

# Longitudinal Stability and Growth in Literacy and Numeracy in Australian School Students

Katrina L. Grasby<sup>1,2,3</sup> · William L. Coventry<sup>1</sup>

Received: 15 November 2015 / Accepted: 27 May 2016 / Published online: 17 June 2016  
© Springer Science+Business Media New York 2016

**Abstract** We explored the genetic and environmental influence on both stability and growth in literacy and numeracy in 1927 Australian twin pairs from Grade 3 to Grade 9. Participants were tested on reading, spelling, grammar and punctuation, writing, and numeracy. In each domain, performance across time was highly correlated and this stability in performance was primary due to genes. Key findings on growth showed that reading followed a compensatory growth pattern that was largely due to genetic effects, while variation in growth in the other literacy domains was predominantly due to environmental influences. Genes and the shared environment influenced growth in numeracy for girls, while for boys it was influenced by the shared and unique environment. These results suggest that individual differences in growth of reading are primarily due to a genetically influenced developmental delay in the acquisition of necessary skills, while environmental influences, perhaps including different schools or teachers, are more important for the other domains.

**Keywords** Reading · Numeracy · Academic achievement · Twins · Growth · Genetic influences

---

Edited by Yoon-Mi Hur.

---

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

<sup>1</sup> Discipline of Psychology, School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale, NSW, Australia

<sup>2</sup> Australian Research Council Centre of Excellence for Cognition and Its Disorders, Sydney, NSW, Australia

<sup>3</sup> Queensland Institute of Medical Research, 300 Herston Rd, Herston, Brisbane, QLD, Australia

## Introduction

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 et al. 2010), classroom and teacher (Byrne et al. 2010; Nye et al. 2004), study (Cooper et al. 2006), confidence (Stankov et al. 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 et al. 2015; Olson et al. 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, growth trajectories have been found to differ and may result in either increasing or decreasing achievement gaps (Pfost et al. 2014; Shin et al. 2013). Variation in growth trajectories have also been linked with SES, cognitive resources, and behavior (Baumert et al. 2012; Morgan et al. 2011). However, 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 et al. 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. (2016) 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 et al. 2007; Wadsworth et al. 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 Project (WRRP; Logan et al. 2013; Petrill et al. 2010), The International Longitudinal Twin Study (ILTS; Christopher et al. 2013a, b), 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 WRRP (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, b). For reading comprehension, only the shared environment substantially and significantly influenced growth among participants in the WRRP (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. WRRP, 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. 2005, 2007, 2009). Given the substantial contribution of genes to stability in performance in younger Australian children and in students from the USA and the UK we expect the longitudinal phenotypic correlations in reading and spelling from Grade 3 to Grade 9 will mostly be due to genes. 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, .58 for grammar and punctuation and .45 for writing (when averaged across all four grades; Grasby et al. 2016). Longitudinal genetic influence on math 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 the genetic and environmental influence on growth are similar for females and males. Some phenotypic studies have found differences between reading and numeracy in growth. A compensatory growth pattern has been found in reading, such that poorer readers catch up a little to strong readers, while a more stable or even fan growth pattern has been found in numeracy, such that those who initially perform better also improve faster than those who initially performed poorer (Baumert et al. 2012; Morgan et al. 2011; Shin et al. 2013). Thus, it is feasible that the influence of genes and the environment might differ across academic domain. In contrast with the studies from the ILTS, WRRP 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 et al. 2008). Previous biometric growth studies that show substantial influence from the shared environment might do so due to variation in 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 environmental effects in a twin study, will contribute to growth beyond the initial rapid development in reading. In summary for growth, we will assess the extent of genetic and environmental influences on growth, if genetic and environmental influences are similar for females and males, and if compensatory or fan patterns of growth are due to the influence of genes, the environment, or both. In determining the extent that variation in growth is attributable to environmental influence, we will thus evaluate the appropriateness of attributing variation in growth to teachers and schools.

## Methods

### Participants

Twins and triplets born from 1993 to 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 et al. 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). Table 1 details the number of twin pairs by gender and zygosity for each test. 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. Thus, most of the longitudinal missing data is missing by design. The number of twin pairs at each grade with longitudinal numbers specified in parenthesis was: 1178 in Grade 3; 1101 in Grade 5 (773 in Grade 3); 990 in Grade 7 (717 in Grade 5 and 413 in Grade 3), and 809 in Grade 9 (653 in Grade 7, 408 in Grade 5, and 149 in Grade 3). The average age in Grade 3 at the time of testing was 8.6 years.

### 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 calendar year, test calibration and scaling was based on the Rasch model. Scores for each achievement domain were equated across grades, based on common items in tests from adjacent grades, and

**Table 1** Descriptive statistics, intraclass correlations by sex and zygosity, and heritability estimates for each domain and grade

Variable	Descriptives			Intraclass correlations									
	<i>M</i>	<i>SD</i>	<i>n</i>	MZ <sub>F</sub>		MZ <sub>M</sub>		DZ <sub>F</sub>		DZ <sub>M</sub>		DZ <sub>OS</sub>	
				<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>	<i>R</i>	<i>n</i>	<i>r</i>	<i>n</i>
Reading													
Grade 3	447	85	1175	.74	261	.74	244	.41	181	.42	162	.43	299
Grade 5	526	76	1098	.67	227	.67	220	.54	165	.49	170	.33	279
Grade 7	582	66	988	.70	209	.71	201	.51	162	.47	152	.42	223
Grade 9	634	64	809	.72	194	.72	169	.44	122	.45	119	.46	163
Spelling													
Grade 3	426	74	1177	.77	261	.75	244	.43	180	.47	163	.34	301
Grade 5	507	68	1099	.78	229	.76	221	.44	166	.48	171	.41	279
Grade 7	565	62	988	.78	212	.73	197	.43	161	.29	154	.35	216
Grade 9	610	63	808	.79	194	.72	165	.40	123	.27	120	.26	162
GP													
Grade 3	457	88	1177	.70	259	.64	241	.44	179	.48	161	.42	295
Grade 5	532	83	1099	.71	230	.67	221	.46	165	.46	168	.33	278
Grade 7	579	74	989	.63	212	.69	199	.41	160	.37	152	.41	220
Grade 9	615	71	806	.69	193	.62	165	.35	124	.43	122	.40	166
Writing													
Grade 3	436	58	1177	.49	261	.47	240	.28	177	.36	161	.29	297
Grade 5	508	62	1098	.60	229	.53	219	.38	164	.38	167	.27	272
Grade 7	546	69	989	.48	210	.50	197	.34	162	.25	151	.19	221
Grade 9	609	76	806	.52	193	.44	164	.18	123	.41	122	.19	166
Numeracy													
Grade 3	424	70	1174	.74	258	.72	238	.52	182	.45	162	.39	297
Grade 5	517	70	1097	.79	228	.72	218	.53	166	.48	170	.42	279
Grade 7	574	71	987	.84	210	.79	199	.47	162	.45	151	.37	222
Grade 9	632	69	802	.79	195	.77	165	.43	120	.43	114	.46	163

Variable	Standardised variance estimates <sup>a</sup>			Total variance
	A	C	E	
Reading				
Grade 3	.67*	.08*	.24*	7217
Grade 5	.58*	.11*	.30*	5734
Grade 7	.55*	.18*	.27*	4416
Grade 9	.56*	.19*	.25*	4294
Spelling				
Grade 3	.73*	.05	.22*	5477
Grade 5	.77*	.03	.21*	4574
Grade 7	.76*	.02	.21*	4100
Grade 9	.76*	.01	.22*	4314
GP				
Grade 3	.51*	.19*	.30*	7748
Grade 5	.66*	.05	.29*	6697
Grade 7	.57*	.11*	.32*	5571
Grade 9	.52*	.16*	.32*	5195
Writing				
Grade 3	.42*	.10*	.48*	3277
Grade 5	.47*	.11*	.43*	3719

**Table 1** continued

Variable	Standardised variance estimates <sup>a</sup>						Total variance	
	A		C		E			
Grade 7	.41*		.08		.52*		4476	
Grade 9	.39*		.10*		.52*		5600	
Numeracy	F	M	F	M	F	M	F	M
Grade 3	.63*	.64*	.11	.13*	.26*	.23*	4387	5628
Grade 5	.54*	.71*	.21*	.06	.25*	.23*	4140	5229
Grade 7	.79*	.73*	.04	.09	.17*	.17*	4381	5753
Grade 9	.59*	.61*	.20	.18*	.21*	.21*	4501	5135

All correlations were significant at  $p < .05$

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

\* Significance calculated from 95 % confidence intervals not including zero

<sup>a</sup> Estimates were obtained from the correlated factors model, estimates for males and females were significantly different for numeracy

equated with historical years, based on administering a subsample of students equating tests and the NAPLAN tests. Raw scores were transformed into a score on a common scale from 0 to 1000. This scaled score spans all years of the test and was designed to measure growth within cohorts (i.e. calendar year) and to compare across cohorts. Technical information and test administration details were obtained from the Australian Curriculum Assessment and Reporting Authority (2015a, b; R. Randall, personal communication, July 10, 2013). Example test papers and writing prompts are available at [www.nap.edu.au](http://www.nap.edu.au).

#### Reading comprehension

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 min, and for Grades 7 and 9 there were 45–50 items to be completed in 65 min. 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 min 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 min 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 to 2010 the maximum score was 47, from 2011 to 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 min. 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 min. A Cronbach's alpha of .84 (or above) for each test in each year indicates a high internal reliability.

### Procedure

Along with consent forms, parents completed a questionnaire detailing the school attended at the time of each NAPLAN test, the home environment, child medical history, and zygosity. 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 (comprised 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

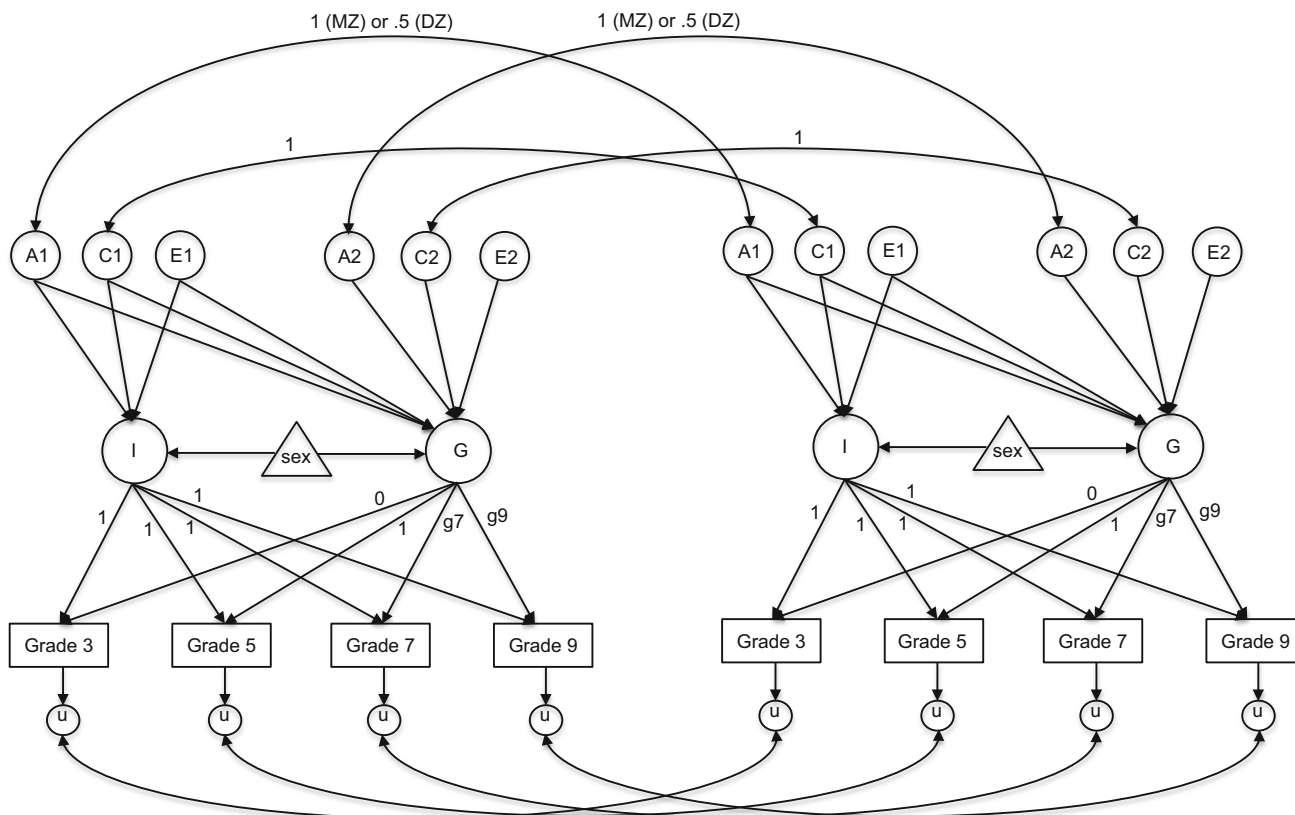
Phenotypic growth curves were estimated with the R package lavaan (Rosseel 2012). Biometric models were

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 to covary with mean performance, and sex has been shown to moderate heritability in some of the domains and grades (details in Grasby et al. 2016). 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 et al. 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 log 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 (Rijsdijk and Sham 2002). 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.

A phenotypic growth curve was initially estimated to determine the shape of growth for each domain. All participants were used in estimation of the phenotypic growth curves; non-independence between twins in a pair will bias standard errors but not parameter estimates, which are of interest in these phenotypic analyses. Growth was centered at Grade 3; growth to Grade 5 was fixed to 1, and growth to Grade 7 and to Grade 9 were estimated (parameters  $g_7$  and  $g_9$  respectively in Fig. 1). The estimates from the phenotypic models were used to specify the shape of growth for the biometric growth curve models.

The biometric latent quadratic growth curve model is depicted in Fig. 1. The model estimates variation in performance at an intercept (I) and for growth (G) (McArdle et al. 1998 discusses in detail). Variation in the intercept and growth and their covariance are decomposed into additive (A), shared environmental (C), and unique environmental (E) sources. In line with classic twin design methodology (see Plomin et al. 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 deviations of scores from that predicted by the growth curve are modelled as time-point specific error variance (u). 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



**Fig. 1** Biometric latent growth model. The left side of the figure represents Twin 1 and the right side represents Twin 2. I = intercept; G = growth; g7 = Grade 7 growth parameter from the phenotypic model; g9 = Grade 9 growth parameter from the phenotypic model. Each path from latent additive genetic (A1, A2), shared environment

(C1, C2), and unique environment (E1, E2) 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

between twins in a pair (Christopher et al. 2013a, b). Moreover, Christopher et al. (2013b) showed that models with correlated errors resulted in a better fit than models with uncorrelated errors. 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 allowed the errors between twin pairs to correlate. 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 log likelihood ratio test.

## Results

### Descriptives and longitudinal correlations

Descriptive statistics, intraclass correlations, heritability, shared and unique environmental estimates and total variance for each grade and subject are reported in Table 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 environmental estimates were small, and the unique environmental estimates (including measurement error) were greatest in the writing domain. Total variance decreased with increasing grade in all domains except for writing, which increased in total variance with increasing grade. Phenotypic correlations

were high among most grades for each domain, though the correlations across grades for writing were more moderate (see Table 2). Correlations were equally high for grades further apart in time as for adjacent grades.

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 for numeracy, correlations could be equated across sex,  $\chi^2(50) = 34.99$ ,  $p = .947$ , but not path loadings as well,  $\chi^2(62) = 87.84$ ,  $p = .017$ . For numeracy, 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 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 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 these domains. 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 environmental estimates at each grade level may have been due to measurement error.

### Biometric growth curve

The shape estimates from the phenotypic growth curves were rounded to one decimal place and tested for fit against the observed data. Shape estimates and fit indices are reported in Table 4; fit was very good in each domain, where an RMSEA below .01, .05, and .08 indicate excellent, good, and acceptable fit, respectively (MacCallum et al. 1996).

For the biometric growth curve models 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 5 indicate that girls scored higher on the literacy domains and boys scored higher on numeracy. Sex effects on growth indicated that growth in performance for boys was significantly steeper than for girls in reading, spelling, and numeracy. In combination these sex effects indicate that for reading and spelling, 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 grammar and punctuation and writing 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.

**Table 2** Phenotypic correlations among the grades in each domain

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

All correlations were significant at  $p < .05$



**Table 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
<b>Reading</b>									
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]
<b>Spelling</b>									
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]
<b>GP</b>									
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]
<b>Writing</b>									
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]
<b>Numeracy</b>									
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]

95 % confidence intervals are in brackets

**Table 4** Phenotypic growth estimates for Grade 7 and Grade 9 and fit indices for each domain

Variable	Growth estimates		Fit indices	
	g7	g9	CFI	RMSEA
Reading	1.7	2.4	.998	.015
Spelling	1.7	2.3	.999	.016
GP	1.6	2.1	.999	.008
Writing	1.5	2.4	1.00	.000
Numeracy	1.6	2.2	.990	.039

g7 and g9 are the growth estimates for Grade 7 and Grade 9 respectively, where Grade 3 is the intercept and growth to Grade 5 is fixed at unity

Variation in the intercept and growth are reported in Table 5. All domains are scored on a scale of 0–1000 with similar means across the domains at each grade. As such, writing had a noticeably constricted variation in performance on the intercept when compared to the other domains; this was also the only domain to increase in variation as the grades progressed. For spelling, writing, and numeracy, where sex could not be dropped from the covariance structure, boys had greater variance than girls on both intercept and growth. The variation in growth in numeracy was narrower than the literacy domains, indicating that growth in NAPLAN numeracy was quite similar particularly among girls.

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 not be equated in any domain 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 and growth are reported in Table 6 for each

domain, where A1, C1, and E1 represent the intercept and A2, C2, and E2 represent growth. On the off diagonal are reported the standardised A, C, and E covariances between the intercept and growth. On the intercept, genetic influences were significant for each domain, standardised variance estimates ranged from .56 for numeracy performance among girls to .89 for spelling performance among girls. Shared environmental influences were modest and significant for all domains except for spelling among girls, estimates ranged from .06 for numeracy among boys to .41 for numeracy among girls. The unique environmental influence on the intercept for each domain was generally small and significant, estimates ranging from .04 for numeracy among girls to .20 for writing among boys.

For reading, genetic influences on growth were substantial and significant, with a modest influence from the unique environment. Much of the genetic influence on growth was shared with that of the intercept. The genetic correlation ( $r_A$ ) between the intercept and growth was  $-.76$  (calculated from the genetic covariance divided by the square root of the product of the genetic variance of the intercept and growth [ $-.59/\sqrt{.80 * .74} = -.76$ ]). About half of the unique environmental influences on growth were shared with the intercept (unique environmental correlation,  $r_E = -.54$ ). Negative covariation between intercept and growth indicated that (on average) individuals who scored higher in Grade 3 had a slower rate of growth than individuals who scored lower in Grade 3. The genetic and unique environmental correlations between intercept and growth indicate many of the genes and about half of the unique environmental factors that influence individuals to score higher in Grade 3 also contribute to less growth.

For spelling, variation in growth for girls was significantly influenced by genes, the shared environment, and the unique environment. The genetic influence was modest, while the shared environment was the most substantial influence. This shared environmental influence on growth was independent of any factors influencing performance in Grade 3, as there was no influence from the shared

**Table 5** Estimated mean, variance and sex effect on intercept and growth, and error correlations for each domain

Variable	Intercept			Growth			$r_U$	
	Mean	Sex	Variance	Mean	Sex	Variance	MZ	DZ
Reading	455	–12.9*	5574	75	4.4*	301	.14*	.04
Spelling	435	–16.4*	4144 F 5066 M	78	3.2*	136 F 252 M	.28*	.09*
GP	466	–19.3*	5310	74	3.4	249	.24*	.10*
Writing	446	–21.0*	1702 F 2269 M	73	–.9	217 F 393 M	.13*	.01
Numeracy	419	12.7*	2980 F 4263 M	91	3.0*	60 F 198 M	.29*	.19*

Sex specifies the adjustment to the mean intercept or shape for males;  $r_U$  = error correlation between twins; F = females; M = males

\* Significance calculated from 95 % confidence intervals not including zero

**Table 6** Estimates of standardised A, C, E variance and covariance components of the intercept (A1, C1, E1) and growth (A2, C2, E2) for each domain

Domain	Variable	A1	A2	C1	C2	E1	E2
Reading							
	Intercept	.80 [.77, .85]*		.11 [.07, .16]*		.09 [.05, .11]*	
	Growth	-.59 [-.66, -.48]*	.74 [.48, .99]*	.03 [-.04, .08]	.01 [0, .08]	-.08 [-.08, -.01]*	.25 [.23, .31]*
Spelling							
Female	Intercept	.89 [.80, .91]*		0 [0, .01]		.11 [.08, .14]*	
	Growth	-.38 [-.52, -.20]*	.19 [.15, .20]*	-.03 [-.16, .05]	.73 [.59, .82]*	-.05 [-.06, -.02]*	.08 [.02, .21]*
Male	Intercept	.65 [.55, .72]*		.21 [.20, .25]*		.13 [.10, .15]*	
	Growth	-.15 [-.36, .10]	.50 [0, .92]	-.29 [-.51, -.08]*	.41 [.07, .44]*	-.03 [-.12, .04]	.09 [0, .09]
Grammar and punctuation							
	Intercept	.71 [.69, .87]*		.22 [.20, .26]*		.08 [.05, .10]*	
	Growth	-.26 [-.54, -.02]*	.10 [0, .62]	-.25 [-.50, -.02]*	.67 [.19, .94]*	-.05 [-.05, -.03]*	.24 [.18, .52]*
Writing							
Female	Intercept	.58 [.27, .59]*		.24 [.01, .53]*		.19 [.17, .31]*	
	Growth	.23 [.22, .61]*	.094 [.091, .64]*	.05 [.03, .20]*	.12 [0, .63]	-.35 [-.41, -.28]*	.78 [.02, .94]*
Male	Intercept	.58 [.25, .83]*		.23 [.03, .53]*		.20 [.06, .32]*	
	Growth	.04 [-.22, .28]	0 [0, .52]	-.17 [-.36, -.03]*	.56 [.04, .91]*	-.12 [-.32, .09]	.44 [.17, .78]*
Numeracy							
Female	Intercept	.56 [.39, .63]*		.41 [.18, .53]*		.04 [.03, .05]*	
	Growth	.52 [.23, .66]*	.49 [.14, .85]*	-.35 [-.44, -.20]*	.39 [.14, .71]*	.07 [.05, .09]*	.12 [.00, .19]
Male	Intercept	.86 [.75, .94]*		.06 [.02, .16]*		.08 [.03, .08]*	
	Growth	-.07 [-.28, .13]	.02 [0, .97]	-.04 [-.04, -.01]*	.62 [.03, .99]*	-.02 [-.16, .11]	.36 [.02, .87]*

A = genetic; C = shared environment; E = unique environment; 1 = intercept; 2 = growth; 95 % confidence intervals are in brackets

\* Significance calculated from 95 % confidence intervals not including zero

environment on performance in spelling for girls in Grade 3. For boys, the shared environment was the only significant influence on variation in growth of spelling, and the same factors that influenced growth also influenced spelling in Grade 3 (shared environmental correlation,  $r_C = -.99$ ).

For grammar and punctuation, genetic influences were not significant. The shared environment was the most substantial influence on growth with a more modest influence from the unique environment. There was both overlap and independence of factors influencing growth and performance in Grade 3 ( $r_C = -.67$  and  $r_E = -.33$ ).

For writing, genetic influences on growth for girls were small but significant. The unique environment was the most substantial influence on growth for girls, with most factors that influenced growth also influencing performance in Grade 3 ( $r_E = -.92$ ). For boys, the shared and unique environments influenced growth to a similar extent, with approximately half of the factors influencing growth also influencing performance in Grade 3 ( $r_C = -.47$  and  $r_E = -.42$ ).

For numeracy, genes and the shared environment influenced growth to a similar extent among girls, and the factors were mostly the same as those that influenced

variation in Grade 3 ( $r_A = 1.0$  and  $r_C = -.89$ ). The positive genetic correlation in numeracy indicates that, unlike reading, the genes that influenced girls to score higher in Grade 3 also influenced girls to have a faster rate of growth. However, no overall fan effect in growth was evident for girls, as this positive correlation between intercept and growth on the genetic factor was counterbalanced by a negative correlation between intercept and growth on shared environmental factors. No genetic influence on growth in numeracy was evident among boys; instead both the shared environment and unique environment were substantial influences. Unlike the other domains, most of the factors influencing growth in numeracy among boys were independent of those influencing performance in Grade 3 ( $r_C = -.18$  and  $r_E = -.09$ ).

## 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. In contrast, the etiology of variation in growth differed by domain. For reading, genetic factors were the strongest influence on growth with some influence from unique environmental factors. For spelling, the shared environment was the strongest influence on growth for both girls and boys, with a smaller influence from genes and the unique environment for girls. For grammar and punctuation, shared environmental factors were again the strongest influence on growth with some influence from unique environmental factors. For writing, unique environmental influences were the strongest influence on growth for girls, while both shared and unique environmental factors were significant for boys. For numeracy, both shared and unique environmental factors were again significant for boys, while genetic and shared environmental factors were substantial influences on growth for girls.

### 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). From grade specific estimates, 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 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 WRRP (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. In contrast to our results and those from the USA, novel genetic influences on later ages have been found for both reading and math 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, WRRP, 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 in 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 most substantial contributor to individual differences in growth in reading, and these were largely same genes that influenced performance at Grade 3. The unique environment had a more modest influence on growth in reading, indicating that the role of specific

environmental impacts on variation in reading growth from Grade 3, such as different teachers, instruction methods, or interests, is much less substantial than that of genes.

The negative covariation between the intercept and growth 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 largely due to genetic factors. In Pfof 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 assessed in these NAPLAN data, is not a constrained reading skill. However, in Pfof et al.'s review reading comprehension was also associated more with compensatory growth than 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 may largely be due to an inherent developmental delay in the mastery of necessary reading skills.

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, which lessened the 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 in Grade 3. Grammar and punctuation also showed compensatory growth, such that the poorer performing students in Grade 3 tended to catch-up a little to the higher performing students. However, the results differed from reading regarding the influences on growth; for grammar and punctuation the environment contributed to variation in growth. Therefore, the relative improvement in poorer performers in the case of grammar and punctuation is due to environmental influences, most of which are shared environmental factors. These would include potential effects from the home environment, teachers, and schools.

For the remaining literacy domains, spelling and writing, girls and boys significantly differed in the relative

contribution of genes and the environment to variation in growth. As with the other literacy skills, girls scored higher than boys in Grade 3, but there were no significant sex differences in average growth, indicating a relatively stable achievement gap between girls and boys in spelling and writing over time. Although there were small genetic influences on growth in both spelling and writing for girls, like grammar and punctuation the environment was the strongest influence. For spelling, the shared environment was the most substantial and significant influence on growth. This was also a compensatory growth pattern, indicating a possible instructional influence on relative improvement in poorer spellers over time. For writing, unlike the other literacy domains, there was no overall compensatory growth for poorer writers. For girls the principal influence was the unique environment, while for boys both the shared and unique environments were important influences on variation in growth of writing. This sex difference in environmental influence is interesting; however, around each of these estimates there are wide confidence intervals. It is possible there is a fundamental difference to the way that girls and boys are responding to their environments, either home or school, which impacts growth of their writing performance. However, given the lack of any other biometric longitudinal analysis in writing, replication would be desirable before refining too much on the influence of unique versus shared experiences on gender differences in growth of writing.

In contrast to the literacy domains, boys scored higher than girls in Grade 3 numeracy and had a higher rate of growth. This resulted in increasing sex differences over time. At initial testing in Grade 3, shared environmental influences were stronger for girls than boys and genetic influences were stronger for boys than girls. Like writing there was no evidence of compensatory growth. For girls both genes and the shared environment fairly evenly influenced variation in growth of numeracy, while for boys it was influenced by the shared and unique environments. Moreover, the environmental influences on growth for boys were predominantly independent from the environmental influences on performance in Grade 3. This independence of environmental influences on growth indicates that the factors influencing boys are either being experienced after Grade 3 or are only relevant to performance after Grade 3. This could include the impact of different teachers or classes over time through these middle years of school.

## Implications

For the most part many of the influences on variation in growth are already present and influencing performance

in Grade 3. In many cases this is part of a compensatory growth pattern where those who were performing higher in Grade 3 are then growing more slowly, essentially the higher performing students are not experiencing the same increase in performance as those who performed poorer in Grade 3. This might reflect a ceiling effect for high performers in these NAPLAN tests, at least in our sample.

The influence of genetic factors on growth for many of the domains suggests a tempering of the claim that growth in NAPLAN performance reflects the value added by the school. This is particularly important as far as reading is concerned, where variation in growth was strongly influenced by genes. 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 environmental effect. Other than reading, the shared environment was a significant influence on growth in NAPLAN performance, so our results are consistent with schools influencing growth. Unfortunately our results are not conclusive on this interpretation, as the shared environment includes everything that influences growth in performance that twins have in common, not least of which is the same family environment.

### Limitations

This inability to tease apart the influence of the home from the school environment, particularly regarding the shared environmental influence on variation in growth of spelling, grammar and punctuation, writing, and numeracy limits the extent that these findings can be interpreted. There was considerable overlap in the environmental influence on growth and on performance in Grade 3, indicating that factors affecting growth are stable for individuals over time or have a long-term influence. However, with Grade 3 as the initial testing time it is not possible to assess if any long-term influences on growth are from family factors, perhaps even pre-dating formal schooling, or if they are from the early years of school. Even where there is no substantial overlap in environmental factors between growth and initial performance, as with numeracy among boys, it is not possible to determine from these models what the specific factors are that influence growth and whether they are educational in origin or not.

### 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. Many of 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. For the other literacy domains of spelling, grammar and punctuation, and writing environmental factors were the principal influences on growth, and in the case of spelling and grammar and punctuation these environmental influences contributed to a lessening of the achievement gap over time. In contrast to the literacy domains, boys outperformed girls at initial testing in numeracy and the achievement gap increased over time. For girls, genetic and shared environmental factors influenced variation in growth of numeracy and these factors were the same as those that influenced initial performance. Whereas for boys, shared environmental factors influenced variation in growth and these factors were essentially different to those that influenced initial performance.

**Acknowledgments** This research was supported by an Australian Research Council Grant (DP120102414). The Australian Twin Registry is supported by an enabling Grant (628911) from the National Health and Medical Research Council. We thank the Australian Twin Registry, and all of the twins, triplets and parents involved.

### Compliance with ethical standards

**Conflict of interest** Katrina L. Grasby and William L. Coventry declare that they have no conflict of interest.

**Ethical standard** All procedures were performed in accordance with the ethical standards of the University of New England (HE12-150).

**Informed consent** Informed consent was obtained from all individual participants included in the study.

### References

- Australian Curriculum Assessment and Reporting Authority (2015a) National assessment program—literacy and numeracy 2014: technical report. ACARA, Sydney
- Australian Curriculum Assessment and Reporting Authority (2015b) National protocols for test administration. ACARA, Sydney

- 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 Dev* 83(4):1347–1367
- Bejemann R, Willcutt E, Olson R, Keenan J, DeFries J, Wadsworth S (2008) Word reading and reading comprehension: stability, overlap and independence. *Read Writ* 21(5):539–558
- Bloom HS, Hill CJ, Black AR, Lipsey MW (2008) Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *J Res Educ Eff* 1(4):289–328
- 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, 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, Samuelsson S, Corley R, Willcutt EG, Wadsworth S, DeFries JC (2009) Genetic and environmental influences on aspects of literacy and language in early childhood: continuity and change from preschool to Grade 2. *J Neurolinguist* 22(3):219–236
- 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
- Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington B, DeFries JC, Wadsworth SJ, Willcutt E, Olson RK (2013a) The genetic and environmental etiologies of individual differences in early reading growth in Australia, the United States, and Scandinavia. *J Exp Child Psychol* 115(3):453–467
- Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington B, DeFries JC, Wadsworth SJ, Willcutt EG, Olson RK (2013b) Modeling the etiology of individual differences in early reading development: evidence for strong genetic influences. *Sci Stud Read* 17:350–368
- Cooper H, Robinson JC, Patall EA (2006) Does homework improve academic achievement? A synthesis of research, 1987–2003. *Rev Educ Res* 76(1):1–62
- 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
- Grasby KL, Coventry WL, Byrne B, Olson RK, Medland SE (2016) Genetic and environmental influences on literacy and numeracy performance in Australian school children in Grades 3, 5, 7, and 9. *Behav Genet*. doi:10.1007/s10519-016-9797-z
- 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
- Hart SA, Logan JAR, 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. *Dev Psychol* 49:1971–1981
- Hattie J (2008) *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Routledge, London
- Kovas Y, Haworth CMA, Dale PS, Plomin R, Weinberg RA, Thomson JM, Fischer KW (2007) The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monogr Soc Res Child Dev* 72(3):i-156
- Logan JAR, Hart SA, Cutting L, Deater-Deckard K, Schatschneider C, Petrill S (2013) Reading development in young children: genetic and environmental influences. *Child Dev* 84(6):2131–2144
- Lykken DT, Bouchard TJ, McGue M, Tellegen A (1990) The Minnesota twin family registry: some initial findings. *Acta Genet Med Gemellol (Roma)* 39(1):35–70
- MacCallum RC, Browne MW, Sugawara HM (1996) Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods* 1(2):130–149
- Masters G, Rowley G, Ainley J, Khoo S (2008) Reporting and comparing school performances. [http://research.acer.edu.au/ar\\_misc/8](http://research.acer.edu.au/ar_misc/8)
- McArdle JJ, Prescott CA, Hamagami F, Horn JL (1998) A contemporary method for developmental-genetic analyses of age changes in intellectual abilities. *Dev Neuropsychol* 14(1):69–114
- Morgan PL, Farkas G, Wu Q (2011) Kindergarten children’s growth trajectories in reading and mathematics: who falls increasingly behind? *J Learn Disabil* 44(5):472–488
- 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 An* 26(3):237–257
- Olson RK, Keenan JM, Byrne B, Samuelsson S, Coventry WL, Corley R, Wadsworth SJ, Willcutt EG, DeFries JC, Pennington BF, Hulslander J (2011) Genetic and environmental influences on vocabulary and reading development. *Sci Stud Read* 15(1):26–46
- Olson RK, Keenan JM, Byrne B, Samuelsson S (2014) Why do children differ in their development of reading and related skills? *Sci Stud Read* 18(1):38–54
- Paris SG (2005) Reinterpreting the development of reading skills. *Read Res Quart* 40(2):184–202
- Petrill SA, Deater-Deckard K, Thompson LA, Schatschneider C, DeThorne L, Vandenberg DJ (2007) Longitudinal genetic analysis of early reading: the Western Reserve Reading Project. *Read Writ* 20(1–2):127–146
- Petrill SA, Hart SA, Harlaar N, Logan J, Justice LM, Schatschneider C, Thompson LA, DeThorne LS, Deater-Deckard K, Cutting L (2010) Genetic and environmental influences on the growth of early reading skills. *J Child Psychol Psych* 51(6):660–667
- 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. *Rev Educ Res* 84(2):203–244
- Plomin R, DeFries JC, Knopik VS, Neiderhiser JM (2013) *Behavioral genetics*. Worth, New York
- Rijsdijk FV, Sham PC (2002) Analytic approaches to twin data using structural equation models. *Brief Bioinform* 3(2):119–133
- Rosseel Y (2012) Lavan: an R package for structural equation modeling. *J Stat Softw* 48(2):1–36
- 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. Canberra: Senate Printing Unit. [http://www.aph.gov.au/Parliamentary\\_Business/Committees/Senate/Education\\_and\\_Employment/Naplan13/Report/index](http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Education_and_Employment/Naplan13/Report/index)
- Shin T, Davison ML, Long JD, Chan C-K, Heistad D (2013) Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learn Individ Differ* 23:92–100

- Sirin SR (2005) Socioeconomic status and academic achievement: a meta-analytic review of research. *Rev Educ Res* 75(3):417–453
- 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 Indic Res* 3(1):85–104
- Soden B, Christopher ME, Hulslander J, Olson RK, Cutting L, Keenan JM, Thompson LA, Wadsworth SJ, Willcutt EG, Petrill SA (2015) Longitudinal stability in reading comprehension is largely heritable from Grades 1 to 6. *PLoS One* 10(1):e0113807
- Stankov L, Morony S, Lee YP (2014) Confidence: the best non-cognitive predictor of academic achievement? *J Educ Psychol* 34(1):9–28
- Wadsworth SJ, Corley RP, Hewitt JK, DeFries JC (2001) Stability of genetic and environmental influences on reading performance at 7, 12, and 16 years of age in the Colorado Adoption Project. *Behav Genet* 31(4):353–359