

University of Wollongong

**Research Online**

---

Faculty of Arts, Social Sciences and Humanities  
- Papers

Faculty of Arts, Social Sciences & Humanities

---

2021

**Relative age effects on academic achievement in the first ten years of formal schooling: A nationally representative longitudinal prospective study**

Myrto F. Mavilidi

Herbert W. Marsh

Kate M. Xu

Phillip D. Parker

Pauline W. Jansen

*See next page for additional authors*

Follow this and additional works at: <https://ro.uow.edu.au/asshpapers>

---

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Relative age effects on academic achievement in the first ten years of formal schooling: A nationally representative longitudinal prospective study

## Abstract

The effects of school starting age and relative age effects (RAEs) have generated much interest from parents, teachers, policymakers, and educational researchers. Our 10-year longitudinal study is based on a nationally representative ( $N = 4,983$ ) prospective sample from the Longitudinal Study of Australian Children. The primary outcomes are results from the high-stake, Australia-wide National Assessment Program-Literacy and Numeracy tests in Years 3, 5, 7, and 9, controlling for demographic characteristics (gender, socioeconomic status, school type, and childhood cognition measured before the start of kindergarten). We evaluated how direct and mediated RAEs vary over the first 10 years of schooling for these longitudinal data. Results revealed significant RAEs in primary school years for both numeracy and literacy test outcomes. Effects were large in primary school years but declined in secondary school years. Although the direct effects of RAEs declined over time, there continued to be significant indirect effects over the whole 10-year period. RAEs in primary school had enduring effects that were mediated through the effects of earlier achievement. We juxtapose our results with previous RAE research on achievement and a range of other noncognitive outcomes where the RAEs are enduring into adolescence and even adulthood. We position our research within this broader research literature and discuss implications for educational policy, practice, theory, and future research.

## Publication Details

Mavilidi, M. F., Marsh, H. W., Xu, K. M., Parker, P. D., Jansen, P. W., & Paas, F. (2021). Relative age effects on academic achievement in the first ten years of formal schooling: A nationally representative longitudinal prospective study. *Journal of Educational Psychology*.

## Authors

Myrto F. Mavilidi, Herbert W. Marsh, Kate M. Xu, Phillip D. Parker, Pauline W. Jansen, and Fred Paas

---

Relative Age Effects on Academic Achievement in the First Ten Years of Formal Schooling: A Nationally Representative Longitudinal Prospective Study

**Abstract (241 words)**

The effects of school starting age and relative age effects (RAEs) have generated much interest from parents, teachers, policymakers, and educational researchers. Our 10-year longitudinal study is based on a nationally representative ( $N = 4983$ ) prospective sample from the Longitudinal Study of Australian Children (LSAC). The primary outcomes are results from the high-stake, Australia-wide National Assessment Program-Literacy and Numeracy (NAPLAN) tests in Years 3, 5, 7, and 9, controlling for demographic characteristics (gender, socio-economic status, school type, and childhood cognition measured before the start of kindergarten). We evaluated how direct and mediated RAEs vary over the first ten years of schooling for these longitudinal data. Results revealed significant RAEs in primary school years for both numeracy and literacy test outcomes. Effects were large in primary school years but declined in secondary school years. Although the direct effects of RAEs declined over time, there continued to be significant indirect effects over the whole ten-year period. RAEs in primary school had enduring effects that were mediated through the effects of earlier achievement. We juxtapose our results with previous RAE research on achievement and a range of other non-cognitive outcomes where the RAEs are enduring into adolescence and even adulthood. We position our research within this broader research literature and discuss implications for educational policy, practice, theory, and future research.

**Keywords:** *relative age effect; academic achievement; primary and secondary schools; longitudinal structural equation models; social comparison processes*

**Educational Impact and Implications Statement**

Our 10-year longitudinal study showed that children who are younger than their classmates score lower on national achievement tests in primary school (Years 3 and 5). Although these relative age effects declined in secondary school (Years 7 and 9), the indirect relative age effects persist over the first ten years of schooling. Hence, the effect of relative age in early primary school has long-term implications.

Parents, teachers, policymakers, and researchers have become increasingly concerned about the physical, mental, social, and emotional maturity in relation to a child's readiness for starting school and relative age effects (RAEs). Children in most school systems start their schooling at different ages because most schools accept entering students only once a year based on a fixed cut-off date of birth. There is increasing public recognition of RAEs in which relatively older students are potentially advantaged for a range of physical, social, emotional achievement, motivation, and self-belief outcomes that are related to age and maturity (Barnsley et al., 1985; Thompson et al., 1999). In recognition of this phenomenon, in many educational systems, parents may decide to hold back their children for an additional year following normal schooling starting age—a practice known as redshirting that is on the increase (Hanly et al., 2019; Larsen et al., in press; Marsh, 2016; Mergler & Walker, 2017). Also, in some school systems, grade retention of underperforming students is widespread (Marsh, 2016). All these contribute to RAEs in complex ways. Marsh (2016) noted that educational administrators and policymakers apply ad hoc policies without understanding their implications. In our 10-year longitudinal, prospective study, we evaluate RAEs on the development of achievement in the first ten years of formal schooling.

In Australia, like many countries, children's admission in school is based on cut-off dates (i.e., specific dates based on children's month of birth) for allowing students to enroll in either kindergarten or primary school. The typical age of school entrance lies between 4 and 7 years, whereas the most predominant starting age is 5 or 6 years (Bertram & Pascal, 2002). In general, the cut-off dates categorize children into cohorts when they enter formal schooling and create age differences up to one year for children in the same cohort or class (DeCos, 1997; Lawlor et al., 2006; Musch & Grondin, 2001; Peña, 2017; Sprietsma, 2010). Research has shown that the youngest children in a cohort, who are born close to the next cut-off date, are less cognitively, physically, and emotionally developed compared to the oldest children, who are born shortly after the cut-off date, defining the start of their cohort (Musch & Grondin, 2001).

There is an abundance of research evidence demonstrating these RAEs in sports (e.g., baseball; Thompson et al., 1991; dance; Van Rossum, 2006; football; Barnsley et al., 1992; hockey; Barnsley et al., 1985; soccer; Helsen, Van Winckel, & Williams, 2005), as well as in academic context (e.g., Carroll, 1992; Copley et al., 2009b; Dhuey et al., 2019; Sharp, 1995; Tymms, 2006). It is assumed that trainers and teachers falsely associate physical and psychological maturity with actual skill differences in both contexts. In this sense, the effects are driven by social comparison processes that fail to consider relative age and maturity.

Consequently, teachers and trainers tend to show a preference for the older children within a year group, considering them as more skillful and talented, thereby classifying the younger children as less skillful and talented (Hancock et al., 2013). These perceived skill differences are assumed to affect the trainer or teacher's behavior towards the students. They provide more feedback, praise, and instruction to the 'more talented' students (e.g., Solomon et al., 1998). Hence, the 'less-talented' younger entrants experience psychosocial disengagement (e.g., reduced self-esteem, self-efficacy, less active social life; Pellizzari & Billari, 2012; Thompson et al., 2004), lower academic self-concept (Marsh, 2016; Marsh, Pekrun, et al., 2017) as well as difficulties in adapting to the school settings (Gledhill et al., 2002; Martin et al., 2004). In their review, Cobley et al. (2009a) concluded that RAEs in educational settings are pervasive and systematic, implicating maturational and psychological mechanisms.

Noting the importance of academic self-concept to educational outcomes, Marsh (2016) found that RAEs for academic self-concept generalized over nationally representative samples of 15-year-olds from 41 countries. They proposed that these differences are due to social comparison processes in which relatively younger students compare themselves to classmates without considering relative age differences. These effects were not substantially explained or moderated by control variables (e.g., gender, school starting age, repeating grades, home language, immigrant status, SES, achievement, school-average achievement). Marsh concluded that educational policymakers in different countries use diverse strategies concerning school starting age, repeating grades, and acceleration, apparently without fully understanding the implications of these policy practices in relation to academic self-concept, motivation, and a range of affective variables that have long-term implications for academic choice and accomplishments. Parker, Marsh et al. (2019) pursued these suggestions in a longitudinal study of RAE's long-term effects. They replicated the RAEs on academic self-concept. However, they also found RAEs on subsequent university entry that could be explained in terms of academic self-concept. These results demonstrate the long-term implications of RAEs and suggest a mechanism through which this occurs based on the social-comparison process and academic self-concept.

In the present study, we used the data from the Longitudinal Study for Australian Children (LSAC; Australian Institute of Family Studies, 2015; Sanson et al., 2002) to explore longitudinal RAEs on literacy and numeracy performance on standardized achievement tests across the first ten years of schooling. To the best of our knowledge, no previous study has linked RAEs with the developmental trajectory of academic performance across the whole span of primary and secondary school using multiple repeated achievement

measures and controlling for children's cognitive abilities measured before the start of school, along with a range of demographic and school characteristics, as well as state-level differences within the same country.

### **Empirical evidence on RAE**

There is considerable empirical evidence in support of RAEs on academic achievement for students from kindergarten (Tymms et al., 2000), primary school (Norbury et al., 2016; Sakic et al., 2013; Sprietsma, 2010), secondary schools (Cobley et al., 2009b; Marsh, 2016), or primary and secondary schools (Peña, 2017). The majority of studies focused on mathematics and language (Bedard & Dhuey, 2006; Cobley et al., 2009b; Marsh, 2016; Norbury et al., 2016; Peña, 2017; Sakic et al., 2013; Sprietsma, 2010; Tymms et al., 2000), but also considered a variety of other learning domains such as science, chemistry, biology, geography, physical education (Cobley et al., 2009b; Peña, 2017; Sakic et al., 2013; Tymms et al., 2000). These learning domains have been assessed by standardized assessments (Bedard & Dhuey, 2006; Marsh, 2016; Peña, 2017; Sprietsma, 2010), final school marks by teachers (Cobley et al., 2009b), or both (Sakic et al., 2013). Finally, research on RAEs has predominantly used cross-sectional observational designs (Cobley et al., 2009b; Norbury et al., 2016; Peña, 2017; Sakic et al., 2013; Sprietsma, 2010), with some notable exemptions using data based on more than one grade of achievement assessments (Bedard & Dhuey, 2006; Tymms et al., 2000).

In educational settings, relatively older students achieve more than their younger classmates (Allen & Barnsley, 1993, Borg & Falzon, 1995; Kawaguchi, 2011; Lien et al., 2005). They are also less likely to repeat a grade (Elder et al., 2009; Peña, 2017), are less likely to be diagnosed as having specific learning disabilities (Elder, 2010; Martin et al., 2004), have more positive self-beliefs (Marsh, 2016), and are more likely to be class leaders in high school (Dhuey & Lipscomb, 2010). In a cross-national study of nationally representative samples, Bedard and Dhuey (2006) reported RAEs were significant and consistent across students in Years 4 and 8 students. Based on other data, they found RAEs for participation in pre-university academic programs in Canada and the US and enrollment in leading universities in the US. Following Dhuey and Lipscomb, 2010), they suggested that RAEs in early schooling persist into adulthood. Kawaguchi (2011) also found positive RAEs in Years 4 and 8. The results were consistent for both boys and girls, suggesting that gender did not moderate the RAE. Kawaguchi further emphasized that early RAEs were maintained, leading to higher levels of subsequent educational attainment.

Elder et al. (2009) found that socioeconomic status (SES) moderated RAEs. Thus, being one year older when starting kindergarten resulted in a 10.65 percentile advantage for children from the lowest SES quartile, but a 23.66 percentile advantage in the highest quartile. They found that these effects generalize over primary school years. However, although the effect sizes declined with age, the effect was relatively more long-lasting for the student from high-SES families. Elder et al. suggested that older children's added skills drive the advantages associated with being relatively older. These are likely to be more substantial for students from high-SES families.

Norbury et al. (2016) assessed 7267 UK students (4.9 - 5.1 years) from 161 primary schools on a range of behavioral and academic measures. Children enrolled in school in September of the academic year that they turned five, and were assessed at the end of the year. Control variables included gender, language spoken at home, SES, medical/clinical diagnosis, and special education needs. Teachers used standardized assessments to identify language impairment, screening of social, emotional, and behavioral functioning, and academic progress. Results revealed that, even when the youngest children were not significantly disadvantaged before the school entry, they showed more behavior problems and had lower academic progress assessments than the older children in the cohort.

Peña (2017) examined the RAE in test scores of low-income Mexican students for grades 3 to 9 ( $n = 162,186$  students in the 2009 data and 161,295 students in the 2012 data). Peña compared students within the same cohort on performance on national standardized tests. These tests were administered at the end of every academic year (i.e., Spanish, math, and science or ethics) and recorded across seven continuous grades (3-9). Individual test score data were available for students in 2009 and 2012. The results showed that older students had a sizable advantage over their younger classmates in Spanish and maths achievement. In complementary analyses based on national (Mexican) data, Peña showed that adults who were older in their class attained higher education levels and higher earnings, and married more educated spouses.

Sakic et al. (2013) examined differences in a crude dichotomous measure of starting age for two cohorts of students completing national standardized in Years 4 (44,479) and 8 (43,338). They assessed achievement with school grades and standardized achievement tests. Analyses were multiple two-group comparisons conducted separately for each outcome at each grade level. In Year 4, there were small but largely significant differences in favor of older students that generalized reasonably well across the school grades and test scores in four domains. However, in Year 8, the differences were mostly non-significant

across school grades and test scores in seven subjects. Sakic et al. noted the limitations of cross-sectional designs and recommended longitudinal studies to provide stronger tests of school starting age.

Langer et al. (1984) evaluated the effects of relative age on national test scores in grades 4, 8, and 11 in a series of cross-sectional analyses, controlling for background variables and class-average age. For Year 4 students, controlling for class-average age, there was a positive RAE. The sizes of the RAEs were smaller but still significant in Year 8, but were no longer significant by Year 11. Langer et al. also reported negative RAEs for grade retention (i.e., relatively older students were less likely to be retained; also see Elder et al., 2009).

In a cross-sectional study based on a single age cohort, Sprietsma (2010) evaluated the long-term RAEs on mathematics and reading test scores for 15-year olds based on Programme of International Student Assessment (PISA) data, adjusting for separate cut-off dates for starting school. Results indicated an RAE on long-term academic achievement, with the older school entrants having 20% of a standard deviation higher test scores at age 15. Effect sizes varied from .07 and .22, and were statistically significant in 15 (for reading) and 17 (for mathematics) of the 34 regions. Sprietsma suggested that these effects are explained in part by the spreading of students across grades (i.e., grade retention and skipping grades), forms of ability grouping, and other unobserved differences in school systems. Sprietsma noted the need for longitudinal data that included primary school data to disentangle effects of early maturity, teacher expectations, ability grouping, and peer effects on long-term RAEs.

In a recent Australian study, Larsen et al. (2020) considered results from a large database of twins to evaluate the effects of delayed-entry (i.e., "red-shirting") at the start of school. Rather than RAEs per se, their focus was on relative starting age indexed by a dichotomous variable representing delayed entry, excluding students who repeated a grade. Thus, their study did not distinguish between relatively older or younger students unless they delayed their entry into the school by a year. The population was twins from a national twin registry, multiple age cohorts who completed NAPLAN tests in Years 3, 5, 7, and 9. Noting issues in using twin data, they randomly selected one twin from each twin pair rather than explicitly modeling the twin effects. Larsen et al. conducted multiple cross-sectional models separately for each age cohort rather than evaluating longitudinal data, but they did compare results for different age cohorts. They found significant effects of delayed-entry for numeracy and reading for younger children in 3<sup>rd</sup> grade, but those effects diminished for older age cohorts.



Tymms et al. (2000) assessed UK students of 4-5 years ( $n = 1000+$  in 62 classes) at the start of school (reception) and again three years later in Year 2. Analyses included controls for prior schooling (nursery, playgroup, age, gender, cultural capital, and English as a second language). Tymms et al. noted that relatively older students tended to score consistently higher than younger students on all assessments. Differences in age of 6.8 months resulted in differences at Year 2 of .38 SD (mathematics) and .31 (reading). The authors warned of potential harm to young children by starting school too early.

Schnorrbusch et al. (2020) noted that converging evidence that RAEs are risk factors for a range of outcomes, including academic achievement, grade retention, and social impairment. In a meta-analysis of the RAEs and their relation to hyperactivity, Schnorrbusch found support for RAE effects. In addition, they noted that relatively younger children are significantly more likely to be diagnosed with attention deficit hyperactivity disorder (ADHD), and to be prescribed medication (also see Elder et al., 2009; Elder, 2010). They noted that this RAE on ADHD and medication was relatively independent of the child's age.

Summarizing the current research on RAEs in relation to academic achievement, there is wide variation in existing studies in study design, methodology, outcomes, and control variables. Most studies use cross-sectional comparisons that do not control for cognitive differences before children first start school. Nevertheless, particularly for young students in the early primary school years, there is consistent evidence of RAEs for achievement in favor of students who are older than their classmates. Although there is also evidence that these early differences decline with age, this research is based mainly on multiple cross-sectional analyses that do not allow researchers to distinguish between direct, mediated, and total effects over time. RAEs have been found especially in mathematics (Bedard & Dhuey, 2006; Cobley et al., 2009b; Marsh, 2016; Peña, 2007; Sakic et al., 2013; Sprietsma, 2000; Tymms et al., 2000). For language, the results are less consistent (Sharp et al., 1994; Sprietsma, 2010; Cobley et al., 2009b; Tymms et al., 2006). However, research has not consistently tracked the longitudinal RAEs on achievement from early primary school into high school.

### **The Present Study**

Overall, a shared limitation of the previous literature is the lack of a developmental approach through repeated achievement measures across schooling years. This lack of longitudinal studies makes it difficult to provide empirical information about long-term RAEs. Although there is broad agreement on RAEs in early school years, it is not yet clear how long these effects last, the extent to which the RAEs are

present across the span of primary school, and whether RAEs extend to secondary school. Longitudinal studies are needed to reflect better the changes in RAEs on students' academic performance over time.

Furthermore, there is limited evidence on how RAEs are affected by demographic variables such as gender, family SES (Blackmore et al., 2009), and school type (i.e., public vs. private). An important contribution of our study is the inclusion of early childhood measures of cognitive skills collected before the start of kindergarten. Research has shown these to be related to academic achievement and school performance (Bull & Scerif, 2001; Diamond & Lee, 2011; Greene et al., 2018). Marks (2016) found that some cognitive attributes (i.e., persistence) may moderate student achievement, weakening its relationship with SES.

### **Research Hypotheses and Research Questions**

Based on a nationally representative, longitudinal sample using the LSAC data, we evaluate RAEs on achievement (literacy and numeracy) covering the first ten years of formal schooling. Achievement outcomes are high-stakes NAPLAN tests administered nationally across all schools in Years 3, 5, 7, and 9. More specifically, based on our review of the existing literature, we hypothesize that:

1. The direct effects of RAEs will be statistically significant and positive in early school years. Furthermore, we expect that RAEs will have indirect (mediated) effects across all primary and secondary school years. We leave as a research question whether the total RAEs (direct plus mediated) will still be significant at Year 9.
2. Some of the RAEs, can be explained by the set of five covariates (gender, SES, type of school, percent of girls in the school, preschool cognitive skills). However, due to limited evidence that SES might moderate RAEs (e.g., Elder et al., 2009), we leave as a research question as to whether any of these covariates also moderate RAEs.

## **Method**

### **Participants**

We conducted this study using data from Waves 1 to 6 of the Longitudinal Study of Australian Children (LSAC), a large-scale government project run by the Australian Institute of Family Studies (Australian Institute of Family Studies, 2015). Eligible children were identified through Australia's universal healthcare database, Medicare. Multi-stage cluster sampling was used to select the sample (Soloff et al., 2005). The study commenced in 2004 by recruiting participants ( $n = 4,983$ ) who were aged 4-5 years (the so-called Kindergarten cohort) and were born between March 1999 and February 2000. Data were collected

every two years with direct anthropometric measurements and parental questionnaires. Also, students' academic achievement was measured by the National Assessment Program - Literacy and Numeracy (NAPLAN, 2016). In Wave 6, 3,537 participants were still retained in the study (Australian Institute of Family Studies, 2015). Students' academic achievement was measured by the National Assessment Program - Literacy and Numeracy (NAPLAN, 2016). NAPLAN test scores were not part of the LSAC study but were made accessible for data linkage through the data server for the LSAC study.

### **Ethics statement**

This study was conducted following the principles expressed in the World Medical Association Declaration of Helsinki and was approved by the Australian Institute of Family Studies Ethics Committee. Parents or legal guardians provided written informed consent for their participating child at the first wave of data collection.

### **Measures**

**Relative age.** The school year for all Australian states starts at the end of January or the start of February. However, the age a child can start formal schooling is subject to different cut-off dates specific to the state in which the child resides (see Supplemental Material for more information). Because of these differences between states, we defined relative age as a student's age relative to other students in the same school year based on the age at which each student first completed a NAPLAN test. Specifically, we calculated relative age by computing each student's age minus the mean age in their corresponding state in Year 3. We note that a few children moved to another state following Year 3. However, based on the NAPLAN data, the change in location from one state to another over any two-year cycle was small, varying between 1.3% and 1.7%. Also, in some cases, students moved to a different state that had the same age cut-off for starting school and a correspondingly similar state-average age. Furthermore, there was very little movement involving Tasmania (because it is small in size) with the most discrepant cut-off starting age of the different states. Nevertheless, we note that this is an inevitable but potential limitation of a longitudinal study covering the first ten years of schooling.

**Academic achievement.** Academic achievement was based on a standardized test, the National Assessment Program Literacy and Numeracy (NAPLAN), provided by the Australian Curriculum, Assessment and Reporting Authority (ACARA). NAPLAN is an Australian-wide, standardized testing program completed each year by students in Years 3, 5, 7, and 9. Tests in numeracy and literacy (reading, spelling, grammar, and

writing) align with the Australian National Curriculum. All scores are highly reliable (Cronbach's alpha: .86 - .96, ACARA, 2014).

The NAPLAN scores are developed to track Australian students' academic developments over time and across different cohorts. Thus, the scales are comparable for the same student over time and different test administrations. Each year, vertical and horizontal test equating procedures are conducted by the national testing agency to ensure each cohort's results are comparable to all others (ACARA, 2014), across both time and cohorts. Initially, ACARA uses item- and test-analyses to construct reporting scales and score-equivalence tables. Vertical test equating is accomplished using a sample of Australian students and scaled to represent a common metric. This approach is then applied to the entire cohort used to inform national reporting used to report back to schools and parents. The NAPLAN technical report (ACARA, 2014) provides details of the test calibration, test equating and scale construction, differential item functioning analyses, test reliabilities, and item discrimination. ACARA notes that these procedures are similar to those used by the OECD Programme for International Student Assessment (PISA). These procedures are critical for the present investigation. They allow achievement to be tracked over time (longitudinally) for a given cohort of students and cross-sectionally to compare different student cohorts. Hence, NAPLAN test scores' critical feature is that they provide measures on a common metric that makes scores comparable over time (within the same age cohort) and comparable over cohorts (for different cohorts).

The NAPLAN scores were linked to the LSAC dataset. In this way, NAPLAN test scores are provided by ACARA for all students who are part of the LSAC study. Each year the same NAPLAN tests were completed by all students in the same grade (93-97% of all Australian students; ACARA, 2014). NAPLAN tests were assessed bi-annually: Year 3 data were collected mainly in May 2008 or May 2009, depending on the student's year in school. Emphasizing the importance of the scaling done with this data by ACARA (2014), we note that students in the LSAC sample were spread across up to three different school years at any given point in time. However, the NAPLAN data are corrected for fixed levels of the year in school rather than fixed ages. Hence, there is considerable variation in students' relative ages who complete each NAPLAN test wave. Data from Year 5, 7, and 9 were collected 2, 4, and 6 years later (Daraganova et al., 2013).

The NAPLAN tests broadly reflect students' academic achievement in literacy (reading, writing, spelling, grammar, and punctuation) and numeracy (numbers, measurement, and thinking mathematically).

Test scores across the four Year levels (Years 3-9) were standardized to a mean of 500 (range: 1 to 1000) to provide a common metric across different cohorts and over time. It is important to emphasize that collection, test equating, and construction of NAPLAN test scores by ACARA are separate from the standard collection of LSAC data.

*Cognitive measures.* In early childhood, cognitive capacity was based on the Peabody Picture Vocabulary Test (3<sup>rd</sup> edition, PVT). The PVT evaluated children's receptive vocabulary abilities (Dunn & Dunn, 1997) to measure verbal skills. The test was administered when children were 4 or 5 years of age, typically before the start of school.

*Control variables.* Parental questionnaires provided information on several socio-demographic variables. In Wave 1, parents reported their child's gender. Family socio-economic status (SES) was based on wave 1 information on parents' annual family income reports and parental occupation and education. Information on the schools attended from Wave 3 (age 8 or 9) was available through a linked dataset *MySchool* (Baker et al., 2016). School variables used in the present study included the type of school (government, non-government) and the percentage of girls per school. We used the percentage of girls to indicate single-sex schools (but only 1% of children attended schools with more than 90% female pupils). SES, PVT, and the percentage of girls in the school were standardized ( $M = 0$ ,  $SD = 1$ ) to facilitate interpretation of the results.

## **Statistical Analyses**

*Structural equation models.* Structural equation models is a commonly used method for treating longitudinal datasets (Marsh, 2016; Skrondal & Rabe-Hesketh, 2004; Tabachnick & Fidell, 2006). Here, it was used to estimate RAE effects on NAPLAN scores in school Years 3, 5, 7, and 9. We used four models to evaluate longitudinal RAEs from Year 3 till Year 9 (Figure 1). We present both standardized and unstandardized parameter estimates (and standard errors) to facilitate interpretation in relation to an effect size metric.

All our predictor variables are single-indicator constructs, as are the four numeracy measures collected in Years 3, 5, 7, and 9. However, four literacy subtests (reading, writing, spelling, and grammar) were measured each year. Rather than simply taking an average of these four indicators to reflect literacy, we treated literacy as a latent variable. Preliminary analyses indicated that models with a single literacy factor fit the data well. We note that because of the scaling of these literacy indicators by ACARA, which are

measured on a common metric that allows us to compare scores over time (longitudinally) and different age cohorts. Hence, we evaluated the relative contribution of each component to overall literacy as a latent variable construct. We note that this consistency over time in the relative importance is not a necessary condition to our primary analyses, but facilitates the literacy scores' interpretation.

In Supplemental Materials (section 1), we show that the factor loadings relating each of the literacy components to overall literacy are invariant over the four years (metric invariance), indicating that the relative contribution of each of the literacy components is similar across the four years. Although not a necessary requirement for our primary regression analyses, this support for metric invariance facilitates interpretations. We then evaluated support for scalar invariance of indicator intercepts (see related discussion of growth models in Supplemental Materials, section 2). There was also support for scalar invariance for three of the four indicators (all but grammar). Following these analyses, we based subsequent analyses on a model with full metric invariance (factor loadings over time) and partial scalar invariance (invariance of intercepts for all but also the grammar score). This approach to partial invariance is consistent with traditional (e.g., Byrne et al., 1989) and current recommendations (e.g., van de Schoot et al., 2012; Putnick & Bornstein, 2016). Thus, Steenkamp and Baumgartner (1998) suggested that ideally, more than half of indicators on a factor should be invariant. Similarly, Vandenberg and Lance (2000) suggested that a factor is considered to be partially invariant if most of the factor indicators are invariant. Van de Schoot et al. (2012) note that it is possible to make valid inferences about latent means as long as at least two intercepts are constrained to be equal (see Supplemental Materials, section 1 for further discussion). We used this quasi-invariant model for all subsequent analyses.

*Effect Sizes (ESs).* In the present investigation, we present the ESs to aid in the interpretation of the relative importance of parameter estimates. Cohen's *d* is an appropriate measure of effect size, calculated on the adjusted difference of relative age divided by the pooled standard deviation of change (Vacha-Haase & Thompson 2004). Consistent with typical practice, we present ESs as the estimated change in achievement as a function of a one SD change in relative age (.317), which we refer to as ES1. This is typical when the variables vary on an arbitrary metric that is idiosyncratic to the study and has no meaning external to the study. However, this is not the case for relative age that is based on a well-defined metric (age in years) that is common to all waves of the study and has external meaning. Hence, we also present ES2 that is based on the difference between the students who differ by one year of age (i.e., the difference in age associated with

being held back one year; also see related discussion of a growth model approach to the analyses of these data in Supplemental Materials, section 2).

**Missing data.** In a 10-year longitudinal study, there is an inevitable issue of missing data (see summary in Table 1). Multiple imputation is an appropriate approach to overcoming the limitations of traditional approaches to missing data (Graham, 2009). Here, we used the Mplus statistical package to construct ten multiple-imputation data sets. Multiple imputation produces unbiased estimates for missing values even in the case of large numbers of missing values (Enders, 2010). It is an appropriate method to manage missing data in large longitudinal studies (Jeličić et al., 2009). More specifically, as emphasized in classic discussions of missing data (e.g., Newman, 2014), under the missing-at-random (MAR) assumption, the basis of multiple imputations, missingness is allowed to be conditional on all variables included in the analyses but does not depend on the values of variables that are missing. As noted by Graham, even when the data are not purely missing at random, the assumption can be met by the inclusion of auxiliary covariates that can be expected to be related to the missingness. This implies that missing values can be conditional on both covariates included in the analysis, as well as the same variable's values collected in a different wave in a longitudinal panel design. This makes it unlikely that MAR assumptions are seriously violated, as the key situation of not MAR is when missingness is related to the variable itself. Hence, having multiple waves of parallel data provides strong protection against this violation of the MAR assumption.

**Weighting, Stratification, and Clustering.** The sampling procedure of LSAC was stratified to ensure proportional geographic representation for states and territories (Australian Institute of Family Studies, 2015). Participants from 311 postcodes (clusters) were recruited with equal probability, and sample weights were derived to account for non-response. Thus, all analyses in the present investigation were adjusted with the sample weights according to the sampling procedures of LSAC. We used Mplus 8.4 (Muthén & Muthén, 2015) for statistical analyses with the maximum likelihood estimator with robust standard errors. We handled missing data by the construction of multiple imputations based on a missing at random assumption. We used stratification to adjust standard error for stratum in the model estimation and complex design to adjust clustering for postcodes. The analysis was also weighted with child sample weight at wave 1, the sum of which equals the sample size. In Australia, schools and school policy is largely a function of the state or territory. Hence, there are many idiosyncratic state-to-state differences in schools' organization, including cut-off dates for starting school. Following recommendations by McNeish and Stapleton (2016; also see

Larsen et al., 2020), differences between states were controlled with fixed-effect models using dichotomous variables representing each state as a fixed-effect variable.

*Statistical Models for Addressing A Priori Hypotheses.* Our primary analyses were based on a series of multivariate SEM regression models conducted with Mplus (with the appropriate weighting of cases and adjusting for clustering within stratum and postcodes). Specifically, we used this temporal ordering to construct a set of four SEMs (Figures 1 and 2) to evaluate the direct and indirect effects of relative age NAPLAN test scores, controlling for background covariates.

Our primary focus was the direct and indirect (mediated) effects of RAE on subsequent achievement. Direct effects refer to the total RAEs on each of the eight achievement scores (NAPLAN test scores on numeracy and literacy in Years 3, 5, 7, 9). Indirect RAEs are those effects that were mediated through each subsequent wave of achievement. Particularly when controlling for potentially confounding variables and mediation, it is important to establish the appropriate temporal ordering of variables (see Figures 1 and 2). We included student gender, family SES, and school-type as covariate control variables. The PVT was administered when participants were aged four or five, whereas the mean starting age was 5.3 years ( $SD = 0.35$ ). Hence, most of the sample completed this test before starting school. In this sense, it is reasonable to consider this as a covariate even though a few students would have been tested shortly after starting school. Relative age was assessed in Year 3 but is largely determined by school starting age. Thus, relative age follows the covariates and precedes the NAPLAN test administrations in the temporal ordering.

The analytic approach of the present investigation is based on a series of SEMs. In describing the set of SEMs, we begin with a fully articulated cross-lagged panel model (Figure 1; also see Model 3 in Figure 2) that contains all paths, residual variances, and correlations. This cross-lagged panel model is typical of those used in longitudinal panel models (e.g., Kenny, 1975; Marsh, 1990; Marsh, et al., 2018; Marsh, Pekrun, et al., 2017, 2018; Orth et al., 2020). The primary outcome variables are the set of NAPLAN test scores (numeracy and literacy scores in Years 3, 5, 7, and 9). The primary predictor variable is RAE. The critical control variables are a set of five background covariates used to control for pre-existing differences (PVT of preschool cognitive skills, SES, gender, school type, and the percent of females in the school, as well as dummy variables representing state/territory to control for state/territory differences). In this model, the covariates and dummy state/territory variables directly affect relative age and Year 3 achievements, and the effects of covariates on Year 3 achievements are also mediated through relative age. There are direct and



indirect effects of the covariates and relative age on Year 3 outcomes, but all their effects on subsequent achievement outcomes are mediated through Year 3 achievements.

Because there are multiple parallel indicators for the literacy construct, these are latent factors. As is typical in panel studies (e.g., Marsh, Pekrun, et al., 2017, 2018), correlated uniqueness is associated with the same literacy indicator measured on different occasions. The critical parameters in relation to our a priori hypotheses are the effects of relative age on achievement (Hypothesis 1; the direct and mediated effects, but also the total effects—the sum of the direct and mediated effects) and how much of the RAEs can be explained in terms of the set of five covariates (Hypothesis 2). In pursuit of these aims, we used the Mplus "indirect model" option to compute RAE's total indirect effects.

As in the typical cross-lagged panel model (Figure 1), the effects of covariates and relative age on achievement in Years 5, 7, and 9 are posited to be mediated entirely through achievement in Year 3. Although this is the traditional representation of the cross-lagged panel model, this is a highly restrictive assumption (i.e., the covariates and relative age have no direct effects on achievement in Years 5, 7, 9 beyond the effects that mediated through Year 3 achievement). To fully evaluate this model's appropriateness and illustrate different aspects of the RAEs, we juxtaposed the results of four models, portrayed in Figure 2. Each of the four models is a variation of the fully articulated model in Figure 1 (which corresponds to Model 3 in Figure 2) – including the control for state/territory dummy variables. To avoid excessive clutter and complexity, we present them in schematic form and excluded dummy variables representing states and territories. Model 1 (Figure 2) tests RAEs without controlling the effects of covariates. In Model 2, we include the effects of the covariates. Model 3 is the cross-lagged panel model already discussed in Figure 1. Finally, in the full-forward Model 4 (Figure 2), we include direct effects covariates and relative age on all subsequent achievement measures, and effects of each achievement measure on all subsequent achievement measures (e.g., the effects of Year 3 achievement on achievement in Years 5, 7 and 9 as in Model 3). As part of the presentations of results, we articulate the juxtaposition and rationale for these models in more detail.

## Results

### Descriptive statistics

Table 1 presents descriptive statistics of the variables included in this study. Among the 4,983 participants with baseline information, 49% were female. The mean school entry age of the first year in primary school (based on calculation as described in the methods section) was 5.3 years with a minimum age

of 4 years and a maximum of 8.04 years ( $SD = 0.35$ ). There were 67% (2,811 of 4,177) children attending government schools in 2008. In terms of NAPLAN scores, there was an increase in achievement over the four waves across Years 3, 5, 7, and 9. The average literacy score was 424.48 points ( $SD = 69.56$ ) at Year 3, 500.35 ( $SD = 66.25$ ) at Year 5, 547.32 points ( $SD = 63.48$ ) at Year 7, and 586.26 ( $SD = 65.27$ ) at Year 9. The average numeracy score at Year 3 reached 417.42 ( $SD = 74.16$ ) points, 498.98 ( $SD = 72.90$ ) at Year 5, 552.00 ( $SD = 74.14$ ) at Year 7, and 601.46 ( $SD = 73.83$ ) at Year 9. The trend of NAPLAN scores from Year 3 to Year 9 showed slower progression in secondary than primary schools for both literacy and mathematical achievement.

Relative age correlated positively with the Peabody Picture Vocabulary Test (PVT;  $r = .06$ ) at Wave 1 (age =4/5). There were also small, but statistically significant gender differences, with girls being slightly younger than boys. Relative age was also positively correlated with NAPLAN achievement across all four grades ( $r = .12, .09, .05$ , and  $.04$  for numeracy;  $.12, .09, .06$ , and  $.04$  for literacy). However, correlations with achievement declined over time and were not statistically significant at Year 9.

Correlations among the NAPLAN scores were extremely high. Particularly for literacy, correlations between adjacent waves (lag-1 correlations) varied from  $.95$  to  $.97$ . Although not as high, lag-1 correlations for numeracy were also substantial ( $r_s = .73$  to  $.88$ ). The literacy scores were more stable due at least in part because they were latent factors that were corrected from measurement error. We also note that numeracy and literacy scores at each of the four years were also highly correlated with each other ( $r_s = .77$  to  $.80$ ). In summary, there was an extremely positive manifold among the NAPLAN test scores – particularly the lag-1 correlations for literacy.

NAPLAN test scores are also significantly correlated with SES and the PVT of preschool cognitive skills. Although SES and PVT correlated positively with each other ( $r = .27$ ), both these covariates were even more highly correlated with subsequent NAPLAN tests. Overall, NAPLAN test scores are somewhat more highly correlated with SES than the PVT, and these correlations are somewhat higher for literacy than numeracy.

### **The Effect of Relative Age on Achievement across Year 3, 5, 7, and 9: Direct, Indirect, and Total Effects**

#### ***Direct, Indirect, and Total Effects.***

As described earlier, we tested our a priori hypotheses by evaluating the direct and indirect (mediated) RAEs with a set of four SEM regression Models (Figure 2; also see Figure 1). The key variables are a set of five background covariates used to control for pre-existing differences (PVT of preschool cognitive skills, SES, gender, school type, and percent of females in the school), our primary predictor variables (RAE), and the set of NAPLAN test scores (numeracy and literacy scores in Years 3, 5, 7, and 9). We focus on the direct and total RAEs on the eight achievement scores and the total indirect RAEs mediated through each subsequent data wave achievement.

In Model 1 (Figure 1), we evaluate RAEs with no controls for covariates or intervening achievement (there are only paths from relative age to all achievement outcomes). These results are like testing these effects separately for each wave of data with no controls. Because of this model's nature with only direct effects of RAE (see Figure 2), direct and total effects are necessarily the same (i.e., there are no indirect effects). Total RAE's and particularly ES2 estimates are moderate in size and statistically significant for all achievement measures (numeracy and literacy) in Years 3, 5, and 7). However, these effects decline over time, and the total effects are not statistically significant in Year 9.

Model 2 (Figure 2) is like Model 1, with the inclusion of five control covariates. The pattern of effects in Model 2 is similar to that in Model 1. Given the small correlations between the covariates and relative age (Table 1,  $|rs| \leq .06$ ), we did not expect the results to change substantially. Nevertheless, not surprisingly, most of the effects are somewhat smaller after controlling for the set of covariates for all years. In particular, total RAE effects are non-significant for Year 9 achievements and numeracy in Year 7.

Model 3 (Figure 2; also see Figure 1) is the classical cross-lagged panel model in which only lag-1 effects are included for the NAPLAN test scores. In this model, covariates and relative age directly influence achievement in Year 3. However, all their subsequent effects on achievement effects in Year 5 are posited to be mediated through Year 3 achievement. Likewise, effects on Year 7 achievement are mediated through achievement in Years 3 and 5, and effects on Year 9 achievement are mediated through achievement in Years 3, 5, and 7. In this way, RAEs only directly affect Year 3 achievement, and all other effects are mediated through subsequent achievement measures. Thus, RAE's total, direct, and mediated effects continue to be substantial for Year 3, but the direct effects are necessarily zero for direct effects to achievement at Years 5, 7, and 9. For achievement in Years 5, 7, and 9, the total effects are necessarily the same as the indirect. These are all statistically significant and tend to be reasonably similar over time.

Model 4 (Figure 2) is a full-forward model in which paths lead from every variable to all subsequent variables in the temporal ordering schema. The covariates and relative age all directly affect achievements in Years 3, 5, 7, and 9; indirect effects are mediated through intervening achievement measures. Likewise, Year 3 achievements directly affect achievements in Years 7 and 9 (lag-2 and lag-3 paths) and achievement in Year 5 (a lag-1 path). Also, all the indirect RAEs are statistically significant. However, the positive total RAEs are limited primarily to primary school (Years 3 and 5). For secondary school, the total RAEs are only significant for literacy achievement in Year 7.

#### *Selected Parameter Estimates from Model 4*

It is also relevant to examine selected parameter estimates based on Model 4 (Table 4). The two school-level covariates (government school and percent of females) had little or no effect on any achievement tests. We have already evaluated the RAEs in detail. However, of special interest are the effects of the other three covariates—female gender, SES, and PVT. In Year 3 all three are significant for both numeracy and literacy. The effects of PVT and SES were positive for achievements. However, girls score higher on literacy, but lower on numeracy.

Beyond Year 3, there are no additional direct effects of PVT beyond those mediated through the NAPLAN achievement tests. However, SES continues to have additional effects beyond those of achievement in previous years for all four waves. This means that there were new, additional effects of SES beyond those in the previous year in each year. Similarly, particularly for numeracy, there were additional gender differences beyond those evident in Year 3. Thus, each year, there were new, additional negative effects of gender beyond those in the previous year. Although there were counterbalancing positive effects for literacy, these are only significant in Year 5 (i.e., new positive effects of gender beyond those in Year 3).

Not surprisingly, the largest effects for achievement in each year are achievements in the immediately previous year. However, it is interesting to note that numeracy achievement contributes little to the prediction of literacy. However, literacy scores contribute substantially to the predictions on numeracy. Indeed, in some cases, previous literacy's effect on subsequent numeracy approaches or even exceeds the previous numeracy effects (also see the related pattern in correlations in Table 2).

#### *Potential Moderators of the Relative Age Effects*

Finally, we extended Model 4 to determine whether any of the five covariates (gender, SES, type of school, percent of girls in the school, cognitive skills prior to starting school) moderated the RAEs. We

tested these interaction effects in two ways (see Table 5). First, we consider each interaction term separately and then a final model with all five interaction terms. However, for both approaches, all the interaction effects were statistically non-significant for all eight NAPLAN scores. Hence, the RAE effects are robust in relation to the covariates considered here.

### Discussion

Our population-based Australian study investigated RAEs for academic achievement (i.e., literacy and numeracy) across primary and secondary schools in Years 3, 5, 7, and 9, as well as potential confounders and moderators. Consistent with Hypothesis 1, SEMs showed that relatively older students had higher levels of achievement in primary school. This difference declined as students grew older and were no longer statistically significant for numeracy at Year 9. As was the case with raw means (Table 1), RAEs were more substantial in primary school years (Years 3 and 5) than in secondary school years (Years 7 and 9).

Relevant to Hypothesis 2, several of the covariates (SES, PVT, and gender) had substantial effects on achievement in early school years that were maintained or increased across the school years. Gender and cognitive measures (PVT) administered before the start of school also influenced relative age. Thus, girls tended to be somewhat younger than boys, suggesting that delayed entry was more common for boys than girls. Interestingly, the cognitive measure was positively related to relative age, suggesting that delayed entry was not a function of lower cognitive skills. Thus, consistent with Hypothesis 2, some of the RAEs are explained by the set of covariates. However, in response to our associated research question 2, none of the covariates (including school-level covariates) moderated the RAEs (also see the growth model representation presented in Supplemental Section 2.) In this respect, the RAEs were robust. Interestingly, we did not find a significant relation between SES and relative age, nor did SES moderate the RAEs on achievement Years 3, 5, 7, or 9. In summary, our results support both our a priori hypotheses and provided preliminary answers to our research questions.

### *Implications for Research, Policy, and Practice*

#### *Cross-lagged Panel Model 3 and Full-Forward Model 4.*

The differences in Models 3 and 4 are particularly important. The widely used cross-lagged model (Model 3) suggests RAEs that are more substantial and more long-lasting than RAEs in the full-forward model (Model 4). If we had focused only on Model 3, our conclusions would have been quite different. The critical difference between the models is that lag-1 effects cannot explain the longitudinal effects of the

covariates, relative age, or achievement. There are statistically significant lag-2 and lag-3 effects, and for the covariates, even lag-4 effects (i.e., direct effects of covariates on Year 9 NAPLAN tests) that detract from RAEs. Importantly, direct RAEs tend to be negative beyond the first year's substantially positive effects. Hence, there are positive direct RAEs in Year 3, and these positive effects are mediated through Year 3 to achievement in subsequent years. In Model 3, RAEs in subsequent years only reflect these positive mediated effects. However, in Model 4, these indirect RAEs in Year 3 are counter-balanced by new, negative direct RAEs in subsequent years. Thus, in Model 4, the mediated RAEs continue to be positive across all achievement tests. However, the total effects decline over time (also see related growth model approach in Supplemental Materials, Section 2). Based on our results, we recommend that researchers juxtapose the results from alternative models to more fully explicate RAEs.

*The Magnitude of Effect and Effect Size (ES) Metric.* It is also relevant to evaluate the magnitude of RAEs in relation to an interpretable ES metric. In much educational and psychological research, ESs are in relation to an arbitrary metric that is idiosyncratic to a particular study and has no intrinsic meaning. In this situation, it is typical to compute effect sizes in relation to SD units. For example, in our study ES1 is the traditional ES (e.g., Cohen, 1988) — the difference in achievement expected of two children who differ in relative age by one SD (.32 years). ES1 is useful in evaluating the magnitude of effect in relation to natural variation within a typical classroom and comparing results with other observational studies the report ESs in this form.

However, Kelley and Preacher (2012), Wilkinson (1999), and many others emphasize that if the metric is meaningful and has practical implications, then it is preferable to present ESs in relation to this metric. For example, in our study, relative age is based on a meaningful metric with intrinsic meaning — the child's age in years. Consistent with this logic, we also estimated ES2 as the difference in achievement for students who differ in age by one year. ES2 is more useful in evaluating the results of a random-control-trial intervention that held students back for one year or for parents who want to know the likely effects of delaying the entry into school for their child for one year. In each case, the associated change in relative age would be one year (i.e., the age difference that is the basis of ES2).

We do not argue that either of these ES metrics is inherently superior. We note that they reflect simple transformations of each other ( $ES2 = ES1 * 1 \text{ year} / .317 \text{ years}$ , where .317 is the SD of relative age differences). Instead, they provide alternative perspectives on the practical significance of our RAE

results. Notably, the traditional ES1 underestimates the effects associated with holding back or accelerating a child. Based on our results, we recommend that researchers carefully consider the implications of alternative ESs in describing the magnitude of RAEs.

*Why Do Young Children Perform Poorly In Early School Years?* Possible reasons are: relative cognitive and social immaturity; the inappropriateness of formal curriculum; teachers' lower expectations (Crawford et al., 2007; Sakic et al., 2013; Sharp, Hutchinson, & Whetton, 1994; Sprietsma, 2010), or the inevitable social comparison processes that influence perceptions of teachers as well as those of the children themselves (Marsh, 2016; Marsh et al., 2018).

The current teaching methods and curriculum may better suit older students, with younger students struggling to keep up with the requirements (Sharp et al., 2009). Teachers are sometimes unable to distinguish between maturity and ability when children start school (Allen & Barnsley, 1993). Consequently, younger school entrants may experience stress and failure to adapt to the school settings (Gledhill et al., 2002; Martin et al., 2004; Morrison et al., 1997; Sharp et al., 1994). This represents a social comparison process in relation to teacher judgments (e.g., rating younger children as having more language deficits, poorer academic progress, and more behavior problems at the end of the year; Norbury et al., 2016). Apart from the cognitive immaturity of younger children, other explanations for the RAE can result from grade repetition, selection of students into different grades, and co-existence of different ability groups in the same class (Sprietsma, 2010).

Our findings suggest that there are substantial RAEs for standardized tests in primary schools in Australia, but the direct effects of RAEs decline and largely disappear over secondary school years. Larger RAEs in primary than secondary school were also found in other countries comparing different cohorts and testing periods, with early enrolment contributing to lower scores across the entire education path (e.g., Italy; Ordine & Rose, 2019; Mexico; Peña, 2017). It is important to note that the first study included only primary years (Ordine & Rose, 2019), whereas the latter was similar to our study (i.e., grades 3-9; Peña, 2017). It will be insightful if future research includes measures of later years of secondary school (e.g., Year 11), as well as rates of students entering University (e.g., Parker et al., 2018), and labor market outcomes (e.g., Peña, 2017).

The focus of our study was on nationally administered standardized achievement tests. However, if not managed, RAEs can cause detrimental effects on children's and adolescents' cognitive and personal

development: Being younger relative to the modal grade also negatively affects self-concept and self-beliefs (Marsh, 2016; Parker et al., 2019). Cobley et al. (2009b) found that among secondary students, the older students in the same year were more likely to be identified as gifted-and-talented. In contrast, younger students were more likely to be diagnosed with learning difficulties or required learning support. Younger students also had worse attendance rates than older students, possibly due to decreased motivation. More broadly, long-term RAEs are associated with grade repetition; emotional, psychological, and behavioral problems; less educational attainment; fewer leadership positions; less earning; and having less educated spouses (Black et al., 2011; Cobley et al., 2009b; Dhuey & Lipscomb, 2010; Dobkin & Ferreira, 2010; Goodman et al., 2003; Kawaguchi, 2011; Peña, 2017; Sprietsma, 2010). Indeed, in one of the few meta-analyses associated with RAEs, Schnorrbusch et al. (2020) found that relatively younger children were disproportionately diagnosed with ADHD and prescribed medication. Schnorrbusch et al. suggested that at least some of this difference was due to the psychological, cognitive, and emotional immaturity of relatively younger children rather than underlying psychological problems per se.

***Social Comparison Processes.*** Implicit in many explanations of RAEs are social comparison processes in which students within the same classroom are compared to each other—by the teacher, their parent, or by themselves. Thus, social comparison processes influence teachers' perceptions but also students' self-perceptions. This focus on social comparison processes was made explicit in the Marsh (2016) study of academic self-concept. In the formation of academic self-concept, students compare their own academic performances with those of other students (social comparisons), changes in their own performances over time (temporal comparisons), their relative performances in different school subjects (dimensional comparisons), or comparison to an absolute standard (Marsh, Parker, et al., 2020). Particularly in educational settings, social comparisons are the most salient comparison process. Thus, for example, unlike a sporting context, in educational settings, students are rarely given feedback in relation to an absolute metric that facilitates comparisons over time or in relation to an absolute standard (e.g., how fast you can run, how high you can jump, or how many push-ups you can do).

One of the most well-documented social comparison processes in the academic self-concept literature is the big-fish-little-pond effect (BFLPEs). According to the BFLPE, being a student in a school or class where classmates' average achievement is high negatively affects their self-concept. Thus, equally able students will have lower academic self-concepts in classrooms where the average ability level is high, and



lower academic self-concepts in classrooms where the average achievement is low. Indeed, cross-national support for the BFLPE (see Marsh et al., 2018, 2019, 2020) makes it one of educational psychology's most universal findings.

Marsh (2016) made explicit the social comparison basis of the RAE. He proposed that relatively younger students compare their accomplishments with relatively older classmates who are physically, cognitively, socially, and emotionally more mature. Based on a large cross-national study of 15-year olds using PISA data, he juxtaposed BFLPE and RAE as two distinct forms of social comparison processes in the formation of academic self-concept. In each case, there were adverse effects on the academic self-concept of school-average achievement (BFLPEs) and being relatively younger than classmates (RAEs) that generalized over 41 countries in the PISA database. Marsh emphasized that the RAEs and BFLPEs were independent of each other and represented two distinct sources of social comparison processes. Parker, Marsh (2018) subsequently pursued these suggestions in a longitudinal study using an instrumental-variable approach that provided stronger causal claims of RAEs. In addition to replicating the RAEs on academic self-concept, they also demonstrated RAEs on subsequent university entry that could be explained in terms of academic self-concept. These results validate long-term RAEs and suggest that social comparison processes are the mechanism through which this occurs. Marsh (2016) demonstrated that relative age was a critical variable in forming self-concept rather than how children came to be younger or older than their classmates (e.g., birthdate, starting early or late, retention, acceleration). This led Parker, Marsh, et al. to conclude that "the ever-present comparison with same class peers becomes an indomitable influence in the way in which children come to view themselves academically" (p. 128).

Our study of RAEs is consistent with a broader perspective of the central importance of cultural traditions, practices, policies, or theories about school selection (Parker, Marsh, et al., 2018, 2019). Social comparison processes are central in the way children form their self-beliefs. Any decision that affects who a child's peers will be will play a significant role in how they feel about themselves. This acknowledgment of macro-forces on the individual's intrapsychic constitution is a central feature of theoretical models in educational and developmental psychology (Bronfenbrenner, 1979; Elder, 1998; see discussion by Parker, Marsh, et al., 2019). However, these theoretical models were not strong in articulating the processes by which it happens. Here, and in related research, we articulate social comparison mechanisms through government policy at the macro-contextual level, and mediating structures at the meso-contextual level.

These then influence children's formation of self-belief and identity in educational settings as well as other life domains.

*Strengths, Weaknesses, and Directions for Further Research*

Particularly in relation to existing RAE research, our study's most important strength is the design and analysis of longitudinal data covering the first ten years of schooling. Unlike most RAE studies that are based on cross-sectional data, ours is a true longitudinal study. Unlike many studies based on idiosyncratic samples that undermine generalizability, our sample is a nationally representative sample of Australian students. Although most RAE studies include covariates like some of those we considered, a unique aspect of our study is the inclusion of a measure of cognitive skills administered before students began kindergarten. A particular strength is the availability of the Australian-wide NAPLAN tests for students in Years 3, 5, 7, and 9. These achievement measures are high-stakes tests within the Australian context that align closely with the national curriculum.

Hence, the tests are likely to be completed more seriously by students and better represent true optimal performance than low-stakes tests used in many RAE studies. Furthermore, because of ACARA's extensive psychometric scaling work, NAPLAN test scores vary along a common metric that facilitates comparisons across different cohorts and over time. Coupled with strong SEMs, our study provides a particularly strong basis for testing our research hypotheses. Indeed, we know of no other RAE study based on a prospective, true longitudinal sample covering the development of achievement over the first ten years of schooling.

Nevertheless, readers need to consider several weaknesses when interpreting the results that also provide further research directions. We based our relative age measure on the NAPLAN test's student age and location in Year 3. This provides a strong measure of relative age at that point in time, but is not fully adequate in establishing the age at which students first started school or relative age in the subsequent school years. For present purposes, we considered relative age relative to the average age of students within each Australian state and territory. This is important because different states have somewhat different policies about starting school, including different cut-off ages. Hence, the absolute average age of students in Year 3 differed for each state, and our measure of relative age accounted for these differences. Although this might not fully account for students who moved to a different state following Year 3, we note that the numbers of students doing this were small (between 1.3 and 1.7% across every two-year cycle of NAPLAN tests, or 3.8% across the entire Year 3 to Year 9 period).

Different Australian states and territories had slightly different cut-off ages for starting school. This situation would seem to be heuristic in establishing differences in absolute and relative ages. However, such differences are completely confounded with the many differences that exist between the states. Furthermore, because we included multiple dummy variables for each state (the fixed-effect approach recommended by McNeish & Stapleton, 2016, also see related discussion by Larsen et al., 2020), all state-specific differences were eliminated. Nevertheless, we note that such naturally occurring variation can be an important source for quasi-experimental research that would provide an alternative to the design used here. A true experimental design with random assignment might be even more robust in relation to internal validity. However, such a study would likely be weak in terms of external validity and, perhaps, ethically dubious.

There is increasing evidence to suggest that the local school and even the class within a school would have provided a stronger basis comparison in relation to social comparison processes posited as one basis of RAEs (see Marsh, 2016, for discussion of local dominance effects for academic self-concept formation). Future research should establish the relative age for more local contexts than the entire state or whole countries. More broadly, it would be useful to juxtapose RAEs based on standardized achievement measures like those used here with a more diverse set of outcomes. This might include school grades, academic self-concept, motivation, school belonging, and satisfaction of psychological needs (e.g., autonomy, belonging, and competence posited as central in self-determination theory; see Deci & Ryan, 1985).

### **Conclusions**

Our study highlighted the positive influence of relative age in primary school. As more years of schooling result in better academic scores, relative age differences decline or vanished at later secondary education stages. Initial cognitive abilities and SES explained part of the RAEs in academic performance. Teachers must identify early developmental or cognitive discrepancies among their students, ensuring that they provide all their students with the necessary skills. Such skills are needed to improve achievements in numeracy and to enable a swift catch up in literacy. Concerning state policy, it is fundamental that equal opportunities for knowledge are provided to students regardless of their school. Importantly, before rushing to judge students' cognitive deficiencies and psychological problems, children's relative age needs to be considered. We posited that social comparison processes are a fundamental mechanism driving RAEs, which has important implications for future research, policy, practice, and educational choice.

### References

- Allen, J., & Barnsley, R. (1993). Streams and tiers: The interaction of ability, maturity, and training in systems with age-dependent recursive selection. *The Journal of Human Resources*, 28(3), 649-659.
- Australian Government Department of Social Services [ACARA] (2014). *Growing Up in Australia: The Longitudinal Study of Australian Children: 2011–12 Annual Report*. Canberra, Commonwealth of Australia: Australian Government. Retrieved December 20, 2018 from:  
[https://www.dss.gov.au/sites/default/files/documents/10\\_2013/the\\_longitudinal\\_study\\_of\\_australian\\_children\\_2011-12\\_annual\\_report.pdf](https://www.dss.gov.au/sites/default/files/documents/10_2013/the_longitudinal_study_of_australian_children_2011-12_annual_report.pdf)
- Australian Institute of Family Studies. (2015). *Data user guide. The Longitudinal Study of Australian Children: An Australian government initiative*. Canberra, Commonwealth of Australia: Retrieved December 20, 2018, from <https://growingupinaustralia.gov.au/sites/default/files/userguide.pdf>
- Baker, K., Maguire, B., Daraganova, G., & Siphthorp, M. (2016). *Using My School data in the Longitudinal Study of Australian Children*. National Centre for Longitudinal data, Department of Social Services: Canberra, Australia. Retrieved June 22, 2020 from  
<https://growingupinaustralia.gov.au/sites/default/files/tp16.pdf>
- Barnsley, R. H., Thompson, A. H., & Barnsley, P. E. (1985). Hockey success and birthdate: The relative age effect. *Journal of Canadian Association of Health, Physical Education and Recreation*, 51, 23-28.
- Barnsley, R. H., Thompson, A. H., & Legault, P. (1992). Family planning: Football style. The relative age effect in football. *International Review for the Sociology of Sport*, 27(1), 77-87.
- Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, 121(4), 1437-1472.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.
- Bertram, T., & Pascal, C. (2002). *Early years education: An international perspective*. London, UK: QCA.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics*, 93(2), 455-467.
- Blakemore, T., Strazdins, L., & Gibbings, J. (2009). Measuring family socioeconomic position. *Australian Social Policy*, 8, 121-168.
- Borg, M. G., & Falzon, J. M. (1995). Birth date and sex effects on the scholastic attainment of primary school children: A cross-sectional study. *British Educational Research Journal*, 21(1), 61-74.

- Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, 34(10), 844-850.
- Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, 19(3), 273-293.
- Byrne, B., Shavelson, R., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466.
- Carroll, H. C. M. (1992). Season of birth and school attendance. *British Journal of Educational Psychology*, 62(3), 391-396.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504.
- Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods*, 4(3), 236-264.
- Crawford, C., Dearden, L., & Meghir, C. (2007). *When you are born matters: The impact of date of birth on child cognitive outcomes in England*. The Institute for Fiscal Studies: London, UK. Retrieved June 22, 2020 from [https://www.ifs.org.uk/docs/born\\_matters\\_report.pdf](https://www.ifs.org.uk/docs/born_matters_report.pdf)
- Cobley, S., Baker, J., Wattie, N., & McKenna, J. (2009a). Annual age-grouping and athlete development: a meta-analytic review of relative age effects in sport. *Sports Medicine*, 39(3), 235-256.
- Cobley, S., McKenna, J., Baker, J., & Wattie, N. (2009b). How pervasive are relative age effects in secondary school education? *Journal of Educational Psychology*, 101(2), 520-528.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, Erlbaum.
- Crawford, C., Dearden, L., & Meghir, C. (2007). *When you are born matters: The impact of date of birth on child cognitive outcomes in England*. Centre for the Economics of Education (CEE) Discussion Paper 0093.
- Daraganova, G., Edwards, B., & Siphthorp, M. (2013). *Using National Assessment Program Literacy and Numeracy (NAPLAN) data in the Longitudinal Study of Australian Children (LSAC)*. Australian Institute of Family Studies. Australian Government, Department of Families, Housing, Community Services and Indigenous Affairs: Canberra, Australia. Retrieved June 22, 2020 from <https://core.ac.uk/download/pdf/30676984.pdf>

- Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality, 19*(2), 109-134.
- DeCos, P. L. (1997). Readiness for kindergarten: What does it mean?. Sacramento, CA: California Research Bureau.
- Dhuey, E., & Lipscomb, S. (2010). Disabled or young? Relative age and special education diagnoses in schools. *Economics of Education Review, 29*(5), 857-872.
- Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (2019). School starting age and cognitive development. *Journal of Policy Analysis and Management, 38*, 538–578.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333*(6045), 959-964.
- Dobkin, C., & Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes?. *Economics of Education Review, 29*(1), 40-54.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test (3rd ed.)*. Bloomington, MN: Pearson Assessments.
- Elder, G. H. (1998). The life course as developmental theory. *Child Development, 69*, 1–12.
- Elder, T. E. (2010). The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates. *Journal of Health Economics, 29*(5), 641–656.
- Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement. *The Journal of Human Resources., 44*(3), 641–683.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Frey, N. (2005). Retention, social promotion, and academic redshirting: What do we know and need to know? *Remedial and Special Education, 26*(6), 332-346.
- Gledhill, J., Ford, T., & Goodman, R. (2002). Does season of birth matter? The relationship between age within the school year (season of birth) and educational difficulties among a representative general population sample of children and adolescents (aged 5–15) in Great Britain. *Research in Education, 68*(1), 41-47.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576.

- Graue, M. E., & DiPerna, J. (2000). Redshirting and early retention: Who gets the gift of time and what are its outcomes? *American Educational Research Journal*, *37*(2), 509-534.
- Greene, J. A., Cartiff, B. M., & Duke, R. F. (2018). A meta-analytic review of the relationship between epistemic cognition and academic achievement. *Journal of Educational Psychology*, *110*(8), 1084-1111.
- Goodman, R., Gledhill, J., & Ford, T. (2003). Child psychiatric disorder and relative age within school year: Cross-sectional survey of large population sample. *BMJ*, *327*, 1-4.
- Hancock, D. J., Adler, A. L., & Côté, J. (2013). A proposed theoretical model to explain relative age effects in sport. *European Journal of Sport Science*, *13*(6), 630-637.
- Hanly, M., Edwards, B., Goldfeld, S., Craven, R. G., Mooney, J., Jorm, L., & Falster, K. (2019). School starting age and child development in a state-wide, population-level cohort of children in their first year of school in New South Wales, Australia. *Early Childhood Research Quarterly*, *48*, 325-340.
- Hauck, A. L., & Finch, A. J. (1993). The effect of relative age on achievement in middle school. *Psychology in the Schools*, *30*(1), 74-79.
- Helsen, W. F., Van Winckel, J., & Williams, A. M. (2005). The relative age effect in youth soccer across Europe. *Journal of Sports Sciences*, *23*(6), 629-636.
- Jeličić, H., Phelps, E., & Lerner, R. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*(4), 1195-1199.
- Katz, L. G. (2000). Academic redshirting and young children. Washington, DC: Office of Education Research and Improvement.
- Kawaguchi, D. (2011). Actual age at school entry, educational outcomes, and earnings. *Journal of the Japanese and International Economies*, *25*(2), 64-80.
- Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137-152.
- Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*(6), 887-903. <https://doi.org/10.1037/0033-2909.82.6.887>
- Langer, P., Kalk, J. M., & Searls, D. T. (1984). Age of admission and trends in achievement: A comparison of Blacks and Caucasians. *American Educational Research Journal*, *21*(1), 61-78.
- Larsen, S. A., Little, C. W., & Coventry, W. L. (in press). Exploring the associations between delayed school entry and achievement in primary and secondary school. *Child Development*.

- Lawlor, D. A., Clark, H., Ronalds, G., & Leon, D. A. (2006). Season of birth and childhood intelligence: Findings from the Aberdeen children of the 1950s cohort study. *British Journal of Educational Psychology, 76*(3), 481–499.
- Lien L, Tambs K, Oppedal B, Heyerdahl S, Bjertness E. (2005). Is relatively young age within a school year a risk factor for mental health problems and poor school performance? A population-based cross-sectional study of adolescents in Oslo, Norway. *BMC Public Health, 5*, 1-8.
- Marks, G. N. (2016). The relative effects of socio-economic, demographic, non-cognitive and cognitive influences on student achievement in Australia. *Learning and Individual Differences, 49*, 1-10.
- Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology, 108*(2), 256-273.
- Marsh, H. W., Parker, P. D., Guo, J., Pekrun, R., & Basarkod, G. (2020). Psychological comparison processes and self-concept in relation to five distinct frame-of-reference effects: Pan-human cross-cultural generalizability over 68 countries. *European Journal of Personality, 34*(2), 180-202.
- Marsh, H. W., Parker, P. D., & Pekrun, R. (2018). Three paradoxical effects on academic self-concept across countries, schools, and students. *European Psychologist, 24*(3), 231-242.
- Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology, 54*(2), 263–280.
- Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology, 109*(3), 425-438.
- Marsh, H., Seaton, M., Dicke, T., Parker, P., & Horwood, M. (2019). 2 The centrality of academic self-concept to motivation and learning. *The Cambridge handbook of motivation and learning*, 36-62.
- Martin, R. P., Foels, P., Clanton, G., & Moon, K. (2004). Season of birth is related to child retention rates, achievement, and rate of diagnosis of specific learning disabilities. *Journal of Learning Disabilities, 37*(4), 307-317.
- McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research, 51*(4), 495-518.



- Mergler, A., & Walker, S. (2017). 'This is possibly THE hardest decision a parent has to make.' Deciding when your child is ready to start Prep. *Australasian Journal of Early Childhood*, 42(2), 97-104.
- Morrison, F. J., Alberts, D. M., & Griffith, E. M. (1997). Nature–nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology*, 33(2), 254-262.
- Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology*, 31(5), 789-799.
- Musch, J., & Grondin, S. (2001). Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental Review*, 21(2), 147-167.
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide*. Los Angeles: Muthen & Muthen.
- National Assessment Program [NAPLAN], 2016. Key dates. NAPLAN. Retrieved December 20, 2018 from: <https://www.nap.edu.au/information/key-dates>
- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4), 372-411.
- Norbury, C. F., Gooch, D., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2016). Younger children experience lower levels of language competence and academic progress in the first year of school: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(1), 65-73.
- Ordine, P., & Rose, G. (2019). Early entry, age-at-test, and schooling attainment: evidence from Italian primary schools. *Economia Politica*, 36(3), 761-784.
- Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2020). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*. Advance online publication. <http://dx.doi.org/10.1037/pspp0000358>
- Parker, P. D., Marsh, H. W., Jerrim, J. P., Guo, J., & Dicke, T. (2018). Inequity and excellence in academic performance: Evidence from 27 countries. *American Educational Research Journal*, 55(4), 836-858.
- Parker, P. D., Marsh, H. W., Thoemmes, F., & Biddle, N. (2019). The negative year in school effect: Extending scope and strengthening causal claims. *Journal of Educational Psychology*, 111(1), 118-130.
- Peña, P. A. (2017). Creating winners and losers: date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, 56, 152-176.
- Pellizzari, M., & Billari, F. C. (2012). The younger, the better? Age-related differences in academic performance at university. *Journal of Population Economics*, 25(2), 697-739.

- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-
- Sakic, M., Burusic, J., & Babarovic, T. (2013). The relation between school entrance age and school achievement during primary schooling: Evidence from Croatian primary schools. *British Journal of Educational Psychology, 83*(4), 651-663.
- Sanson, A. V., Nicholson, J., Ungerer, J., Zubrick, S., Wilson, K., Ainley, J., et al. (2002). *Introducing the Longitudinal Study of Australian Children: LSAC Discussion Paper N0. 1*. Melbourne: Australian Institute of Family Studies.
- Schnorrbusch, C., Fabiano, G. A. Aloe, A. M. & Rodriguez, R. C. T. (2020). Attention deficit hyperactivity disorder and relative age: A meta-analysis. *School Psychology Review, 49*, 1, 2-19.
- Sharp, C. (1995). What's age got to do with it? A study of patterns of school entry and the impact of season of birth on school attainment. *Educational Research, 37*(3), 251-265.
- Sharp, C., George, N., Sargent, C., O'Donnell, S., & Heron, M. (2009). International Thematic Probe: The Influence of Relative Age on Learner Attainment and Development. *National Foundation for Educational Research*. Retrieved December 20, 2018 from: <https://files.eric.ed.gov/fulltext/ED508563.pdf>
- Sharp, C., Hutchison, D., & Whetton, C. (1994). How do season of birth and length of schooling affect children's attainment at key stage 1? *Educational Research, 36*(2), 107-121.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*: CRC Press.
- Soloff, C., Lawrence, D., & Johnstone, R. (2005). Sample design (LSAC technical paper no. 1). *Melbourne: Australian Institute of Family Studies, 1*, 30.
- Solomon, G. B., DiMarco, A. M., Ohlson, C. J., & Reece, S. D. (1998). Expectations and coaching experience: Is more better?. *Journal of Sport Behavior, 21*(4), 444-455.
- Sprietsma (2010) Effect of relative age in the first grade of primary school on long-term scholastic results: International comparative evidence using PISA 2003, *Education Economics, 18*(1), 1-32.
- Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach.

*Multivariate Behavioral Research*, 25(2), 173-180.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics*. Needham Heights, MA: Allyn & Bacon, Inc.

Thompson, A. H., Barnsley, R. H., & Battle, J. (2004). The relative age effect and the development of self-esteem. *Educational Research*, 46(3), 313-320.

Thompson, A. H., Barnsley, R. H., & Dyck, R. J. (1999). A new factor in youth suicide: The relative age effect. *Canadian Journal of Psychiatry*, 44, 82-85.

Thompson, A. H., Barnsley, R. H., & Stebelsky, G. (1991). "Born to play ball" The relative age effect and major league baseball. *Sociology of Sport Journal*, 8(2), 146-151.

Tymms, P., Merrell, C., & Henderson, B. (2000). Baseline assessment and progress during the first three years at school. *Educational Research and Evaluation*, 6(2), 105-129.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1-10.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473-481.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492.

Van Rossum, J. H. (2006). Relative age effect revisited: findings from the dance domain. *Perceptual and Motor Skills*, 102(2), 302-308.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.

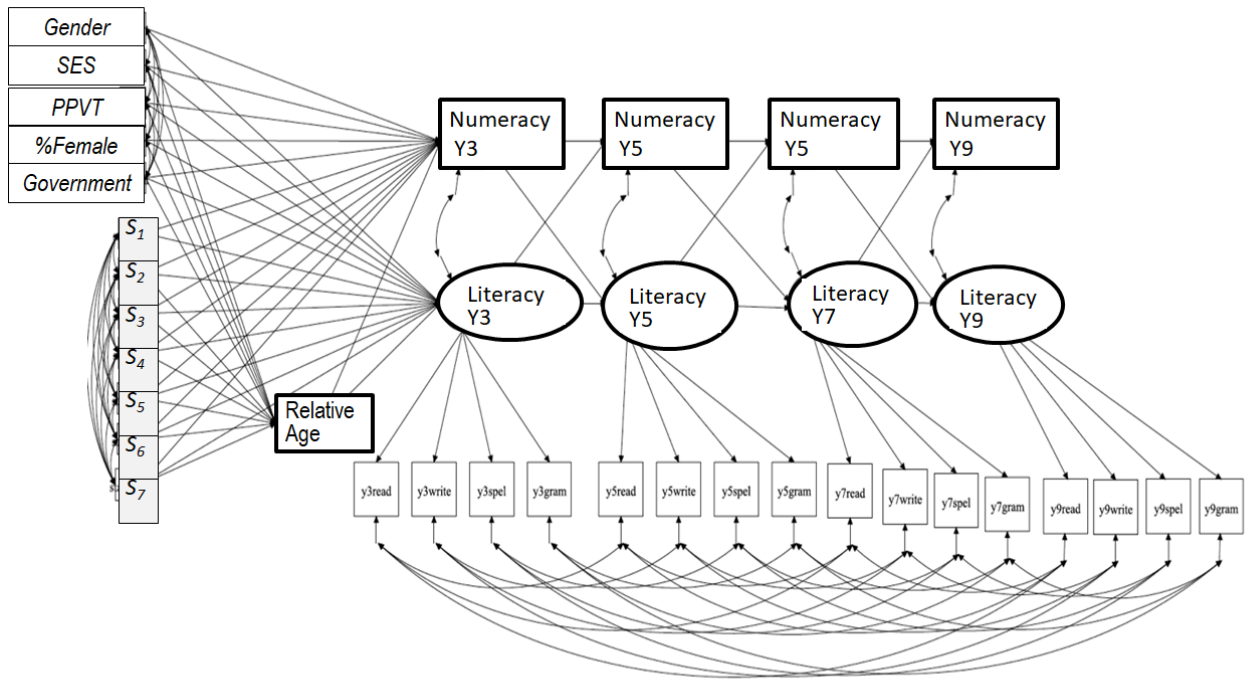


Figure 1. Cross-Lagged Panel Model: Longitudinal Effects of Relative Age Effects.

Square boxes represent manifest variables based on single indicators. Ovals represent latent variables based on multiple indicators. Straight lines represent paths, and curved lines represent covariances. SES = socio-economic status; PPVT = Peabody Picture Vocabulary Test; %Female = percent of students in a school who are girls; Government = school-type (government vs private/independent); S1 – S7 = dummy variables representing the different Australian states and territories.

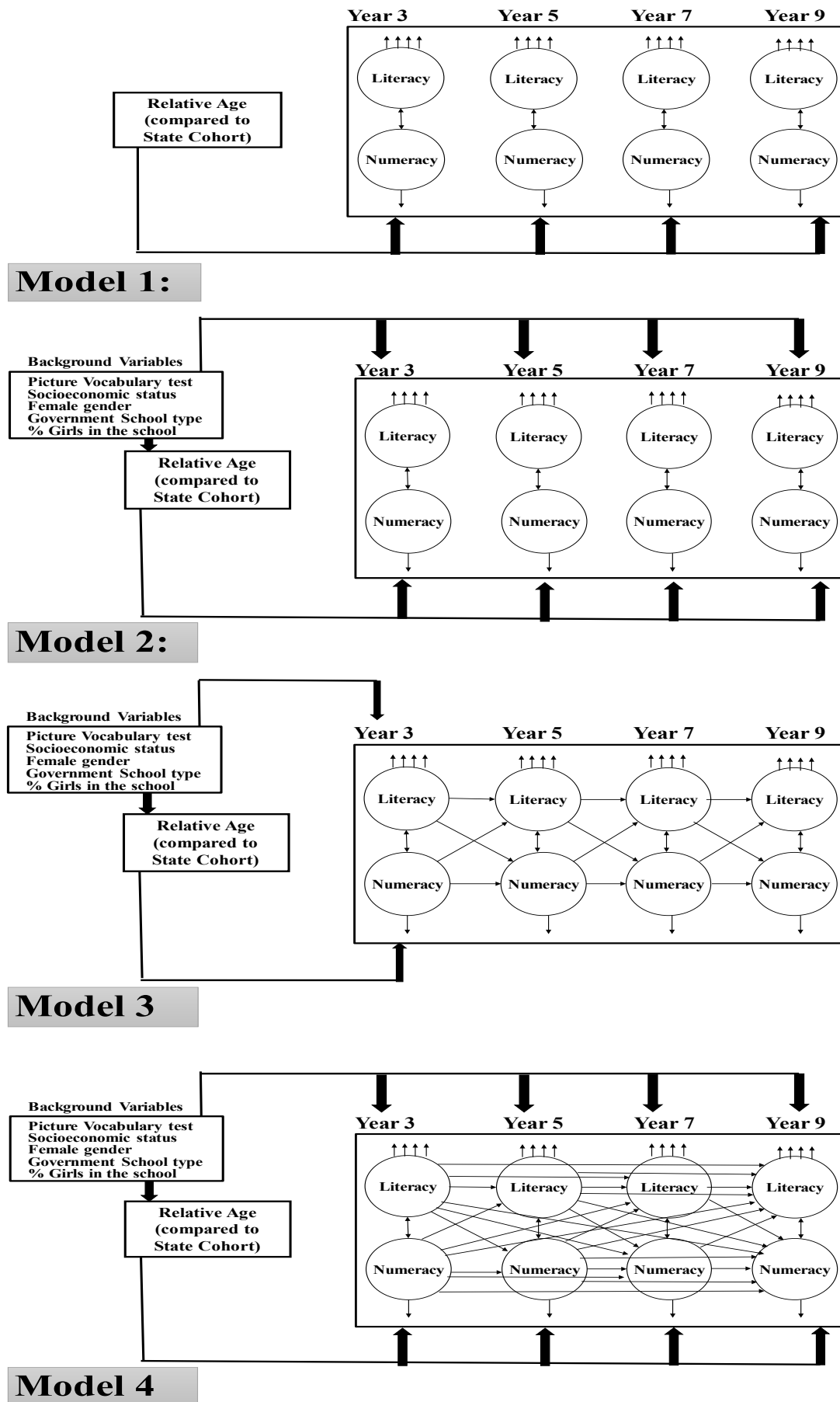


Figure 2. Schematic Diagram of Four Models Evaluating the Effects of Relative Age.

*Note.* Schematic diagrams for four models of the effects of relative age effects (RAs) on subsequent achievement. Model 1 evaluates RAE and repetition effects with no controls for covariates or intervening achievement (there are only paths from RA to all achievement outcomes). Model 2 adds controls for covariates (with paths from each covariate to all achievement outcomes). Model 3 is a classic cross-lagged panel model. It includes lag-1 effects of all variables (e.g., effects of covariates and RAEs directly affect Year 3 achievement which mediates subsequent effects; each achievement only has effects on the immediately subsequent achievement). Model 4 is a full-forward model with paths from all variables to all subsequent variables. Not shown to avoid clutter are the correlated uniqueness relating the matching indicators of literacy from one wave to the next and dummy variables representing states and territories (but see figure 1 for a more fully articulated representation of Model 3).

Table 1. Means and standard deviations of Study Variables.

Variables	N (original)	N (Imputed)	Mean	Median	SD	Skews	Kurts	Mini	Max	20%tile	40%tile	60%tile	80%tile
Female (gender)	4983	4983	1.49	1.00	0.50	0.05	-2.00	1.00	2.00	1.00	1.00	2.00	2.00
Socioeconomic status (age 4/5)	4965	4983	0.00	-0.04	1.00	0.36	-0.19	-3.67	3.15	-0.74	-0.27	0.25	1.05
PVT (age 4/5)	4406	4983	63.64	64.00	6.27	-0.40	2.40	28.00	85.00	59.00	62.17	65.00	70.00
% girls in school	4148	4983	0.00	-0.05	1.00	2.72	25.45	-6.65	7.42	-0.47	-0.17	0.09	0.42
School Type (government)	4177	4983	0.69	1.00	0.46	-0.83	-1.31	0.00	1.00	0.00	1.00	1.00	1.00
Relative Age	3096	4983	0.00	-0.02	0.32	0.17	2.54	-1.36	2.86	-0.27	-0.10	0.06	0.24
NAPLAN numeracy (Year 3)	2987	4983	4.12	4.14	0.75	0.14	0.21	1.38	6.66	3.51	3.95	4.31	4.76
NAPLAN numeracy (Year 5)	3907	4983	4.96	4.95	0.73	0.36	0.53	2.24	8.30	4.39	4.76	5.15	5.57
NAPLAN numeracy (Year 7)	3707	4983	5.46	5.44	0.76	0.35	0.29	2.99	9.23	4.87	5.27	5.64	6.10
NAPLAN numeracy (Year 9)	3110	4983	5.91	5.91	0.76	0.30	0.33	3.00	9.20	5.31	5.73	6.09	6.55
NAPLAN literacy Y3READ	2986	4983	4.14	4.19	0.87	-0.05	0.40	0.05	6.92	3.45	3.97	4.40	4.89
NAPLAN literacy Y3WRITE	2993	4983	4.21	4.23	0.73	-0.28	0.67	0.89	6.85	3.68	4.05	4.43	4.80
NAPLAN literacy Y3SPEL	2993	4983	4.10	4.14	0.79	-0.14	0.18	1.28	6.69	3.49	3.96	4.33	4.74
NAPLAN literacy Y3GRAM	2989	4983	4.20	4.24	0.90	-0.06	0.25	0.19	7.16	3.50	4.05	4.43	4.99
NAPLAN literacy Y5READ	3928	4983	4.98	5.01	0.83	-0.01	0.14	0.90	8.42	4.32	4.78	5.25	5.67
NAPLAN literacy Y5WRITE	3916	4983	4.88	4.95	0.71	-0.35	1.80	0.89	7.78	4.34	4.80	5.09	5.44
NAPLAN literacy Y5SPEL	3922	4983	4.89	4.92	0.72	-0.16	0.05	2.41	6.84	4.33	4.76	5.07	5.49
NAPLAN literacy Y5GRAM	3922	4983	5.06	5.14	0.88	-0.10	0.72	0.94	8.39	4.37	4.94	5.35	5.81
NAPLAN literacy Y7READ	3721	4983	5.49	5.53	0.73	-0.03	-0.21	2.61	7.85	4.91	5.35	5.73	6.15
NAPLAN literacy Y7WRITE	3718	4983	5.25	5.23	0.78	-0.24	1.29	0.95	8.39	4.66	5.12	5.46	5.95
NAPLAN literacy Y7SPEL	3724	4983	5.44	5.47	0.71	-0.19	0.16	2.58	7.63	4.90	5.32	5.65	6.00
NAPLAN literacy Y7GRAM	3724	4983	5.49	5.47	0.80	0.12	0.65	1.93	8.08	4.90	5.33	5.69	6.22
NAPLAN literacy Y9READ	3129	4983	5.85	5.90	0.71	-0.07	0.41	1.96	8.91	5.29	5.71	6.07	6.49
NAPLAN literacy Y9WRITE	3128	4983	5.54	5.58	0.92	-0.61	2.24	0.95	8.77	4.89	5.36	5.83	6.29
NAPLAN literacy Y9SPEL	3135	4983	5.83	5.88	0.74	-0.22	0.39	3.17	8.08	5.27	5.70	6.03	6.45
NAPLAN literacy Y9GRAM	3135	4983	5.77	5.81	0.84	0.01	0.82	1.79	8.94	5.15	5.61	5.95	6.47

**Note:** N original and imputed refer to Ns before and after multiple imputations. All descriptive statistics are based on imputed values. Relative age was the difference between each child's actual age and that state-average age in the state in which they resided for the first NAPLAN test that they completed.

Table 2. *Correlations between Study Variables.*

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Relative age in years	<b>1.00</b>													
2 Female	<b>-0.04</b>	<b>1.00</b>												
3 SES (wave 1 age 4/5)	<b>.02</b>	<b>-0.01</b>	<b>1.00</b>											
4 Picture Vocabulary Test (wave 1)	<b>.06</b>	<b>.06</b>	<b>.27</b>	<b>1.00</b>										
5 %girls in school	<b>.01</b>	<b>.13</b>	<b>.11</b>	<b>.03</b>	<b>1.00</b>									
6 Government school	<b>-0.01</b>	<b>-0.01</b>	<b>-0.17</b>	<b>-0.08</b>	<b>-0.19</b>	<b>1.00</b>								
7 NAPLAN numeracy (Year 3)	<b>.12</b>	<b>-0.04</b>	<b>.36</b>	<b>.36</b>	<b>.06</b>	<b>-0.08</b>	<b>1.00</b>							
8 NAPLAN numeracy (Year 5)	<b>.09</b>	<b>-0.07</b>	<b>.35</b>	<b>.28</b>	<b>.03</b>	<b>-0.05</b>	<b>.73</b>	<b>1.00</b>						
9 NAPLAN numeracy (Year 7)	<b>.05</b>	<b>-0.08</b>	<b>.40</b>	<b>.31</b>	<b>.05</b>	<b>-0.05</b>	<b>.74</b>	<b>.81</b>	<b>1.00</b>					
10 NAPLAN numeracy (Year 9)	<b>.04</b>	<b>-0.07</b>	<b>.43</b>	<b>.28</b>	<b>.05</b>	<b>-0.07</b>	<b>.71</b>	<b>.79</b>	<b>.88</b>	<b>1.00</b>				
11 NAPLAN literacy (Year 3)	<b>.12</b>	<b>.14</b>	<b>.41</b>	<b>.40</b>	<b>.11</b>	<b>-0.11</b>	<b>.80</b>	<b>.73</b>	<b>.73</b>	<b>.70</b>	<b>1.00</b>			
12 NAPLAN literacy (Year 5)	<b>.09</b>	<b>.18</b>	<b>.42</b>	<b>.39</b>	<b>.10</b>	<b>-0.10</b>	<b>.75</b>	<b>.77</b>	<b>.76</b>	<b>.74</b>	<b>.95</b>	<b>1.00</b>		
13 NAPLAN literacy (Year 7)	<b>.06</b>	<b>.18</b>	<b>.45</b>	<b>.39</b>	<b>.11</b>	<b>-0.10</b>	<b>.73</b>	<b>.72</b>	<b>.79</b>	<b>.77</b>	<b>.91</b>	<b>.96</b>	<b>1.00</b>	
14 NAPLAN literacy (Year 9)	<b>.04</b>	<b>.19</b>	<b>.46</b>	<b>.38</b>	<b>.11</b>	<b>-0.11</b>	<b>.71</b>	<b>.70</b>	<b>.77</b>	<b>.79</b>	<b>.88</b>	<b>.94</b>	<b>.97</b>	<b>1.00</b>

**Note:** Values in bold are statistically significant at  $p < 0.05$ ; values in bold and italics are statistically significant at  $p < 0.01$ . Non-significant correlations are shaded in gray. Relative age = student age minus state-average age. SES = socioeconomic status.



Table 3. *Decomposition of Effects Relative Age to Numeracy and Literacy Achievement in Years 3 – 9: Total, Indirect, and Direct Effects and Associated Effect sizes*

Effects	Model 1				Model 2				Model 3				Model 4			
	Est	SE	ES1	ES2	Est	SE	ES1	ES2	Est	SE	ES1	ES2	Est	SE	ES1	ES2
<b>Relative Age to Year 3 Numeracy</b>																
Total Effects	.28	.06	.12	.38	.22	.04	.09	.29	.22	.04	.09	.29	.22	.04	.09	.29
Total Indirect																
Direct	.28	.06	.12	.38	.22	.04	.09	.29	.22	.04	.09	.29	.22	.04	.09	.29
<b>Relative Age to Year 5 Numeracy</b>																
Total Effects	.20	.05	.09	.29	.16	.04	.07	.22	.17	.03	.07	.22	.16	.04	.07	.22
Total Indirect									.17	.03	.07	.22	.19	.04	.08	.25
Direct	.20	.05	.09	.29	.16	.04	.07	.22					-.03	.03	-.01	-.03
<b>Relative Age to Year 7 Numeracy</b>																
Total Effects	.12	.06	.05	.16	.07	.04	.03	.10	.16	.03	.07	.22	.07	.04	.03	.10
Total Indirect									.16	.03	.07	.22	.15	.04	.06	.19
Direct	.12	.06	.05	.16	.07	.04	.03	.10					-.08	.02	-.04	-.12
<b>Relative Age to Year 9 Numeracy</b>																
Total Effects	.09	.05	.04	.13	.04	.04	.02	.07	.15	.03	.07	.22	.04	.04	.02	.07
Total Indirect									.15	.03	.07	.22	.08	.04	.03	.10
Direct	.09	.05	.04	.13	.04	.04	.02	.07					-.04	.03	-.02	-.07
<b>Relative Age to Year 3 Literacy</b>																
Total Effects	.26	.05	.12	.76	.22	.04	.10	.32	.19	.04	.09	.29	.22	.04	.10	.32
Total Indirect																
Direct	.26	.05	.12	.76	.22	.04	.10	.32	.19	.04	.09	.29	.22	.04	.10	.32
<b>Relative Age to Year 5 Literacy</b>																
Total Effects	.17	.05	.09	.29	.14	.04	.07	.22	.16	.04	.08	.25	.14	.04	.07	.22
Total Indirect									.16	.04	.08	.25	.19	.04	.09	.29
Direct	.17	.05	.09	.29	.14	.04	.07	.22					-.05	.02	-.03	-.09
<b>Relative Age to Year 7 Literacy</b>																
Total Effects	.12	.05	.06	.38	.08	.04	.04	.13	.15	.03	.08	.25	.08	.04	.04	.13
Total Indirect									.15	.03	.08	.25	.13	.04	.07	.22
Direct	.12	.05	.06	.38	.08	.04	.04	.13					-.05	.02	-.02	-.07
<b>Relative Age to Year 9 Literacy</b>																

Total Effects	.09	.05	.04	.13	.06	.04	.03	.10	.15	.03	.08	.25	.06	.04	.03	.10
Total Indirect									.15	.03	.08	.25	.08	.04	.04	.07
Direct	.09	.05	.04	.13	.06	.04	.03	.10					-.02	.02	-.01	-.03

**Note.** Est = estimates. SE = standard error. Non-significant estimates are shaded in gray. ES1 = effect size in relation to SD of age (.317 years). ES2 = effect size in relation to students one year older than average and one-year younger than average (a difference of two years). The total effects of relative age to achievement in each year are decomposed into total indirect effect and direct effects RAE in each of four models (see Figure 2): Model 1, Relative Age(RAE)+ No Covariates(Cov)+No prior achievement (ACH); Model 2, RAE+ Cov-All lags+No Ach; Model 3, RAE +Cov-lag1+Ach-lag1; Model 4, RAE +Cov-all lags+Ach-all lags.

Table 4. *Prediction of Literacy and Numeracy in Years 3, 5, 7, and 9: Selected Parameter Estimates from Model 4.*

Numeracy Predictions				Literacy Predictions			
	Est	SE	ES		Est	SE	ES
<b>Year 3 Numeracy ON</b>				<b>Year 3 Literacy ON</b>			
Relative Age	<b>.22</b>	<b>.04</b>	<b>.09</b>	Relative Age	<b>.22</b>	<b>.04</b>	<b>.10</b>
Female	<b>-.09</b>	<b>.02</b>	<b>-.06</b>	Female	<b>.17</b>	<b>.02</b>	<b>.12</b>
Socioeconomic Status	<b>.2</b>	<b>.01</b>	<b>.27</b>	Socioeconomic Status	<b>.22</b>	<b>.01</b>	<b>.32</b>
Picture Vocab	<b>.21</b>	<b>.01</b>	<b>.28</b>	Picture Vocab	<b>.2</b>	<b>.01</b>	<b>.3</b>
School: %Female	<b>.02</b>	<b>.01</b>	<b>.03</b>	School: %Female	<b>.03</b>	<b>.01</b>	<b>.05</b>
School: Government	<b>-.02</b>	<b>.03</b>	<b>-.01</b>	School: Government	<b>-.03</b>	<b>.02</b>	<b>-.02</b>
<b>Year 5 Numeracy ON</b>				<b>Year 7 Literacy ON</b>			
Year3 Literacy	<b>.53</b>	<b>.03</b>	<b>.49</b>	Year3 Literacy	<b>.85</b>	<b>.02</b>	<b>.92</b>
Year3 Numeracy	<b>.33</b>	<b>.02</b>	<b>.33</b>	Year3 Numeracy	<b>.01</b>	<b>.02</b>	<b>.01</b>
Relative Age	<b>-.03</b>	<b>.03</b>	<b>-.01</b>	Relative Age	<b>-.05</b>	<b>.02</b>	<b>-.03</b>
Female	<b>-.18</b>	<b>.02</b>	<b>-.12</b>	Female	<b>.06</b>	<b>.01</b>	<b>.05</b>
Socioeconomic Status	<b>.04</b>	<b>.01</b>	<b>.05</b>	Socioeconomic Status	<b>.02</b>	<b>.01</b>	<b>.04</b>
Picture Vocab	<b>-.03</b>	<b>.01</b>	<b>-.04</b>	Picture Vocab	<b>0</b>	<b>.01</b>	<b>.01</b>
School: %Female	<b>-.02</b>	<b>.01</b>	<b>-.03</b>	School: %Female	<b>-.01</b>	<b>.01</b>	<b>-.01</b>
School: Government	<b>.04</b>	<b>.02</b>	<b>.03</b>	School: Government	<b>.01</b>	<b>.02</b>	<b>.01</b>
<b>Year 7 Numeracy ON</b>				<b>Year 7 Literacy ON</b>			
Year5 Literacy	<b>.43</b>	<b>.08</b>	<b>.36</b>	Year5 Literacy	<b>.91</b>	<b>.06</b>	<b>.92</b>
Year5 Numeracy	<b>.45</b>	<b>.02</b>	<b>.44</b>	Year5 Numeracy	<b>-.04</b>	<b>.01</b>	<b>-.04</b>
Year3 Literacy	<b>-.14</b>	<b>.08</b>	<b>-.13</b>	Year3 Literacy	<b>.02</b>	<b>.05</b>	<b>.03</b>
Year3 Numeracy	<b>.24</b>	<b>.02</b>	<b>.24</b>	Year3 Numeracy	<b>.03</b>	<b>.02</b>	<b>.04</b>
Relative Age	<b>-.08</b>	<b>.02</b>	<b>-.04</b>	Relative Age	<b>-.05</b>	<b>.02</b>	<b>-.02</b>
Female	<b>-.13</b>	<b>.02</b>	<b>-.08</b>	Female	<b>.01</b>	<b>.01</b>	<b>.01</b>
Socioeconomic Status	<b>.05</b>	<b>.01</b>	<b>.07</b>	Socioeconomic Status	<b>.03</b>	<b>.01</b>	<b>.05</b>
Picture Vocab	<b>0</b>	<b>.01</b>	<b>0</b>	Picture Vocab	<b>.01</b>	<b>.01</b>	<b>.01</b>
School: %Female	<b>.01</b>	<b>.01</b>	<b>.02</b>	School: %Female	<b>.01</b>	<b>.01</b>	<b>.01</b>
School: Government	<b>.05</b>	<b>.02</b>	<b>.03</b>	School: Government	<b>.01</b>	<b>.01</b>	<b>.01</b>
<b>Year 9 Numeracy ON</b>				<b>Year 9 Literacy ON</b>			
Year7 Literacy	<b>.42</b>	<b>.08</b>	<b>.34</b>	Year7 Literacy	<b>.42</b>	<b>.08</b>	<b>.34</b>
Year7 Numeracy	<b>.54</b>	<b>.03</b>	<b>.54</b>	Year7 Numeracy	<b>.05</b>	<b>.02</b>	<b>.06</b>
Year5 Literacy	<b>-.19</b>	<b>.11</b>	<b>-.16</b>	Year5 Literacy	<b>-.19</b>	<b>.11</b>	<b>-.16</b>
Year5 Numeracy	<b>.21</b>	<b>.02</b>	<b>.2</b>	Year5 Numeracy	<b>-.05</b>	<b>.02</b>	<b>-.06</b>
Year3 Literacy	<b>-.07</b>	<b>.06</b>	<b>-.07</b>	Year3 Literacy	<b>-.07</b>	<b>.06</b>	<b>-.07</b>
Year3 Numeracy	<b>.08</b>	<b>.02</b>	<b>.08</b>	Year3 Numeracy	<b>.02</b>	<b>.02</b>	<b>.02</b>
Relative Age	<b>-.04</b>	<b>.03</b>	<b>-.02</b>	Relative Age	<b>-.02</b>	<b>.02</b>	<b>-.01</b>
Female	<b>-.04</b>	<b>.02</b>	<b>-.03</b>	Female	<b>.02</b>	<b>.01</b>	<b>.02</b>
Socioeconomic Status	<b>.05</b>	<b>.01</b>	<b>.06</b>	Socioeconomic Status	<b>.02</b>	<b>.01</b>	<b>.04</b>
Picture Vocab	<b>-.03</b>	<b>.01</b>	<b>-.03</b>	Picture Vocab	<b>0</b>	<b>.01</b>	<b>-.01</b>
School: %Female	<b>0</b>	<b>.01</b>	<b>-.01</b>	School: %Female	<b>0</b>	<b>.01</b>	<b>0</b>

**Note:** Est = unstandardized beta estimate; SE = standard error; ES = effect size (standardized beta estimate). Parameter estimates from Model 4 (Figure 2, Table 5). Parameter estimates in bold are statistically significant ( $p < .05$ ) and those shaded in gray are non-significant.

Table 5. *Interactions Between Relative Age Effects (RAE) Five Covariates on NAPLAN Test Scores*

Achievement Outcomes Interaction Effects	Numeracy				Literacy			
	Separately		Combined		Separately		Combined	
	Est	SE	Est	SE	Est	SE	Est	SE
Year 3 ON								
RAE x Female	-.057	.092	-.059	.057	-.016	.075	-.021	.075
RAE x SES	.012	.042	.007	.027	.021	.037	.02	.04
RAE x Picture Vocab	.003	.006	.003	.004	.006	.005	.007	.005
RAE x %Females	.001	.006	.002	.004	.000	.006	.001	.006
RAE x Gov School Type	.001	.080	.022	.019	.080	.075	.101	.078
Year 5 ON								
RAE x Female	.042	.062	.057	.061	.012	.037	.017	.038
RAE x SES	.014	.028	.027	.028	.012	.021	.012	.021
RAE x Picture Vocab	-.003	.005	-.004	.005	.001	.003	.000	.002
RAE x %Females	-.003	.004	-.004	.004	-.002	.002	-.002	.002
RAE x Gov School Type	.027	.054	.019	.053	-.021	.040	-.025	.041
Year 7 ON								
RAE x Female	-.064	.048	-.072	.05	-.005	.035	-.004	.034
RAE x SES	-.003	.026	-.012	.029	-.004	.015	-.006	.016
RAE x Picture Vocab	.004	.004	.005	.004	.000	.002	.000	.003
RAE x %Females	.001	.003	.002	.003	.000	.002	.000	.002
RAE x Gov School Type	.008	.050	.014	.051	-.008	.037	-.011	.039
Year 9 ON								
RAE x Female	.013	.045	.005	.048	-.023	.029	-.023	.031
RAE x SES	.011	.023	.007	.024	-.008	.016	-.008	.017
RAE x Picture Vocab	.002	.004	.002	.004	-.002	.002	-.001	.002
RAE x %Females	.003	.003	.003	.003	.000	.002	.000	.002
RAE x Gov School Type	-.002	.044	.015	.048	-.053	.016	-.018	.035

**Note.** Est = estimates. SE = standard error. SES = socioeconomic status. Based on Model 4 (Figure 2), interaction effects between relative age effects (RAEs) and each of the five covariates was added to the model, predicting effects in relation to each of the eight NAPLAN tests scores (Literacy and numeracy in Years 3, 5, 7 and 9). Five separate analyses were conducted for each interaction effect separately (Separate) and for the combined set of five interactions in a single model (combined).

## Supplemental Materials

### Supplemental Section 1

#### Tests of Invariance of the Literacy Factor

- Supplemental Table 1(SM1)  
Goodness of Fit Statistics in Support of the Quasi-Scalar Model of Invariance over time for the Literacy Factor
- Supplemental Table 2(SM2)  
Factor Loadings and Intercept for the Quasi-Scalar Model of Invariance.

### Supplemental Section 2

#### Relative Age Effects (RAEs) in Years 3, 5, 7, 9 in numeracy and literacy – A growth model

- Supplemental Figure 2 (SM3)  
Growth Curves as a Function of Relative Age Effects: Growth in Numeracy and Literacy in Years 3, 5, 7 and 9
- Supplemental Table 3 (SM4).  
*Marginal Means From Growth Model (Figure 2): Parameter Estimates for the Growth Models: Growth in Numeracy and Literacy in Years 3, 5, 7 and 9.*

# RELATIVE AGE EFFECT AND ACADEMIC ACHIEVEMENT

## Supplemental Section 1

### Tests of Invariance of the Literacy Factor

In longitudinal models, the invariance of the constructs measured at different time points is an important issue. In the present investigation, this refers to the invariance of the indicators of achievement over time. There are two different aspects of invariance over time. For the numeracy test and the four subtests of the verbal component, the invariance over time is based on the NAPLAN tests' construction for Australian students' entire population in each age cohort. Academic achievement was based on a standardized test, the National Assessment Program Literacy and Numeracy (NAPLAN), provided by the Australian Curriculum, Assessment and Reporting Authority (ACARA). NAPLAN is an Australian-wide, standardized testing program completed each year by students in Years 3, 5, 7, and 9. Tests in numeracy and literacy (reading, spelling, grammar, and writing) are aligned to the Australian National Curriculum, and scores are all highly reliable (.86 - .96, ACARA, 2014).

In terms of the four components of literacy, there is a second aspect of invariance over time in terms of the relative contribution of each score to the overall literacy factor. The test construction by ACARA guarantees that each of the components of literacy is measured along a common metric that is fully invariant over time and cohorts. However, this still leaves the less critical question of whether the four components' relative contributions to the latent literacy factor are the same in Years 3, 5, 7 and 9. A reasonable approach might be to take a simple average of the scores each year. However, this would assume that each component contributed equally and that the relative contribution was the same from one year to the next. Hence, we took a more sophisticated approach that provided a test of this assumption. In this sense, the latent reading factor is an empirically weighted average of the individual reading tests that is corrected for measurement error. Each component's relevant contribution is indexed by factor loadings relating each indicator to the latent literacy factor. The test of the invariance of the factor loadings over time tests whether the relative contribution remains constant over time. The results (see Table SM1 in Supplemental Materials) provides good support for this aspect of invariance.

We also tested for the invariance of the factor intercepts. Here, there was some evidence that the grammar component's intercept component was not completely invariant (see Table SM1). To account for this lack of complete invariance, we allowed the grammar subtest intercepts to vary over time. Hence, the four components' relative contributions were invariant over time, and the intercepts of three of the four components were invariant over time. In the final quasi-invariant model, there is complete invariance for factor loadings (the relative contribution of each literacy component to the overall literacy factor) and the intercepts for reading, writing, and spelling components literacy. However, we allowed the grammar indicator intercepts to vary over Years 3, 5, 7, and 9. Inspection of the parameter estimates based on this quasi-invariant model (see Table SM2) shows that the intercept for grammar are similar to the other three components in primary school (Years 3 and 5), but is somewhat lower in secondary school (Years 7 and 9). This suggests that growth in grammar is somewhat smaller than for the other components of literacy.

In our study, the primary analysis is SEMs based on the regression approach. Importantly, there is no assumption of invariance over time of the factor structure underlying the test scores for the regression approach. Thus, controlling the RAEs on NAPLAN achievement in Year 5 by NAPLAN achievement in Year 3 does not require that the two tests are invariant. Of course, support for the factor structure's invariance – particularly the factor loadings - contributes to interpreting the results, but is not a necessary assumption of the statistical analyses. Indeed, because the regression approach is based on latent covariances in structural equation models, the intercepts' invariance is of little consequence (but also see subsequent discussion of a growth model approach in Supplemental Materials, section 2).

## RELATIVE AGE EFFECT AND ACADEMIC ACHIEVEMENT

### Supplemental Table 1(SM1)

Goodness of Fit Statistics in Support of the Quasi-Scalar Model of Invariance over time for the Literacy Factor

Invariance Model	Chi-SQ	Df	Free	RMSEA	CFI	TLI
Configural	829	74	78	0.045	0.989	0.982
Metric	1177	83	69	0.051	0.984	0.977
Scalar	2078	92	60	0.066	0.971	0.963
Quasi-scalar	1398	89	63	0.054	0.981	0.975

*Note.* Free = number of freely estimated parameters; **Chi-SQ** = chi-square; df = degrees of freedom ratio; RMSEA = Root Mean Square Error of Approximation CFI = Comparative fit index; TLI = Tucker-Lewis Index. The goodness of fit of these models was evaluated by a range of recommended indices including the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), and the Comparative fit indexes (CFI; Bentler, 1990). CFIs and TLIs greater than .95 indicate an acceptable model fit, whereas the RMSEAs less than .06 indicate good fit. When comparing a set of increasingly restrictive models, a more restrictive model is preferred if the change in model fit indices is not inferior to those of the less restrictive model. In terms of the RMSEA, the change should be less than .015 (Chen, 2007). For CFI, the change should be less than .01 (Chen, 2007; Cheung & Rensvold, 2001).

## RELATIVE AGE EFFECT AND ACADEMIC ACHIEVEMENT

Supplemental Table 2(SM2)

Factor Loadings and Intercept for the Quasi-Scalar Model of Invariance.

Factors	Factor Loadings			Intercepts		
	Unstand.		Stand.	Unstand.		Stand.
Indicators	Estimate	SE	Estimate	Estimate	SE	Estimate
<b>Y3Literacy BY</b>						
Y3READ	1.00	.00	.81	4.15	.02	4.97
Y3WRITE	.82	.01	.77	4.19	.02	5.76
Y3Spelling	1.01	.01	.84	4.10	.02	5.03
Y3Grammar	1.14	.01	.86	4.19	.02	4.67
<b>Y5Literacy BY</b>						
Y5READ	1.00	.00	.80	4.15	.02	5.29
Y5WRITE	.82	.01	.73	4.19	.02	5.91
Y5Spelling	1.01	.01	.84	4.10	.02	5.41
Y5Grammar	1.14	.01	.83	4.15	.02	4.82
<b>Y7Literacy BY</b>						
Y7READ	1.00	.00	.85	4.15	.02	5.67
Y7WRITE	.82	.01	.69	4.19	.02	5.67
Y7Spelling	1.01	.01	.84	4.10	.02	5.53
Y7Grammar	1.14	.01	.88	3.96	.02	4.93
<b>Y9Literacy BY</b>						
Y9READ	1.00	.00	.87	4.15	.02	5.68
Y9WRITE	.82	.01	.62	4.19	.02	4.96
Y9Spelling	1.01	.01	.84	4.10	.02	5.34
Y9Grammar	1.14	.01	.87	3.82	.03	4.58

**Note:** Parameter estimates (factor loadings and intercepts) and standard errors (SE) for quasi-scalar model (see Supplemental Table SM1) of invariance over the four grades (Y3 = Year 3, Y5 = Year 5, Y7 = Year 7, Y9 = Year 9). Unstandardized factor loading is constrained to be equal across all four grade levels. Indicator intercepts are constrained to be equal across the four grade levels for reading, writing, and spelling subtests, but intercepts for Grammar are freely estimated.



## RELATIVE AGE EFFECT AND ACADEMIC ACHIEVEMENT

### Supplemental Section 2

#### RAE in Years 3, 5, 7, 9 in numeracy and literacy – A growth model

In addition to our main analyses, we also pursued growth models of the latent means. The growth curve model is parallel to full-forward regression model (Model 4 in Figure 2) in the main text. We used the same data and the same set of covariates and control variables. As in the full-forward model, there are direct effects of covariates and relative age on all subsequent achievement measures, and effects of each achievement measure on all subsequent achievement measures (e.g., the effects of Year 3 achievement on achievement in Years 7 and 9 as well as Year 5).

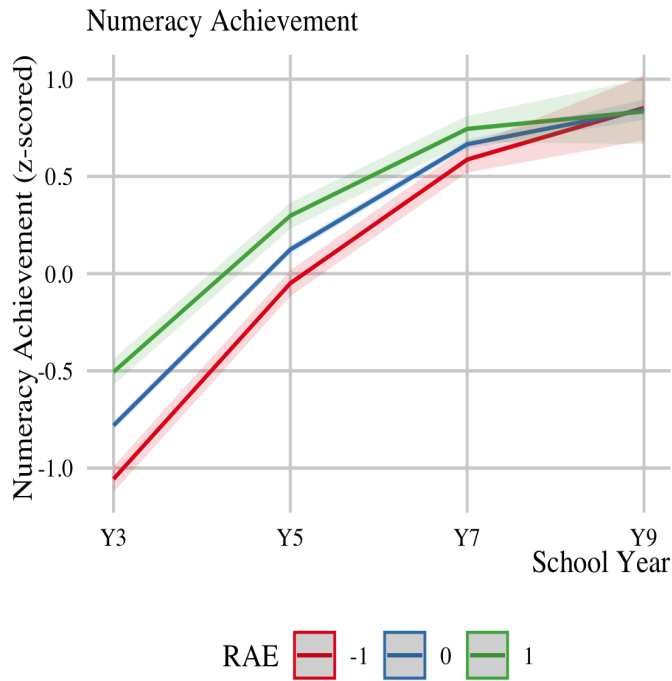
The growth models assume that the subtest scores are scaled along a common metric. Based on ACARA's construction of the test scores, there is good support for this assumption. However, because the growth curves depend on latent means, they depend on the factor structure's invariance over time. From this perspective, the relative lack of invariance for the intercept of the grammar component suggests that the grammar component's growth curve component differs slightly from that based on the other three components and the overall Literacy factor (see earlier discussion of invariance tests in Supplemental Materials, section 1).

Importantly, given that the intercepts are invariant for three of the four components, there is good support for the overall metric's invariance for the literacy measure. This approach to partial invariance is consistent with traditional (e.g., Byrne, et al., 1989) and current recommendations (e.g., van de Schoot et al., 2012; Putnick & Bornstein, 2016). Thus, Steenkamp and Baumgartner (1998) suggested that ideally, more than half of the factors' indicators should be invariant. Similarly, Vandenberg and Lance (2000) suggested that a factor can be considered partially invariant if most of the factor indicators are invariant. Van de Schoot et al. (2012) note that it is possible to make valid inferences about latent means as long as at least two intercepts are constrained to be equal. In this sense, the partial invariance approach is stronger than leaving out the Grammar test altogether. Also, we note that the lack of invariance is not in relation to the metric of the Grammar test, but only in relation to the relative contribution of the Grammar test to the metric underpinning the metric of the overall (latent) Literacy factor. This supports our interpretations of the effect of starting age (concerning invariance assumptions).

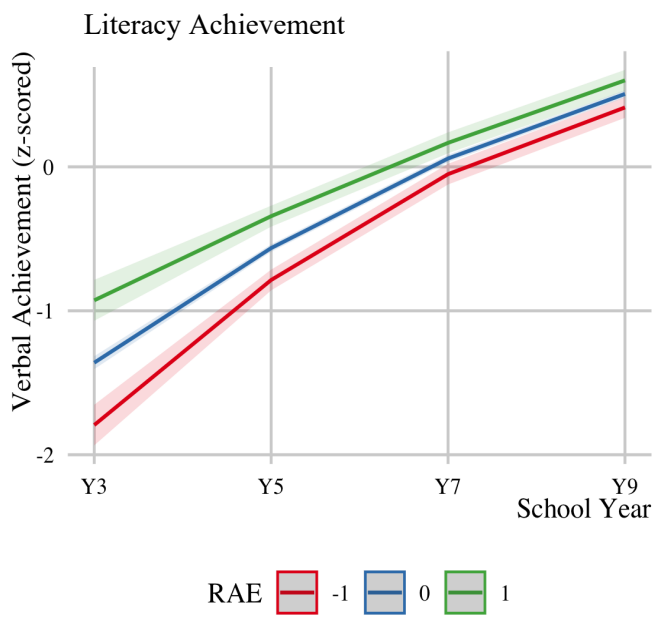
Based on results of the growth model (Supplemental Table 3) we plotted growth curves for students with relative age values of +1 (one year older than the mean age of students in the same state), 0 (the mean age of students in their state), and -1 (one year younger mean age of students in the same state). This graphic representation illustrates NAPLAN test scores (Year 3-Year 9) varying along a meaningful metric (age in years) common across all four waves. The growth curves visually provide a clear separation of achievement for relatively younger and relatively older primary schools (Supplemental Figure 1). The separation declines as students grow older and disappears completely for numeracy at Year 9. The marginal mean differences (Supplemental Table 3) support these visual differences in that differences between the groups are statistically significant for all but numeracy in Year 9.

We also note how the interpretation of the magnitude of the differences based on the growth models relates to earlier discussion (in the main text) of effect sizes. In the main text, we presented ES1 estimates (differences in relation to a 1 SD = .317 years) and ES2 (differences in relation to one year). There we argued that the ES2 metric was more appropriate in relation to the expected results of a random-control-trial intervention that held students back one year or for parents who want to know the likely effects of delaying the entry into school for their child for one year. Because our growth model is based on students who are one year older and one year younger than average, the underlying metric reflects a difference of two years rather than a difference of the one year as in ES2. For this reason, the separation of the effects for the oldest and youngest students is approximately twice the value of ES2s (see growth model results in Supplemental Table 3 compared to ES2 for total effects in Model 4, Table 3 in the main text). This new ES metric implicit in the growth curve models reflects the difference between the oldest and the youngest students (i.e., a difference of two years) or the difference between a student who was held back one year compared to a student who was accelerated one year (i.e., began school early or skipped a grade). Again, we emphasize that none of the different ES metrics are inherently superior, but are relevant to different questions.

### Growth Curve Model by RAE



### Growth Curve Model by RAE



Supplemental Figure 2 (SM3). Growth Curves as a Function of Relative Age Effects: Growth in Numeracy and Literacy in Years 3, 5, 7 and 9.

## RELATIVE AGE EFFECT AND ACADEMIC ACHIEVEMENT

### Supplemental Table 3 (SM4).

*Marginal Means From Growth Model (Figure 2): Parameter Estimates for the Growth Models: Growth in Numeracy and Literacy in Years 3, 5, 7 and 9*

Achievement		Relative Age (RA) in Years			Test of Significance	
		-1	0	+1	RA = -1 vs. RA = +1	
Literacy Year 3	Mn	-1.79	-1.36	-.93	-.87	**
	SE	.07	.02	.07	.10	
Literacy Year 5	Mn	-.79	-.56	-.34	-.44	**
	SE	.04	.01	.04	.05	
Literacy Year 7	Mn	-.05	.06	.17	-.22	**
	SE	.04	.01	.04	.05	
Literacy Year 9	Mn	.41	.51	.60	-.19	**
	SE	.04	.01	.04	.05	
Numeracy Year 3	Mn	-1.06	-.78	-.51	-.55	**
	SE	.03	.01	.03	.05	
Numeracy Year 5	Mn	-.05	.12	.30	-.35	**
	SE	.03	.01	.03	.05	
Numeracy Year 7	Mn	.59	.67	.74	-.16	**
	SE	.03	.01	.03	.05	
Numeracy Year 9	Mn	.85	.84	.83	.02	NS
	SE	.09	.03	.09	.12	

**Note:** Based on the Growth Model (Figure 2) Marginal means (Mn) and standard errors (SEs) were estimated for the eight achievement scores. Achievement scores are standardized (Mn = 0, SD = 1) to a common metric that is common across all years. Achievement scores are estimated for students at relative age (actual age minus the cohort average age) of -1 year, mean, and +1 year. The test of statistical significance is the difference between student one year older and one year younger than their cohort (see Figure 2 SM3).

RELATIVE AGE EFFECT AND ACADEMIC ACHIEVEMENT