2	.653
	βa = βe = 0	779.02	0.97	2	.617
	βc = βe = 0	779.33	1.28	2	.527
	βa = βc = βe = 0	**777.51**	**1.46**	**3**	**.692**
	β$_{SES}$ family	819.67	45.62	4	<.001*

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δ*df* = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

[a]Included sex as a second moderator

* significant loss of fit when parameter(s) dropped

Appendix

Table A18 Model fit statistics for family SES moderation of genetic and environmental influences on performance in Grade 9 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δ*df*	*p*
Reading	Moderated	763.73	-	-	-
	βa = 0	761.74	0	1	.947
	βc = 0	762.66	0.92	1	.336
	βe = 0	762.73	0.99	1	.319
	βa = βc = 0	762.31	2.57	2	.276
	βa = βe = 0	762.93	3.20	2	.202
	βc = βe = 0	760.73	1	2	.607
	βa = βc = βe = 0	**761.02**	**3.29**	**3**	**.349**
	βSES family	839.43	83.7	4	<.001*
Spelling	Moderateda	862.80	-	-	-
	βa = 0	862.08	1.28	1	.258
	βc = 0	861.17	0.37	1	.543
	βe = 0	864.01	3.20	1	.073
	βa = βc = 0	860.30	1.50	2	.473
	βa = βe = 0	863.07	4.27	2	.119
	βc = βe = 0	862.13	3.33	2	.189
	βa = βc = βe = 0	**861.16**	**4.35**	**3**	**.226**
	βSES family	883.77	28.97	4	<.001*
GP	Moderated	881.44	-	-	-
	βa = 0	880.13	0.70	1	.404
	βc = 0	880.65	1.21	1	.271
	βe = 0	880.31	0.87	1	.350
	βa = βc = 0	879.67	2.23	2	.328
	βa = βe = 0	878.35	0.91	2	.633
	βc = βe = 0	879.73	2.29	2	.318
	βa = βc = βe = 0	**878.17**	**2.73**	**3**	**.435**
	βSES family	940.98	67.54	4	<.001*
Writing	Moderated	977.49	-	-	-
	βa = 0	975.70	0.21	1	.645
	βc = 0	976.28	0.79	1	.375
	βe = 0	976.01	0.52	1	.472
	βa = βc = 0	974.59	1.10	2	.577
	βa = βe = 0	974.10	0.62	2	.735
	βc = βe = 0	974.32	0.84	2	.658
	βa = βc = βe = 0	**972.59**	**1.11**	**3**	**.776**
	βSES family	1030.81	61.32	4	<.001*
Numeracy	Moderateda	616.34	-	-	-
	βa = 0	616.75	2.41	1	.121
	βc = 0	621.81	7.47	1	.006*
	βe = 0	617.01	2.67	1	.102
	βa = βc = 0	620.27	7.93	2	.019*
	βa = βe = 0	**615.23**	**2.89**	**2**	**.236**
	βc = βe = 0	617.21	4.87	2	.088
	βa = βc = βe = 0	615.23	4.89	3	.180
	βSES family	695.60	87.26	4	<.001*

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δ*df* = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

aIncluded sex as a second moderator

* significant loss of fit when parameter(s) dropped

Appendix

Table A19 Model fit statistics for school SES moderation of genetic and environmental influences on performance in Grade 3 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δ*df*	*p*
Reading	Moderated	1266.91	-	-	-
	βa = 0	1265.11	0.20	1	.651
	βc = 0	1266.13	1.22	1	.269
	βe = 0	1265.81	0.90	1	.343
	βa = βc = 0	1265.05	2.15	2	.341
	βa = βe = 0	1264.35	1.45	2	.485
	βc = βe = 0	1264.77	1.86	2	.394
	βa = βc = βe = 0	**1263.3**	**2.39**	**3**	**.495**
	βSES school	1329.14	70.23	4	<.001*
Spelling	Moderateda	1274.55	-	-	-
	βa = 0	**1273.05**	**0.50**	**1**	**.480**
	βc = 0	1277.03	4.47	1	.034*
	βe = 0	1275.90	3.35	1	.067
	βa = βc = 0	1275.46	4.91	2	.086
	βa = βe = 0	1273.98	3.42	2	.181
	βc = βe = 0	1275.06	4.50	2	.105
	βa = βc = βe = 0	1273.46	4.91	3	.179
	βSES school	1285.81	19.26	4	.001*
GP	Moderated	1300.19	-	-	-
	βa = 0	1298.19	0	1	.980
	βc = 0	1298.37	0.18	1	.668
	βe = 0	1298.22	0.04	1	.849
	βa = βc = 0	1296.76	0.57	2	.753
	βa = βe = 0	1296.28	0.09	2	.954
	βc = βe = 0	1298.36	2.17	2	.338
	βa = βc = βe = 0	**1296.36**	**2.17**	**3**	**.538**
	βSES school	1323.97	31.78	4	<.001*
Writing	Moderateda	1526.18	-	-	-
	βa = 0	1524.43	0.25	1	.619
	βc = 0	1529.15	4.97	1	.026*
	βe = 0	1526.14	1.96	1	.161
	βa = βc = 0	1529.76	7.58	2	.023*
	βa = βe = 0	**1524.17**	**1.99**	**2**	**.369**
	βc = βe = 0	1527.38	5.20	2	.074
	βa = βc = βe = 0	1528.11	7.93	3	.047*
	βSES school	1582.68	64.5	4	<.001*
Numeracy	Moderateda	1238.31	-	-	-
	βa = 0	1236.34	0.03	1	.869
	βc = 0	1237.01	0.70	1	.403
	βe = 0	1236.49	0.17	1	.677
	βa = βc = 0	1235.17	0.86	2	.651
	βa = βe = 0	1234.50	0.19	2	.910
	βc = βe = 0	1235.09	0.78	2	.679
	βa = βc = βe = 0	**1233.19**	**0.87**	**3**	**.832**
	βSES school	1256.28	25.97	4	<.001*

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δ*df* = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

aIncluded sex as a second moderator

* significant loss of fit when parameter(s) dropped

Appendix

Table A20 Model fit statistics for school SES moderation of genetic and environmental influences on performance in Grade 5 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δdf	p
Reading	Moderated[a]	1227.65	-	-	-
	βa = 0	1225.69	0.04	1	.839
	βc = 0	1226.36	0.70	1	.402
	βe = 0	1225.65	0	1	.991
	βa = βc = 0	1224.46	0.81	2	.666
	βa = βe = 0	1223.71	0.06	2	.972
	βc = βe = 0	1224.36	0.71	2	.701
	βa = βc = βe = 0	**1222.48**	**0.82**	**3**	**.844**
	β$_{SES}$ school	1297.41	77.76	4	<.001*
Spelling	**Moderated**	**1135.5**	**-**	**-**	**-**
	βa = 0	1139.32	5.81	1	.016*
	βc = 0	1136.76	3.26	1	.071
	βe = 0	1137.88	4.38	1	.036*
	βa = βc = 0	1137.57	6.06	2	.048
	βa = βe = 0	1137.36	5.86	2	.054
	βc = βe = 0	1137.56	6.06	2	.048*
	βa = βc = βe = 0	1136.83	7.33	3	.062
	β$_{SES}$ school	1149.87	22.36	4	<.001*
GP	Moderated	1250.06	-	-	-
	βa = 0	1249.04	0.98	1	.322
	βc = 0	1248.22	0.17	1	.684
	βe = 0	1248.48	0.43	1	.514
	βa = βc = 0	1247.65	1.59	2	.451
	βa = βe = 0	1247.30	1.24	2	.538
	βc = βe = 0	1246.90	0.84	2	.657
	βa = βc = βe = 0	**1245.69**	**1.64**	**3**	**.651**
	β$_{SES}$ school	1310.56	68.5	4	<.001*
Writing	Moderated	1327.15	-	-	-
	βa = 0	1325.61	0.46	1	.497
	βc = 0	1325.88	0.73	1	.393
	βe = 0	1325.61	0.46	1	.499
	βa = βc = 0	1324.35	1.20	2	.548
	βa = βe = 0	1323.62	0.46	2	.793
	βc = βe = 0	1324.72	1.56	2	.458
	βa = βc = βe = 0	**1322.73**	**1.58**	**3**	**.665**
	β$_{SES}$ school	1400.6	81.45	4	<.001*
Numeracy	Moderated[a]	1066.82	-	-	-
	βa = 0	1065.05	0.23	1	.633
	βc = 0	1065.63	0.81	1	.368
	βe = 0	1065.38	0.55	1	.457
	βa = βc = 0	1065.07	2.24	2	.326
	βa = βe = 0	1063.85	1.03	2	.599
	βc = βe = 0	1064.40	1.58	2	.454
	βa = βc = βe = 0	**1063.18**	**2.35**	**3**	**.503**
	β$_{SES}$ school	1152.28	93.46	4	<.001*

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δdf = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

[a]Included sex as a second moderator

* significant loss of fit when parameter(s) dropped

Appendix

Table A21 Model fit statistics for school SES moderation of genetic and environmental influences on performance in Grade 7 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δ*df*	*p*
Reading	Moderated	1006.18	-	-	-
	βa = 0	1004.40	0.22	1	.636
	βc = 0	1005.49	1.31	1	.252
	βe = 0	1004.37	0.20	1	.657
	βa = βc = 0	1004.40	2.23	2	.328
	βa = βe = 0	1002.41	0.23	2	.891
	βc = βe = 0	1006.00	3.83	2	.148
	βa = βc = βe = 0	**1004.07**	**3.89**	**3**	**.273**
	βSES school	1066.28	68.10	4	<.001*
Spelling	Moderateda	1028.76	-	-	-
	βa = 0	1027.09	0.33	1	.564
	βc = 0	1027.32	0.56	1	.452
	βe = 0	1027.15	0.39	1	.532
	βa = βc = 0	1025.36	0.61	2	.739
	βa = βe = 0	1025.15	0.40	2	.820
	βc = βe = 0	1026.14	1.38	2	.502
	βa = βc = βe = 0	**1024.16**	**1.40**	**3**	**.704**
	βSES school	1040.11	19.35	4	.001*
GP	Moderated	1127.67	-	-	-
	βa = 0	1126.44	0.77	1	.381
	βc = 0	1127.74	2.07	1	.150
	βe = 0	1127.04	1.37	1	.243
	βa = βc = 0	1126.61	2.93	2	.231
	βa = βe = 0	1125.21	1.53	2	.465
	βc = βe = 0	1126.37	2.70	2	.259
	βa = βc = βe = 0	**1124.76**	**3.09**	**3**	**.378**
	βSES school	1182.16	62.48	4	<.001*
Writing	Moderated	1199.47	-	-	-
	βa = 0	1197.58	0.11	1	.739
	βc = 0	1197.48	0.01	1	.928
	βe = 0	1197.47	0	1	.963
	βa = βc = 0	1195.86	0.39	2	.822
	βa = βe = 0	1196.44	0.97	2	.614
	βc = βe = 0	1195.49	0.03	2	.987
	βa = βc = βe = 0	**1194.47**	**1.01**	**3**	**.800**
	βSES school	1277.15	85.68	4	<.001*
Numeracy	Moderateda	781.09	-	-	-
	βa = 0	782.27	3.18	1	.074
	βc = 0	781.90	2.81	1	.094
	βe = 0	785.97	6.88	1	.009*
	βa = βc = 0	**780.27**	**3.18**	**2**	**.204**
	βa = βe = 0	784.62	7.52	2	.023*
	βc = βe = 0	784.01	6.91	2	.032*
	βa = βc = βe = 0	782.65	7.56	3	.056
	βSES school	856.81	83.71	4	<.001*

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δ*df* = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

aIncluded sex as a second moderator

* significant loss of fit when parameter(s) dropped

Appendix

Table A22 Model fit statistics for school SES moderation of genetic and environmental influences on performance in Grade 9 NAPLAN tests

Domain	Model	AIC	Δ-2LL	Δdf	p
Reading	Moderated	800.35	-	-	-
	βa = 0	**799.03**	**0.68**	**1**	**.409**
	βc = 0	801.73	3.37	1	.066
	βe = 0	802.32	3.96	1	.047*
	βa = βc = 0	800.59	4.24	2	.120
	βa = βe = 0	802.22	5.87	2	.053
	βc = βe = 0	800.32	3.96	2	.138
	βa = βc = βe = 0	800.29	5.94	3	.115
	βSES school	835.71	43.36	4	<.001*
Spelling	Moderateda	866.51	-	-	-
	βa = 0	865.13	0.61	1	.433
	βc = 0	865.03	0.52	1	.470
	βe = 0	868.35	3.84	1	.050
	βa = βc = 0	863.31	0.80	2	.671
	βa = βe = 0	866.93	4.42	2	.110
	βc = βe = 0	866.64	4.13	2	.127
	βa = βc = βe = 0	**865.36**	**4.85**	**3**	**.183**
	βSES school	882.99	24.48	4	<.001*
GP	Moderated	892.65	-	-	-
	βa = 0	891.23	0.58	1	.446
	βc = 0	891.42	0.77	1	.381
	βe = 0	890.99	0.34	1	.559
	βa = βc = 0	889.46	0.81	2	.667
	βa = βe = 0	889.38	0.73	2	.693
	βc = βe = 0	890.46	1.82	2	.403
	βa = βc = βe = 0	**888.59**	**1.94**	**3**	**.585**
	βSES school	929.31	44.66	4	<.001*
Writing	Moderated	972.15	-	-	-
	βa = 0	971.14	0.99	1	.319
	βc = 0	970.18	0.03	1	.869
	βe = 0	971.01	0.85	1	.356
	βa = βc = 0	969.59	1.44	2	.487
	βa = βe = 0	969.21	1.06	2	.589
	βc = βe = 0	969.58	1.43	2	.489
	βa = βc = βe = 0	**968.01**	**1.86**	**3**	**.602**
	βSES school	1043.72	79.57	4	<.001*
Numeracy	Moderateda	626.26	-	-	-
	βa = 0	625.98	1.72	1	.189
	βc = 0	625.73	1.47	1	.225
	βe = 0	626.13	1.87	1	.172
	βa = βc = 0	624.00	1.74	2	.419
	βa = βe = 0	624.13	1.87	2	.392
	βc = βe = 0	625.00	2.75	2	.253
	βa = βc = βe = 0	**623.01**	**2.75**	**3**	**.431**
	βSES school	678.45	60.19	4	<.001*

Note. SES = socioeconomic status; AIC = Akaike's Information Criterion; Δ-2LL = change in -2 log likelihood; Δdf = change in degrees of freedom; a = additive genetic; c = shared environment; e = unique environment. Bold indicates the best-fitting determined from both AIC and tests of individual moderating terms.

aIncluded sex as a second moderator

* significant loss of fit when parameter(s) dropped