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

Katrina L. Grasby[1,2,5] · William L. Coventry[1] · Brian Byrne[1,2,3] ·
Richard K. Olson[4] · Sarah E. Medland[5]

Abstract We examined the extent to which genes and the environment contributed to variation in and covariation among reading, spelling, grammar and punctuation, writing, and numeracy in Australian school children in Grades 3, 5, 7, and 9. Heritability was generally high: reading .58–.71 (excepting Grade 5 girls), spelling .68–.78; grammar and punctuation .52–.66, writing .39–.52, and numeracy .39–.79. Boys' performance varied more than girls in spelling and numeracy, and the common environment was a greater influence in girls than boys in Grade 3 numeracy and Grade 5 reading. 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. The covariation among the domains was largely mediated by genes. Results suggest substantial uniformity in the environmental factors influencing these academic domains.

Keywords Reading · Numeracy · Academic achievement · Twins · Sex differences · Generalist genes

✉ Katrina L. Grasby
   kgrasby2@une.edu.au

[1] Discipline of Psychology, School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale, NSW, Australia

[2] Australian Research Council Centre of Excellence for Cognition and its Disorders, Sydney, NSW, Australia

[3] National Health and Medical Research Council Centre of Excellence in Twin Research, Canberra, ACT, Australia

[4] Department of Psychology, University of Colorado, Boulder, CO, USA

[5] Queensland Institute of Medical Research, 300 Herston Rd, Herston, Brisbane, QLD, Australia

Introduction

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 math 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 et al. 2005a; Petrill et al. 2007; Samuelsson et al. 2008; Taylor and 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 et al. 2005a), a trained test administrator (e.g. Petrill et al. 2007), or via a phone or web 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 et al. 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 et al. 2007; Haworth et al. 2009; Kovas et al. 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 et al. 2008; Hart et al. 2010; Taylor and Schatschneider 2010). The Twin Early Development Study (TEDS) in the UK is by far the largest twin study (Trouton et al. 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 pre-school 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 et al. 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 students 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.

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 math ability in 7- and 9-year-old twins was estimated to be .66 (Oliver et al. 2004) and .68 respectively (Haworth et al. 2007). In slightly older twins, the heritability of various math skills was noticeably lower, ranging from .32 to .45 at age 10 (Kovas et al. 2007) and .49 at age 12 (Haworth et al. 2009). In the younger twins math 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 math 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 math 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 math skills at age 8 was negligible (0–.14; except for a test of math fluency, which was .63; Hart et al. 2009). Although, subsequent analyses at age 10 showed heritability of math skills to range from .34 to .49 (Hart et al. 2010), and when math ability was modelled as a latent variable from various math skills, heritability estimates at age 10 was .41 and at age 11 was .34 (Petrill et al. 2012). These heritability estimates from 10 and 11-year-olds in the WRRMP are close to those obtained from math skills in the UK. However, markedly higher estimates come from the CLDRC where, using similar math tests to the WRRMP and modelling math as a latent variable, heritability estimates of .67 (Knopik and DeFries 1999) and .90 (Alarcón et al. 2000) have been reported. Although the CLDRC project have a sample of twins identified as having at least one twin with learning difficulties, these cited estimates were 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 and DeFries 1999) estimates from the WRRMP were a more substantial. Shared environmental estimates of math skills ranged from .15 to .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 math 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 math knowledge accounts for some of the variance in students' mathematical achievement (Baumert et al. 2010; 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 math content becoming more complex. Consequently, we might find a greater influence of the environment in secondary 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 math (e.g. Hedges and Nowell 1995; Hyde 2005), it does not necessarily follow that the genetic and environmental influences on variation in performance are different. Thus far, when sex differences for literacy or mathematics are found the results across studies are diverse and have not been replicated with any consistency. There is support for qualitative sex-effects in reading at age 7 (Harlaar et al. 2005b). Quantitative sex-effects have also been found: Greater heritability has been indicated in girls for reading but not math (Petrill and Thompson 1994), which contrasts with greater heritability in boys for reading and math (Knopik et al. 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 math (Reynolds et al. 1996; Spinath et al. 2008). Finally, there are numerous studies that find no significant sex-effects in various measures of reading and math (e.g. Davis et al.

2008; Hart et al. 2013; Taylor and Schatschneider 2010; Wadsworth and 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 et al. 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, 2010, 2011, 2012, 2013, 2014). 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 and 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 and 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 et al. 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 and 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 et al. 2009; Harlaar et al. 2012; Hart et al. 2010; Knopik and 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 and DeFries (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 math, and Harlaar et al. (2012) with a correlation of .71 between math and reading comprehension. In a latent factor model, Hart et al. (2010) found no specific unique environmental contribution to a number of math 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 if genes are generalists, but not if unique environments are specialists. In line with findings from the UK and USA we also hypothesize some unique genetic influences on math and reading ability.

Method

Below is a brief description of participants, materials, and procedure, which are described in full in Grasby and Coventry (2016).

Participants

Participants were 1940 twins, recruited through the voluntary Australian Twin Registry. State education departments provided NAPLAN results for the twins.

Zygosity was determined with a short questionnaire (Lykken et al. 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 et al. 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. As NAPLAN testing began in 2008, although these data are longitudinal over Grades 3, 5, 7 and 9 the initial grade assessed and number of waves of data on a given participant depends on their cohort (i.e. calendar year of taking a grade). In Grade 3 there were 1184 twin pairs, in Grade 5 there were 1113 (779 with data in Grade 3), in Grade 7 there were 1001 (727 in Grade 5 and 417 in Grade 3), and in Grade 9 there were 819 (663 in Grade 7, 417 in Grade 5, and 152 in Grade 3). 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 an Australian, standardised assessment in reading, writing, language conventions, and numeracy for

students in Grades 3, 5, 7, and 9. Scores range from 0 to 1000 and are calibrated to compare with previous cohorts and across grades. Tests are predominantly multiple-choice with some short answer questions, except for the writing test, which requires students to write several paragraphs in response to a prompt. The reading test assesses reading comprehension. The language conventions test provides a score for spelling and one for grammar and punctuation. For Grades 3 and 5 there is one numeracy test, for Grades 7 and 9 the students sit two numeracy tests, one that allows the use of calculators and one that does not.

Procedure

The NAPLAN tests are administered over three consecutive days in the second full week of May (approximately 3.5 months into the school year). 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. The saturated model included age, age-squared, age-by-sex, age-squared-by-sex, and cohort effects (which were dummy coded to indicate calendar year) as covariates, and means, variances, and covariances were estimated separately for each zygosity by sex group. Parameters were then 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 nested models to a χ^2 distribution with degrees of freedom equal to the difference in estimated parameters from the nested models (Neale and Maes 2004). To correct for multiple testing, we used an alpha of .01. 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 Fig. 1). When twins are raised together, the difference in same-sex DZ and opposite-sex DZ correlations inform the calculation of both r_A and r_C, thus only one can be estimated in any given model. As these models
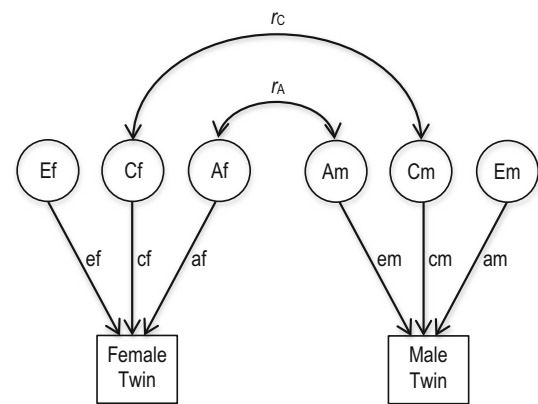

Fig. 1 Path diagram of full sex-limitation model depicting opposite-sex twin pair. Either the shared environment correlation (r_C) or the genetic correlation (r_A) is estimated. Additive genetic (A), shared environmental (C), and unique environmental (E) variance is estimated separately for females (f) and males (m). Paths af, cf, ef, am, cm, em = A, C and E effects on the trait for females and males respectively

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 and 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.

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 et al. 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 (As, Cs, Es). 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.

Results

Assumption Testing

Distributions for each domain in each grade were normal. Scores were truncated at ± 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 onward. Sex significantly influenced means, except for reading in Grades 7 and 9, grammar and punctuation in Grade 7, and spelling in Grade 9. To keep models parsimonious, covariates were included only when they had a significant effect on the means.

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 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 a third of a standard deviation higher than boys. For numeracy, the mean differences increased over time, with boys scoring about a 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 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).

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-fitting statistics available from first author). 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. The best-fitting model for each domain in each grade is reported in Table 3 with univariate genetic and environmental estimates.

For reading, heritability estimates were large for most grades. 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 estimate 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 pattern that did not reach significance to help 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

Table 1 Descriptive statistics by sex for NAPLAN domains at each grade

Variable	Females			Males			Cohen's d	Full sample		
	M	SD	n	M	SD	n		M	SD	n
Reading										
Grade 3	452	84	1207	440	89	1127	0.14	446	86	2334
Grade 5	526	77	1091	521	76	1088	0.07	523	76	2179
Grade 7	580	66	999	576	67	958	0.05	578	66	1957
Grade 9	617	64	831	619	65	766	−0.03	618	64	1594
Spelling										
Grade 3	432	72	1209	417	78	1129	0.21	425	75	2338
Grade 5	512	65	1094	498	70	1087	0.20	505	68	2181
Grade 7	568	59	999	556	65	951	0.19	562	62	1950
Grade 9	606	63	831	599	65	767	0.11	603	64	1592
Grammar and punctuation										
Grade 3	460	88	1203	441	89	1121	0.22	451	89	2324
Grade 5	534	82	1091	519	82	1085	0.18	526	82	2176
Grade 7	576	72	1001	568	77	957	0.11	572	75	1958
Grade 9	612	70	826	604	72	767	0.12	608	71	1593
Writing										
Grade 3	444	56	1206	422	60	1120	0.38	433	59	2326
Grade 5	512	61	1089	493	62	1078	0.31	502	62	2167
Grade 7	563	66	999	539	71	952	0.35	551	70	1951
Grade 9	610	75	827	584	78	765	0.33	597	77	1592
Numeracy										
Grade 3	416	68	1205	426	75	1118	−0.15	421	71	2323
Grade 5	500	66	1093	519	74	1086	−0.27	510	71	2179
Grade 7	565	65	1001	586	76	952	−0.30	575	72	1953
Grade 9	609	65	822	634	71	756	−0.36	621	69	1578

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

MZ monozygotic, *DZ* dizygotic, *n* individuals

boys the estimates remain quite stable over the grades. From looking at the intraclass correlations it seems that the significant sex-effect 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.

Table 2 Intraclass correlations by zygosity and sex for NAPLAN domains at each grade

Variable	Intraclass correlations													
	MZF		MZM		DZF		DZM		DZOS		MZ		DZ	
	r	n	r	n	r	n	r	n	r	n	r	n	r	n
Reading														
Grade 3	.75	262	.75	245	.42	181	.44	162	.43	303	.75	507	.43	646
Grade 5	.68	228	.68	221	.54	166	.46	170	.30	284	.68	449	.41	620
Grade 7	.70	210	.72	201	.52	164	.48	153	.38	230	.71	411	.45	547
Grade 9	.72	195	.73	169	.46	123	.46	120	.43	168	.72	364	.44	411
Spelling														
Grade 3	.78	262	.76	245	.45	180	.48	163	.36	305	.77	507	.41	648
Grade 5	.78	230	.76	222	.40	166	.51	172	.41	281	.77	452	.44	619
Grade 7	.78	213	.74	197	.44	163	.31	155	.36	223	.76	410	.36	541
Grade 9	.80	196	.73	166	.39	125	.29	121	.24	166	.77	362	.30	412
Grammar and punctuation														
Grade 3	.72	260	.65	242	.46	179	.49	161	.41	299	.69	502	.45	639
Grade 5	.73	231	.67	221	.43	165	.46	168	.33	280	.70	452	.40	613
Grade 7	.63	213	.70	199	.42	163	.38	153	.38	229	.66	412	.39	545
Grade 9	.68	193	.62	166	.33	124	.42	123	.40	171	.65	359	.39	418
Writing														
Grade 3	.51	262	.49	242	.32	177	.36	160	.34	302	.52	504	.35	639
Grade 5	.61	228	.53	220	.40	164	.39	168	.29	277	.58	448	.35	609
Grade 7	.50	212	.53	197	.36	162	.27	152	.20	228	.52	409	.28	542
Grade 9	.54	194	.45	164	.22	124	.38	123	.18	171	.51	358	.26	418
Numeracy														
Grade 3	.75	259	.73	239	.55	182	.46	162	.38	301	.74	498	.44	645
Grade 5	.79	229	.73	219	.55	167	.49	171	.42	284	.76	448	.48	622
Grade 7	.84	211	.81	199	.48	164	.45	152	.37	229	.83	410	.43	545
Grade 9	.79	196	.79	166	.44	121	.43	115	.45	168	.80	362	.45	404

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

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. The lowest correlations were with writing, ranging from .51 to .60 and averaged .56 across all grades and domains. Numeracy was as highly correlated with each of the remaining literacy domains (ranging from .57 to .69 and averaged .65 across all grades and domains) as those literacy domains were with each other (ranging from .59 to .72 and averaged .67 across all grades and domains). 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. 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 to .96 with an average genetic correlation of .81 across all domains and grades (see Tables 4, 5, 6 and 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 to .37

Table 3 Univariate estimates from the best-fitting model NAPLAN reading

	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]	
Grammar and punctuation						
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]	

95 % confidence intervals are in square brackets; standard errors are in parenthesis

A, standardised additive genetic variance; C, standardised shared environmental variance; E, standardised unique environmental variance; Common, common effects sex-limitation model; Scalar, scalar effects sex-limitation model; Null model allows no sex effects; F, females; M, males

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 4, 5, 6 and 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

Table 4 Genetic and environmental correlations among NAPLAN domains in Grade 3 (below the diagonal), portion of phenotypic correlation mediated by genetic and environmental factors (above the diagonal), and total variance explained by genetic, shared, and unique environment are in bold on the diagonal

	Reading	Spelling	GP	Writing	Numeracy
Genetic					
Reading	**.70 [.58, .78]**	.53	.57	.43	.56
Spelling	.74 [.68, .80]	**.72 [.61, .79]**	.52	.51	.47
GP	.94 [.86, 1]	.86 [.80, .93]	**.51 [.39, .64]**	.38	.48
Writing	.81 [.69, .91]	.96 [.84, 1]	.84 [.70, .98]	**.39 [.28, .51]**	.37
Numeracy	.83 [.77, .89]	.69 [.62, .77]	.84 [.77, .93]	.74 [.64, .86]	**.64 [.53, .74]**
Shared environment					
Reading	**.06 [0, .17]**	.05	.08	.06	.07
Spelling	.89 [− .70, 1]	**.06 [0, .15]**	.10	.05	.07
GP	.80 [− .32, 1]	.92 [.53, 1]	**.19 [.07, .30]**	.12	.14
Writing	.64 [− .83, 1]	.54 [− .21, 1]	.77 [.32, 1]	**.13 [.04, .23]**	.11
Numeracy	.87 [.29, 1]	.83 [− .07, 1]	.93 [.59, 1]	.92 [.42, 1]	**.11 [.02, .22]**
Unique environment					
Reading	**.24 [.21, .27]**	.07	.07	.05	.06
Spelling	.31 [.23, .38]	**.22 [.20, .26]**	.08	.07	.05
GP	.27 [.18, .35]	.31 [.23, .38]	**.30 [.26, .34]**	.08	.06
Writing	.15 [.07, .23]	.20 [.12, .28]	.20 [.12, .28]	**.48 [.43, .53]**	.03
Numeracy	.25 [.25, .33]	.22 [.14, .30]	.24 [.16, .32]	.08 [0, .16]	**.24 [.21, .28]**

95 % confidence intervals in square brackets

GP grammar and punctuation

Table 5 Genetic and environmental correlations among NAPLAN domains in Grade 5 (below the diagonal), portion of phenotypic correlation mediated by genetic and environmental factors (above the diagonal), and total variance explained by genetic, shared, and unique environment are in bold on the diagonal

	Reading	Spelling	GP	Writing	Numeracy
Genetic					
Reading	**.62 [.49, .71]**	.57	.56	.45	.50
Spelling	.83 [.74, .94]	**.76 [.64, .82]**	.62	.51	.47
GP	.90 [.85, .99]	.90 [.82, .98]	**.62 [.49, .70]**	.46	.51
Writing	.84 [.72, .98]	.85 [.76, .96]	.86 [.76, .98]	**.47 [.33, .58]**	.34
Numeracy	.84 [.75, .94]	.70 [.61, .79]	.85 [.79, .94]	.66 [.56, .77]	**.58 [.47, .70]**
Shared environment					
Reading	**.09 [.02, .21]**	.00	.09	.06	.10
Spelling	.01 [− 1, 1]	**.04 [0, .15]**	.02	.05	.05
GP	.95 [.26, 1]	.31 [− 1, 1]	**.10 [.02, .21]**	.09	.13
Writing	.62 [− .48, 1]	.76 [− 1, 1]	.83 [.20, 1]	**.12 [.03, .23]**	.14
Numeracy	.79 [.07, 1]	.57 [− 1, 1]	.94 [.56, 1]	.97 [.67, 1]	**.19 [.08, .29]**
Unique environment					
Reading	**.29 [.25, .33]**	.03	.05	.03	.06
Spelling	.14 [.04, .22]	**.20 [.17, .23]**	.05	.06	.06
GP	.18 [.10, .27]	.22 [.13, .30]	**.29 [.25, .33]**	.03	.05
Writing	.08 [− .01, .17]	.20 [.11, .28]	.10 [.01, .18]	**.42 [.37, .47]**	.04
Numeracy	.23 [.14, .31]	.30 [.22, .39]	.18 [.10, .27]	.14 [.06, .23]	**.23 [.20, .26]**

95 % confidence intervals in square brackets

GP grammar and punctuation

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

Table 6 Genetic and environmental correlations among NAPLAN domains in Grade 7 (below the diagonal), portion of phenotypic correlation mediated by genetic and environmental factors (above the diagonal), and total variance explained by genetic, shared, and unique environment are in bold on the diagonal

	Reading	Spelling	GP	Writing	Numeracy
Genetic					
Reading	**.59 [.45, .72]**	.45	.53	.42	.53
Spelling	.69 [.61, .77]	**.74 [.64, .80]**	.54	.48	.48
GP	.91 [.85, .98]	.83 [.75, .92]	**.58 [.45, .69]**	.44	.55
Writing	.82 [.73, .93]	.82 [.74, .92]	.86 [.77, 1]	**.46 [.31, .55]**	.45
Numeracy	.80 [.74, .87]	.65 [.59, .70]	.84 [.77, .93]	.77 [.69, .87]	**.75 [.64, .83]**
Shared environment					
Reading	**.14 [.02, .26]**	.07	.12	.09	.10
Spelling	.94 [− 1, 1]	**.04 [0, .12]**	.05	.05	.05
GP	.96 [.48, 1]	.80 [− 1, 1]	**.11 [.01, .22]**	.07	.08
Writing	.98[− 1, 1]	.99 [− 1, 1]	.87 [− 1, 1]	**.07 [0, .18]**	.07
Numeracy	.99 [.38, 1]	.98 [− 1, 1]	.90 [− 1, 1]	1 [− 1, 1]	**.07 [0, .18]**
Unique environment					
Reading	**.28 [.24, .32]**	.09	.08	.05	.05
Spelling	.36 [.28, .44]	**.22 [.19, .26]**	.09	.05	.07
GP	.26 [.18, .34]	.32 [.23, .40]	**.32 [.27, .36]**	.05	.05
Writing	.14 [.06, .23]	.16 [.07, .25]	.14 [.05, .22]	**.48 [.42, .54]**	.01
Numeracy	.23 [.14, .31]	.34 [.26, .42]	.22 [.13, .31]	.05 [− .04, .14]	**.18 [.15, .21]**

95 % confidence intervals in square brackets

GP grammar and punctuation

Table 7 Genetic and environmental correlations among NAPLAN domains in Grade 9 (below the diagonal), portion of phenotypic correlation mediated by genetic and environmental factors (above the diagonal), and total variance explained by genetic, shared, and unique environment are in bold on the diagonal

	Reading	Spelling	GP	Writing	Numeracy
Genetic					
Reading	**.62 [.47, .74]**	.46	.52	.41	.43
Spelling	.67 [.58, .75]	**.75 [.67, .80]**	.56	.48	.45
GP	.91 [.84, .97]	.88 [.80, .95]	**.53 [.40, .66]**	.44	.49
Writing	.78 [.64, .96]	.83 [.73, .99]	.91 [.78, 1]	**.44 [.26, .55]**	.38
Numeracy	.70 [.62, .81]	.66 [.56, .75]	.85 [.75, .94]	.73 [.60, .94]	**.61 [.47, .75]**
Shared environment					
Reading	**.12 [.01, .26]**	.04	.12	.07	.14
Spelling	.98 [− 1, 1]	**.02 [0, .09]**	.04	.02	.05
GP	.86 [.23, 1]	.73 [− 1, 1]	**.15 [.03, .27]**	.07	.13
Writing	.77 [− 1, 1]	.67 [− 1, 1]	.79 [− 1, 1]	**.06 [0, .22]**	.09
Numeracy	.96 [.46, 1]	.94 [− 1, 1]	.80 [.29, 1]	.86 [− 1, 1]	**.18 [.04, .30]**
Unique environment					
Reading	**.26 [.22, .30]**	.08	.07	.09	.09
Spelling	.32 [.23, .41]	**.23 [.20, .28]**	.06	.07	.08
GP	.26 [.17, .35]	.22 [.13, .31]	**.32 [.28, .38]**	.08	.07
Writing	.24 [.15, .33]	.22 [.13, .31]	.19 [.10, .27]	**.50 [.43, .57]**	.06
Numeracy	.37 [.27, .45]	.34 [.25, .43]	.26 [.17, .35]	.17 [.07, .26]	**.21 [.18, .25]**

95 % confidence intervals in square brackets

GP grammar and punctuation

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 2, 3, 4 and 5 show the standardized path estimates of the best-fitting independent pathway models for each grade; dashed lines indicate paths with 95 % confidence intervals that cross zero.

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 grades 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.

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. 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,
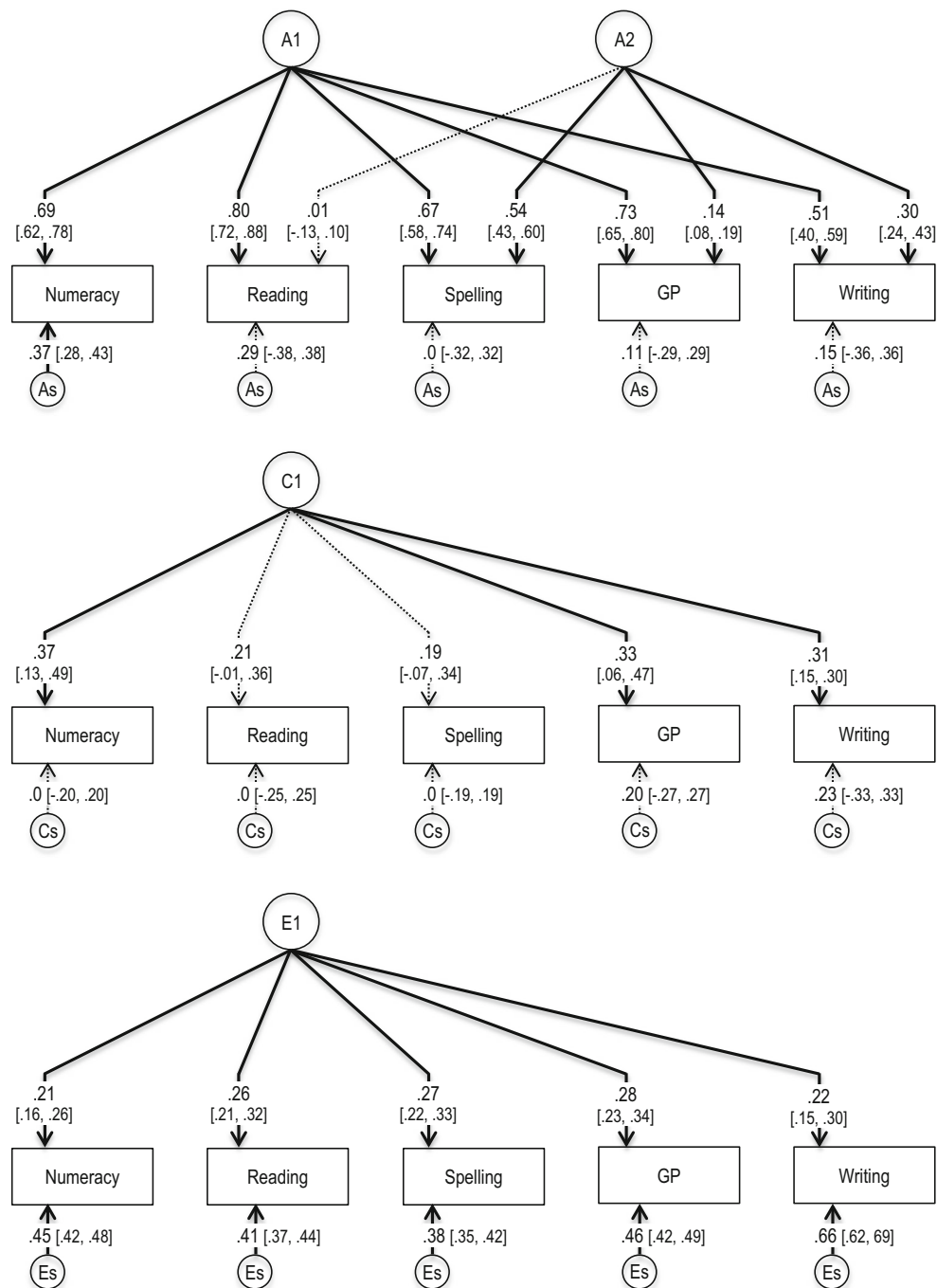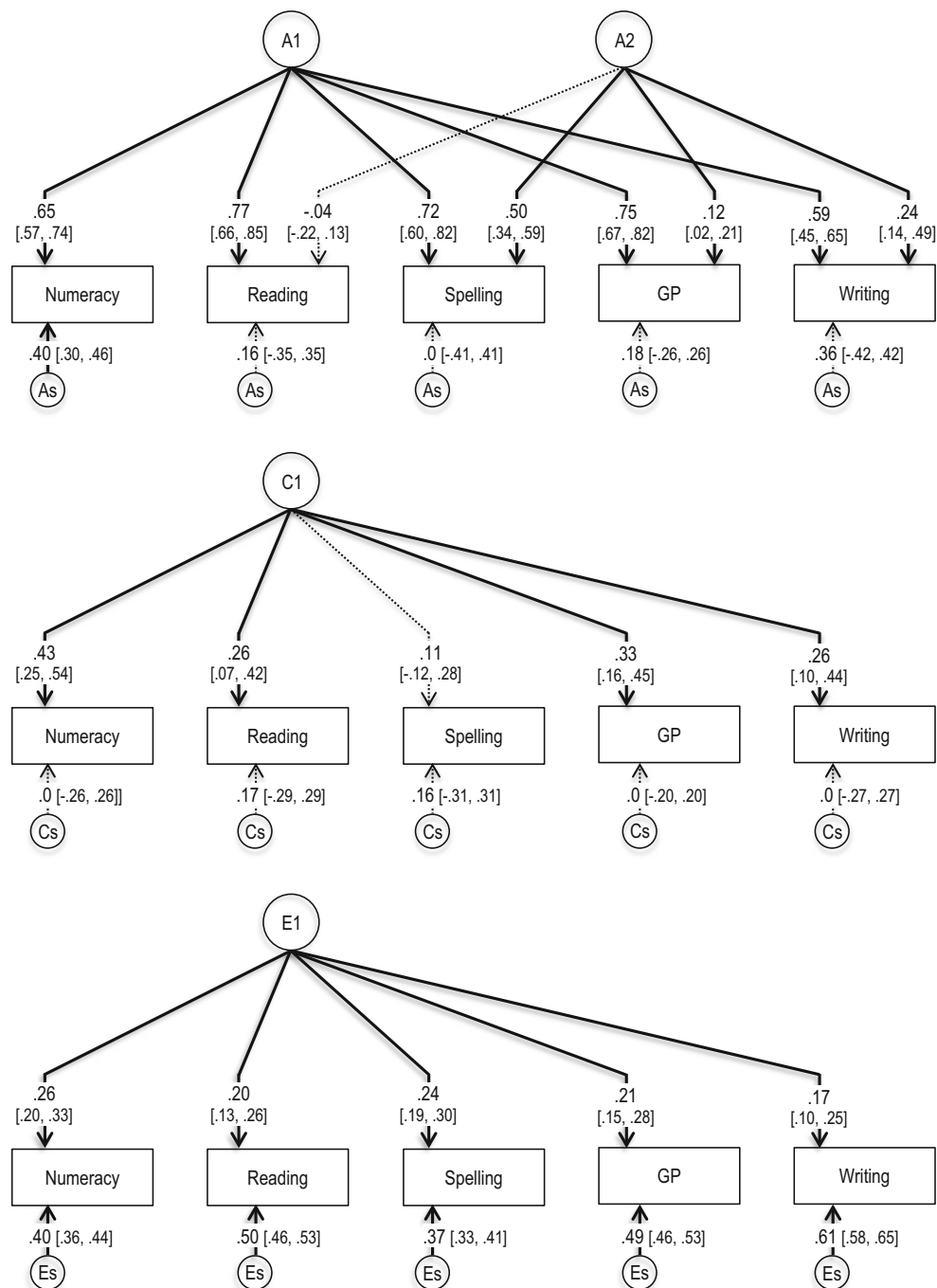
Fig. 2 Path diagram of Grade 3 independent pathway model with standardised parameter estimates. *Solid lines* indicate significant paths, where 95 % confidence intervals did not include zero. The model has a general additive genetic factor (A1), a literacy only factor (A2), a general shared environmental factor (C), a general unique environmental factor (E1), and domain specific factors (As, Cs, Es). Most genetic influences were shared across domains; a second genetic factor influenced spelling, grammar and punctuation (GP), and writing. Some independent genetic effects influenced numeracy. All shared environmental influences were general

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 secondary school compared to primary school.
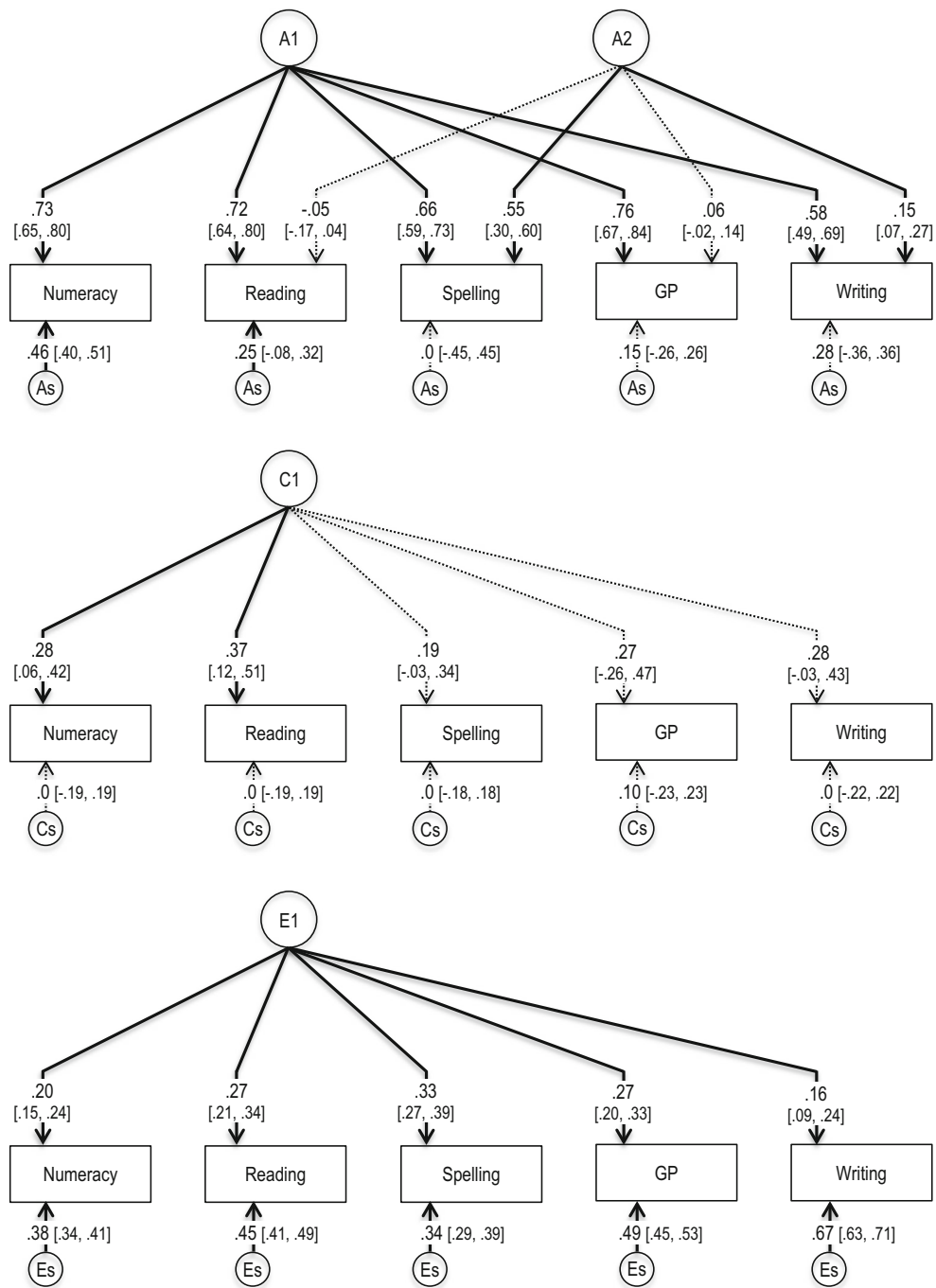
Fig. 3 Path diagram of Grade 5 independent pathway model with standardised parameter estimates. *Solid lines* indicate significant paths, where 95 % confidence intervals did not include zero. The model has a general additive genetic factor (A1), a literacy only factor (A2), a general shared environmental factor (C), a general unique environmental factor (E1), and domain specific factors (As, Cs, Es). Most genetic influences were shared across domains; a second genetic factor influenced spelling, grammar and punctuation (GP), and writing. Some independent genetic effects influenced numeracy. All shared environmental influences were general

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
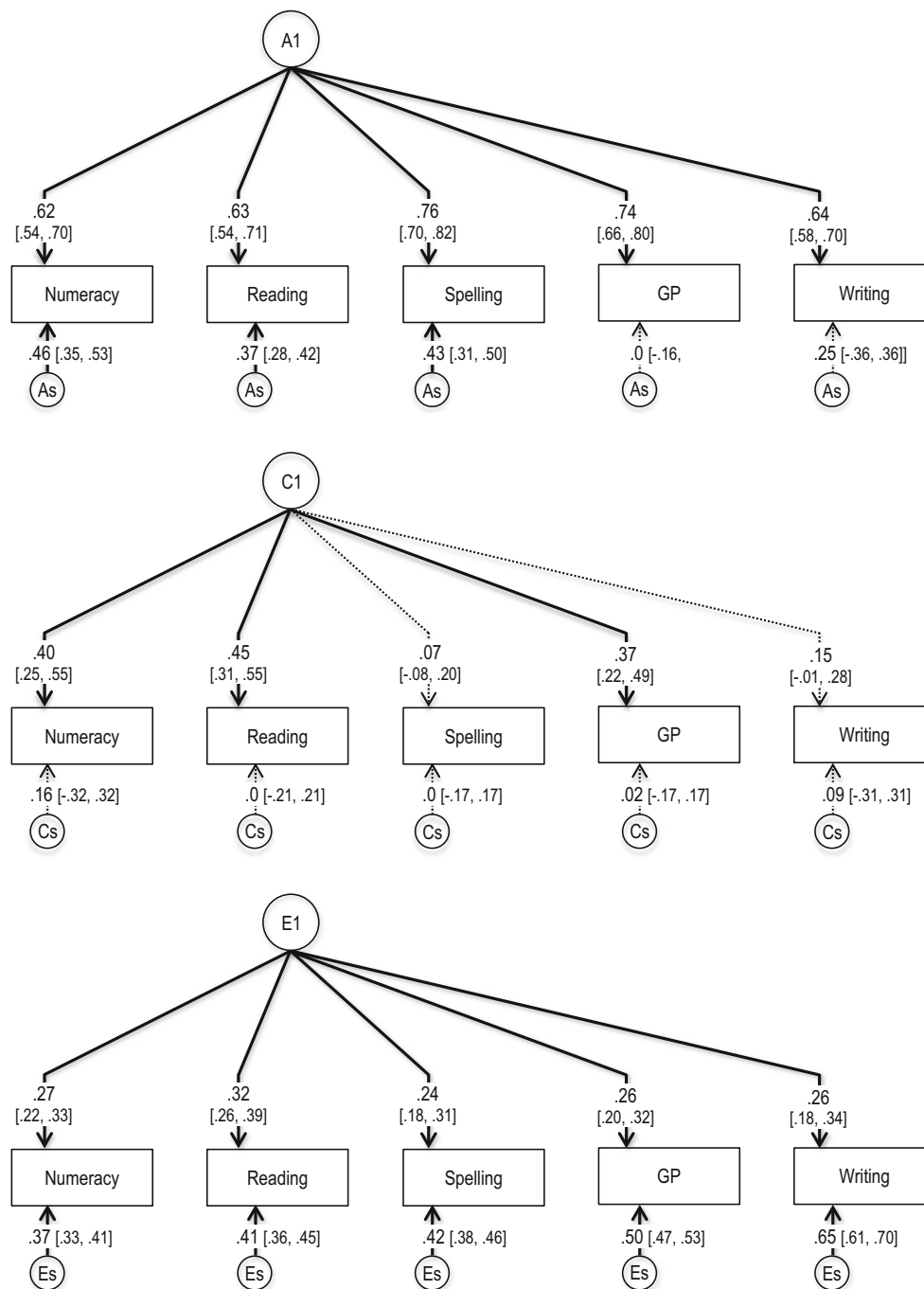
Fig. 4 Path diagram of Grade 7 independent pathway model with standardised parameter estimates. *Solid lines* indicate significant paths, where 95 % confidence intervals did not include zero. The model has a general additive genetic factor (A1), a literacy only factor (A2), a general shared environmental factor (C), a general unique environmental factor (E1), and domain specific factors (As, Cs, Es). 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

style merely differ in difficulty as the grades progress. The dramatic changes in heritability estimates of girls from 71 % in Grade 3 down 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

Fig. 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 (As, Cs, Es). Most genetic influences were shared across domains; there were some literacy specific genetic influences on spelling, grammar and punctuation (GP), and writing. Some independent genetic effects influenced numeracy, reading, and spelling. All shared environmental influences were general

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 a little 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 artifact 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 et al. 2009) and wider research on academic domains (Hedges and Nowell 1995; Machin and 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. Compared to our results, Oliver et al. also had a much lower unique environmental variance component for writing (.27), possibly indicative of less error in their measurement. 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 math knowledge 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' math 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 math anxiety of the teacher was related to the girls' math achievement at the end of the year. This was not the case for boys. In a similar vein, Bleeker and Jacobs (2004) found that mother's beliefs about math and gender had a differential influence on girls and boys, albeit in older participants. These are examples of possible environmental sources of variation in math 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.

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, there was a trend for phenotypic variance to decrease with increasing grade, except writing, which increased in variance. While in numeracy variance was stable across the grades. As such, there was no evidence of increased environmental influence on performance in high school. There are important concerns regarding teachers who end up teaching out-of-field, including teacher morale (Hobbs 2015). However, our findings suggest the specialization of subjects in secondary school in conjunction with possible increased diversity in teacher quality from teaching out-of-field does not have a pervasive and substantial impact on variation in NAPLAN performance.

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 a 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 (Hill et al. 2014). One finding from the study by Hill et al. (2014) particularly pertinent to the generalist genes hypothesis, is that the set of genes, not a single gene, were 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. 2009, 2010; Knopik and 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 et al. 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 consistent. Although reasonably consistent, 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 et al. 2004; Taylor et al. 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. 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 comprised of. However, the domain-level scores 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 the probability of similar findings across grades is greater than if the participants at each grade level were independent samples.

Conclusion

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 harbor 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.

Compliance with Ethical Standards

Conflict of interest Katrina L. Grasby, William L. Coventry, Brian Byrne, Richard K. Olson, and Sarah E. Medland declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent Informed consent was obtained from all individual participants included in the study, and all procedures were performed in accordance with the ethical standards of the University of New England (HE12-150).

References

Akaike H (1987) Factor analysis and AIC. Psychometrika 52(3):317–332

Alarcón M, Knopik VS, DeFries JC (2000) Covariation of mathematics achievement and general cognitive ability in twins. J Sch Psychol 38(1):63–77

Australian Curriculum Assessment and Reporting Authority (2008) NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2008. ACARA, Sydney

Australian Curriculum Assessment and Reporting Authority (2009) NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2009. ACARA, Sydney

Australian Curriculum Assessment and Reporting Authority (2010) NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2010. ACARA, Sydney

Australian Curriculum Assessment and Reporting Authority (2011) NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2011. ACARA, Sydney

Australian Curriculum Assessment and Reporting Authority (2012) NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2012. ACARA, Sydney

Australian Curriculum Assessment and Reporting Authority (2013) NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2013. ACARA, Sydney

Australian Curriculum Assessment and Reporting Authority (2014) NAPLAN achievement in reading, persuasive writing, language conventions and numeracy: National report for 2014. ACARA, Sydney

Baumert J, Kunter M, Blum W, Brunner M, Voss T, Jordan A, Klusmann U, Krauss S, Neubrand M, Tsai Y-M (2010) Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. Am Educ Res J 47(1):133–180

Beilock SL, Gunderson EA, Ramirez G, Levine SC, Smith EE (2010) Female teachers' math anxiety affects girls' math achievement. Proc Natl Acad Sci USA 107(5):1860–1863

Bleeker MM, Jacobs JE (2004) Achievement in math and science: do mothers' beliefs matter 12 years later? J Educ Psychol 96(1):97–109

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J (2011) OpenMx: an open source extended structural equation modeling framework. Psychometrika 76(2):306–317

Byrne B, Delaland C, Fielding-Barnsley R, Quain P, Samuelsson S, Høien T, Corley R, DeFries J, Wadsworth S, Willcutt E, Olson RK (2002) Longitudinal twin study of early reading development in three countries: preliminary results. Ann Dyslexia 52(1):47–73

Byrne B, Wadsworth S, Corley R, Samuelsson S, Quain P, DeFries JC, Willcutt E, Olson RK (2005) Longitudinal twin study of early literacy development: preschool and kindergarten phases. Sci Stud Read 9(3):219–235

Byrne B, Samuelsson S, Wadsworth S, Hulslander J, Corley R, DeFries J, Quain P, Willcutt E, Olson R (2007) Longitudinal twin study of early literacy development: preschool through Grade 1. Read Writ 20(1):77–102

Byrne B, Coventry WL, Olson RK, Wadsworth SJ, Samuelsson S, Petrill SA, Willcutt EG, Corley R (2010) "Teacher effects" in early literacy development: evidence from a study of twins. J Educ Psychol 102(1):32–42

Davis OSP, Kovas Y, Harlaar N, Busfield P, McMillan A, Frances J, Petrill SA, Dale PS, Plomin R (2008) Generalist genes and the internet generation: etiology of learning abilities by web testing at age 10. Genes Brain Behav 7(4):455–462

Davis OSP, Haworth CMA, Plomin R (2009) Learning abilities and disabilities: generalist genes in early adolescence. Cogn Neuropsychiatry 14(4–5):312–331

de Zeeuw EL, de Geus EJC, Boomsma DI (2015) Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. Trends Neurosci Educ 4(3):69–76

Friend A, DeFries JC, Olson RK (2008) Parental education moderates genetic influences on reading disability. Psychol Sci 19(11):1124–1130

Grasby KL, Byrne B, Olson RK (2015) Validity of large-scale reading tests: a phenotypic and behaviour–genetic analysis. Aust J Educ 59(1):5–21

Grasby KL, Coventry WL (2016) Longitudinal stability and growth in literacy and numeracy in Australian school children. Behav Genet. doi:10.1007/s10519-016-9796-0

Harlaar N, Dale PS, Plomin R (2005a) Telephone testing and teacher assessment of reading skills in 7-year-olds: II. Strong genetic overlap. Read Writ 18(5):401–423

Harlaar N, Spinath FM, Dale PS, Plomin R (2005b) Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins. J Child Psychol Psychiatry 46(4):373–384

Harlaar N, Dale PS, Plomin R (2007) From learning to read to reading to learn: substantial and stable genetic influence. Child Dev 78(1):116–131

Harlaar N, Kovas Y, Dale PS, Petrill SA, Plomin R (2012) Mathematics is differentially related to reading comprehension and word decoding: evidence from a genetically sensitive design. J Educ Psychol 104(3):622–635

Hart SA, Petrill SA, Thompson LA, Plomin R (2009) The ABCs of math: a genetic analysis of mathematics and its links with reading ability and general cognitive ability. J Educ Psychol 101(2):388

Hart SA, Petrill SA, Thompson LA (2010) A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. Learn Individ Differ 20(2):63–69

Hart SA, Soden B, Johnson W, Schatschneider C, Taylor J (2013) Expanding the environment: gene × school-level SES interaction on reading comprehension. J Child Psychol Psychiatry 54(10):1047–1055

Hawke JL, Olson RK, Willcut EG, Wadsworth SJ, DeFries JC (2009) Gender ratios for reading difficulties. Dyslexia 15(3):239–242

Haworth CMA, Kovas Y, Petrill SA, Plomin R (2007) Developmental origins of low mathematics performance and normal variation in twins from 7 to 9 years. Twin Res Hum Genet 10(1):106–117

Haworth CMA, Kovas Y, Harlaar N, Hayiou-Thomas ME, Petrill SA, Dale PS, 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. J Child Psychol Psychiatry 50(10):1318–1325

Hedges LV, Nowell A (1995) Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science 269(5220):41–45

Hill HC, Blunk ML, Charalambous CY, Lewis JM, Phelps GC, Sleep L, Ball DL (2008) Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study. Cogn Instr 26(4):430–511

Hill WD, Davies G, van de Lagemaat LN, Christoforou A, Marioni RE, Fernandes CPD, Liewald DC, Croning MDR, Payton A, Craig LCA, Whalley LJ, Horan M, Ollier W, Hansell NK, Wright MJ, Martin NG, Montgomery GW, Steen VM, Le Hellard S, Espeseth T, Lundervold AJ, Reinvang I, Starr JM, Pendleton N, Grant SGN, Bates TC, Deary IJ (2014) Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. Transl Psychiatry 4:e341

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

Hyde JS (2005) The gender similarities hypothesis. Am Psychol 60(6):581–592

Hyde JS (2007) New directions in the study of gender similarities and differences. Curr Dir Psychol Sci 16(5):259–263

Jackson RW, Snieder H, Davis H, Treiber FA (2001) Determination of twin zygosity: a comparison of DNA with various questionnaire indices. Twin Res 4(1):12–18

Knopik VS, DeFries JC (1999) Etiology of covariation between reading and mathematics performance: a twin study. Twin Res Hum Genet 2(3):226–234

Knopik VS, Alarcón M, DeFries JC (1998) Common and specific gender influences on individual differences in reading performance: a twin study. Pers Individ Differ 25(2):269–277

Kovas Y, Plomin R (2007) Learning abilities and disabilities: generalist genes, specialist environments. Curr Dir Psychol Sci 16(5):284–288

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

Kovas Y, Haworth CMA, Petrill SA, Plomin R (2007) Mathematical ability of 10-year-old boys and girls. J Learn Disabil 40(6):554–567

Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault J-B, Asbury K, Harlaar N, Kovas Y, Dale PS, Plomin R (2014) The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. Proc Natl Acad Sci USA 111(42):15273–15278

Limbrick L, Wheldall K, Madelaine A (2010) Estimating gender ratios of poor reading using large-scale assessments. Aust J Educ 54(2):190–222

Lykken DT, Bouchard TJ, McGue M, Tellegen A (1990) The Minnesota twin family registry: some initial findings. Acta Genet Med Gemellol 39(1):35–70

Machin S, Pekkarinen T (2008) Global sex differences in test score variability. Science 322(5906):1331–1332

Neale MC, Maes HH (2004) Methodology for the genetic studies of twins and families Dordrecht. Kluwer, The Netherlands

Neale MC, Røysamb E, Jacobson K (2006) Multivariate genetic analysis of sex limitation and G × E interaction. Twin Res Hum Genet 9(4):481–489

Nye B, Konstantopoulos S, Hedges LV (2004) How large are teacher effects? Educ Eval Policy Anal 26(3):237–257

Oliver B, Harlaar N, Hayiou Thomas ME, Kovas Y, Walker SO, Petrill SA, Spinath FM, Dale PS, Plomin R (2004) A twin study of teacher-reported mathematics performance and low performance in 7-year-olds. J Educ Psychol 96(3):504–517

Oliver B, Dale PS, Plomin R (2007) Writing and reading skills as assessed by teachers in 7-year olds: a behavioral genetic approach. Cogn Dev 22(1):77–95

Olson RK, Hulslander J, Christopher M, Keenan JM, Wadsworth SJ, Willcutt EG, Pennington BF, DeFries JC (2013) Genetic and environmental influences on writing and their relations to language and reading. Ann Dyslexia 63(1):25–43

Petrill SA, Thompson LA (1994) The effect of gender upon heritability and common environmental estimates in measures of scholastic achievement. Pers Individ Differ 16(4):631–640

Petrill SA, Deater-Deckard K, Thompson LA, Schatschneider C, DeThorne L, Vandenbergh DJ (2007) Longitudinal genetic analysis of early reading: the Western Reserve Reading Project. Read Writ 20(1–2):127–146

Petrill SA, Logan J, Hart SA, Vincent P, Thompson LA, 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. J Learn Disabil 45(4):371

Plomin R, Kovas Y (2005) Generalist genes and learning disabilities. Psychol Bull 131(4):592–617

Plomin R, Kovas Y, Haworth CMA (2007) Generalist genes: genetic links between brain, mind, and education. Mind Brain Educ 1(1):11–19

Reynolds CA, Hewitt JK, Erickson MT, Silberg JL, Rutter M, Simonoff E, Meyer J, Eaves LJ (1996) The genetics of children's oral reading performance. J Child Psychol Psychiatry 37(4):425–434

Samuelsson S, Byrne B, Olson RK, Hulslander J, Wadsworth S, Corley R, Willcutt EG, DeFries JC (2008) Response to early literacy instruction in the United States, Australia, and Scandinavia: a behavioral-genetic analysis. Learn Individ Differ 18(3):289–295

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

Shakeshaft NG, Trzaskowski M, McMillan A, Rimfeld K, Krapohl E, Haworth CM, Dale PS, 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

Spinath FM, Spinath B, Plomin R (2008) The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. Eur J Personality 22(3): 211–229

Taylor J, Schatschneider C (2010) Genetic influence on literacy constructs in kindergarten and first grade: evidence from a diverse twin sample. Behav Genet 40(5):591–602

Taylor J, Roehrig AD, Hensler BS, Connor CM, Schatschneider C (2010) Teacher quality moderates the genetic effects on early reading. Science 328(5977):512–514

Trouton A, Spinath FM, Plomin R (2002) Twins Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. Twin Res Hum Genet 5(05):444–448

Wadsworth SJ, DeFries JC (2005) Genetic etiology of reading difficulties in boys and girls. Twin Res Hum Genet 8(6):594–601