The Role of the Socio-economic Compositional Effect in Systemic School Effectiveness and Equity

Michael G. Sciffer (Corresponding Author)

School of Education, Murdoch University, 90 South Street, Murdoch, Western Australia 6150.

ORCID: 0000-0003-2552-8840

research@sciffer.info

Laura B. Perry

Associate Professor Education Policy and Comparative Education, School of Education, Murdoch University.

ORCID: 0000-0003-4252-2379

Andrew McConney

Honorary Research Fellow, School of Education, Murdoch University.

ORCID: 0000-0001-7618-8829

Abstract

This study examines the role of the socio-economic compositional effect in predicting the systemic performance and equity of schools. We build on prior research by comparing the size and linearity of compositional effects in primary and secondary schools and testing whether students differ in sensitivity to school composition, while addressing critiques of potential biases from measurement error and selection effects. A set of residualised-change and cross-sectional models were tested for selection bias, linearity, differential sensitivity and substantiveness of the compositional effect. We found that the relationship between composition and achievement growth was larger in secondary than primary schools while addressing potential biases from measurement error and selection effects. The socio-economic compositional effect was non-linear in several academic domains, reducing as average school SES increased. Secondary students differed in sensitivity to composition by family socio-economic status, with sensitivity reducing as family socio-economic status increased. We found that socioeconomic composition was a substantive effect, both as a comparative predictor of academic achievement and explaining school-level variance in achievement growth. We conclude that the school socio-economic compositional effect substantively influences systemic performance and equity. We argue school composition should be a leading focus of policy reform as reducing school segregation would improve systemic school performance and equity.

The predictive relationship between student socio-economic status (SES) and academic outcomes has been well established in educational research. Sirin's (2005) metanalysis of the relationship between SES and academic achievement found a mean effect size of $r = .29$. National and international testing programs have consistently found SES achievement gaps. Australia's National Assessment Program – Literacy and Numeracy (NAPLAN) 2018 administration found that in all assessed academic domains across each year group, higher levels of parental education and occupation were associated with higher levels of academic achievement (ACARA, 2018). The OECD's Programme for International Student Assessment (PISA) most recent administration found that on average across OECD countries SES accounted for 12% of the variation in reading achievement (OECD, 2019).

Less well established in the literature are the effects of school socio-economic composition (SEC). SEC effects are the predictive relationships between the aggregate socio-economic status of a student body and individual-level student outcomes (Willms, 2010). Student bodies may consist of classes, cohorts, age groups or schools. SEC effects are conceptualised as the difference in outcomes of students who have the same individual SES associated with attending schools with differing socio-economic compositions (Raudenbush & Bryk, 2002). That is, average school SES has an effect separate from individual student SES.

Prior research has found that SEC has a substantial predictive relationship with academic achievement (Willms, 2010) and achievement growth (Rumberger & Palardy, 2005), but many gaps remain in the literature. Few studies have compared the SEC effect in primary and secondary schools. Research findings into the linearity or consistency in the strength of the SEC effect have diverged (Benito, Alegre & González-Ballethò, 2014; Chiu & Khoo, 2005; Rumberger & Palardy, 2005). The little research that has considered differential sensitivity to the SEC effect has found it to be consistent across individual-level SES (Rumberger & Palardy, 2005; Willms, 1986). Methodological questions also remain as to whether measurement error and selection effects have biased past findings of SEC effects or if SEC effects are of a size to warrant policy interest (Armor et al., 2018; Lauen & Gaddis, 2013; Marks, 2015).

The present study addresses some of these gaps in our understanding of the SEC effect with six research questions:

- (1) What is the relationship between SEC and academic achievement growth when accounting for measurement error?
- (2) Does the effect of SEC on academic achievement growth differ in size between primary and secondary schools?
- (3) Do school selection effects bias relationships between SEC and achievement growth?
- (4) To what extent is the SEC effect non-linear?
- (5) Does sensitivity to SEC vary by SES?
- (6) Is the SEC effect of substantive interest?
 - a. How much of the variance between schools in academic achievement growth is due to SEC?
 - b. How does SEC effect compare to other effects associated with academic achievement?

These research questions aim to locate the potential role of the socio-economic compositional effect in predicting the effectiveness of school systems for all students. Testing for the veracity of the SEC effect by accounting for potential measurement errors and selection effects indicates whether school effectiveness research should be informed by compositional effects, including in the development of value-added measures (Marks, 2020). Exploring the relative strength of the SEC effect across school levels, and compared to other predictors of achievement, suggests the priority it should be given in school policy reforms. Testing for the non-linearity of the SEC effect shows whether it predicts the average performance of school

systems while testing for sensitivity indicates whether the learning of particular students is advantaged or disadvantaged by the SEC effect.

Background

School compositional effects came to prominence in the educational literature following the seminal Coleman Report (Coleman et al., 1966). David Coleman and colleagues investigated the quality of US public education afforded to minority ethnic groups compared to European-American students through relating school characteristics to academic achievement (Coleman et al., 1996). Based on a sample of 568,743 students from 3157 schools, they found that school composition was the largest school-level effect on academic performance. These effects have consistently been found over time across international schooling systems predicting a diverse range of educational outcomes.

International large-scale studies have consistently found that SEC is a substantial predictor of student academic achievement. Each cycle of PISA has found that school-level socio-economic factors substantively predict academic achievement separate from individual-level SES in most participating countries (OECD, 2003, 2004, 2007, 2010, 2013, 2016, 2019). Secondary analyses of PISA datasets have shown an inverse correlation between the degree of socio-economic school segregation and national levels of academic achievement (Chiu & Khoo, 2005; Willms, 2010).

SEC has been shown to predict a range of schooling outcomes. Both academic achievement (Lamb & Fullarton, 2002) and achievement growth (Rumberger & Palardy, 2005) across a broad range of subject domains are related to SEC. Academic achievement is predicted by SEC at both primary (Sofroniou et al., 2004) and secondary school levels (Chesters & Daly, 2017). Other school outcomes that have been associated with SEC have been high school graduation and college attendance (Palardy, 2013) and college choice (Palardy, 2015).

Reported effect sizes for SEC vary widely in the literature. Van Ewijk and Sleegers' (2010) meta-analysis of cross-sectional studies found effect sizes ranged 0.03 to 0.59 standard deviations with an average of 0.32. The 2015 PISA administration reported that on average, SEC accounted for 62.6% of variation in academic achievement due to schools (OECD, 2016). Longitudinal studies have tended to find smaller effect sizes for SEC as they control for selection effects where prior achievement is positively correlated with SEC. That is, students with initial higher achievement tend to self-select into schools with higher SEC. Rumberger and Palardy's (2005) analysis of the National Education Longitudinal Survey of 1988 found SEC effect sizes ranged 0.05 to 0.21 standard deviations depending on the academic domain. Marks' (2015) residualised-change model with primary school students found smaller SEC effects ranging from 0.00 to 0.05 standard deviations.

Evidence for differential sensitivity to SEC effects has been inconsistent. Rumberger and Palardy (2005) found that disadvantaged African-American students gained more benefit from attending high SES schools than European-American students but Palardy (2013) found that the school socio-economic compositional effect on graduation and college enrolment was consistent across differing ethnic and socio-economic groups. McPherson and Willms (1987) found that low-SES students may have benefited the most from reductions in school segregation through the comprehensive reorganisation of the Scottish school system while Perry and McConney (2010) found that the SEC effect was relatively consistent across all SES groups.

The linearity of the SEC effect influences its relationship with average school system performance. If the SEC effect is linear, it would not diminish average systemic performance as the negative effects of low SEC schools would be offset by high SEC schools, assuming a normal distribution of school socio-economic composition. But if the SEC effect is non-linear, it may predict higher or lower average systemic performance. In a convex scenario, where the

slope of the SEC effect increases at a growing rate as school SEC increases, the SEC effect would predict higher systemic performance as the academic gains to high SEC schools outweighed the losses to low SEC schools. Alternatively, in the concave scenario, where the slope of the SEC effect increases at a diminishing rate as school SEC increases, the SEC effect would predict lower systemic performance as the academic losses to low SEC schools outweighed the gains to high SEC schools. Cross-national research with PISA 2000 (Chiu & Khoo, 2005) and PISA 2006 (Willms, 2010) have suggested a concave relationship. Simulation research (Benito, Alegre & González-Balletbò, 2014) with PISA 2009 found either convex or concave SEC effects, varying by country.

A limitation of school compositional research has been the potential for selection effects (Lauen & Gaddis, 2013) and measurement error (Marks, 2015) to upwardly bias compositional effects. Selection bias occurs when unmeasured systematic differences between groups confound the relationship between an independent variable and a dependent variable. For example, families with high educational goals for their children may seek to enrol them in schools with good reputations. Due to demand constraints, it is more likely such schools will have above average SEC, such as state schools in expensive suburbs or high-fee private schools. Thus, SEC may act as a proxy for unmeasured factors that explain higher achievement in higher SEC schools. Controlling for prior achievement may partially address selection effects by accounting for the clustering of higher ability, or more motivated, students into higher SEC schools (Hallberg et al., 2018). While longitudinal studies account for biases in prior achievement, achievement growth may also be biased by selection effects as growth rates may differ according to selection effects. One method to mitigate selection bias is to include potential variables that may capture the probability of school selection (Palardy, 2013; Rangvid, 2007). Controlling for school sector in Australian samples likely captures much of the potential selection effect because of the high degree of school choice in Australia (ABS, 2017).

Measurement errors may have upwardly biased prior school compositional research (Harker & Tymms, 2004; Marks, 2015). Measurement error is the difference between the true score of a construct and that captured by its measure. Pokropek (2015) and Televantou et al. (2015) have shown that variables derived from confirmatory factor analysis can address measurement error bias in compositional studies. Variables derived from confirmatory factor analysis account for measurement error as they are derived from only the variance shared by the underlying indicators (Brown & Moore 2012, 361). Very few compositional studies have utilised factor analysis to address measurement errors (Marsh et al., 2009), likely due to the requirement for at least three indicators in single factor models.

Method

Participants

The research questions were answered through secondary analysis of two large-scale assessments. The first was the 2017 student-level deidentified NAPLAN dataset. NAPLAN is Australia's annual population assessment of academic achievement in reading, writing, spelling, grammar and numeracy for students in grades three (313,807 students from 7820 schools), five (311,412 students from 7809 schools), seven (288,946 students from 3422 schools) and nine (281,280 students from 3064 schools). We utilised the 2017 grade five and nine scores and their respective prior grade three and seven scores measured in 2015. NAPLAN is administered by the Australian Curriculum, Assessment and Reporting Authority (ACARA). The NAPLAN dataset allows for the examination of the SEC effect on achievement growth in primary and secondary schools. As a population dataset, NAPLAN also enables the use of descriptive statistical methods to make generalisable comparisons among differing demographic groups.

The second dataset was the 2015 PISA dataset. PISA is the OECD's international triennial assessment of 15-year-olds' reading, mathematical and scientific literacies. Each PISA administration assesses the three academic domains with a focus on scientific literacy in 2015. Approximately 540,000 students from 72 countries and economies participated in PISA 2015 (OECD, 2016). The Australian-specific dataset hosted on the Australian Council for Educational Research's (ACER) website was utilised as it contains additional demographic variables not available in the OECD's international dataset, consisting of 14,530 students from 758 schools. The publicly available PISA datasets provide a rich set of demographic, contextual, systemic, attitudinal, and behavioural measures beyond academic achievement. School compositional research can utilise such variables to control for selection bias and measurement error and compare compositional effects to other constructs.

PISA utilises a two-stage stratified sampling procedure where schools are first sampled, then students within schools (OECD, 2016). The Australian school sample was stratified to be representative of state/territory, sector, grade, urbanisation, gender composition, socio-economic level and senior secondary education (OECD, 2017). The Australian student sample was a random selection of students born between 1 May 1999 and 30 April 2000, and all age-eligible Indigenous students (Thomson et al., 2017).

PISA applies survey weights to participant students and schools to adjust for non-equal survey selection due to oversampling of small subpopulations, inaccuracies in the planning of the assessment design, school-level non-responses without replacement, and non-random student-level non-responses (OECD, 2017). Survey weights adjust the contribution of each student and school in a model to match their known proportion in a population (Gonzalez, 2012). All reported inferential results in this study have taken student and school survey weights into account.

Measures

The dependent variables were reading, writing, spelling, grammar and punctuation, and numeracy in the NAPLAN dataset and scientific literacy in the PISA dataset. Reading and mathematical literacies were not used as dependent variables in our analysis of PISA 2015 as many of the predictors were specific to science teaching and learning. NAPLAN tests are designed to "broadly reflect aspects of literacy and numeracy within the curriculum in all jurisdictions" (ACARA, 2018, p. iv). The knowledge and skills assessed are drawn from the *Australian Curriculum: English* and *Mathematics* and the literacy and numeracy general capabilities of the Australian curriculum (ACARA, 2017). The Australian curriculum defines literacy as "the capacity to interpret and use language features, forms, conventions and text structures in imaginative, informative, and persuasive texts" and numeracy as "the knowledge and skills to use mathematics confidently across all learning areas at school and in their lives more broadly" (ACARA, 2017, pp. 6-7).

PISA implements a cognitively demanding definition of literacy as "students' capacity to apply knowledge and skills in key subjects, and to analyse, reason and communicate effectively as they identify, interpret and solve problems in a variety of situations" (OECD, 2016, p. 25). PISA 2015 calculated 10 plausible values of each student's performance in each literacy domain (OECD, 2017). Plausible values account for the planned missingness in international large-scale assessments where participants complete linked subsets of the total item set (Von Davier et al., 2009). They are multiple imputations of individual scores which allow for the estimation of latent population constructs (Laukaiyte & Wiberg, 2017; Wu, 2005). All reported results in this study have taken plausible values into account unless otherwise specified.

Confirmatory factor analysis was used to develop measures of SES in both datasets. The resulting latent measures address measurement error (Brown & Moore, 2012) and the

potentially attendant inflation of aggregated effects (Pokropek, 2015; Televantou et al., 2015). SEC was derived from SES through manifest aggregation in the NAPLAN dataset due to it sampling almost all students in each school. Latent aggregation was used with the PISA dataset due to the sample averaging 19.25% of 15-year-olds in each school (Lüdtke et al., 2008). *Tables S4 and S5* in the *Technical Appendix* provide item descriptors, factor loadings and fit indices of the latent constructs used in this study.

The analysis of the PISA dataset utilised structural equation modelling where a factor analysis generating the measures of SES and SEC was performed simultaneously with the modelling of the relationship of the variables to each other. The analysis of the NAPLAN dataset was a series of hierarchical regression models where SES was the factor scores derived from a prior confirmatory factor analysis. Structural equation modelling could not be used to derive SEC in the NAPLAN dataset as the indicators of SES were categorical variables and it does not make sense to derive group averages of categorical measures.

Other variables in this study were a range of predictors in the PISA dataset used for comparative analysis with SEC and a binary variable to distinguish public from private schools in the NAPLAN dataset. Descriptive statistics for all variables used in the NAPLAN and PISA datasets are in *Tables S2 and S3* of the technical appendix.

Procedure

Visual analyses of graphs of the unweighted PISA and NAPLAN datasets were performed to explore the relationship between SEC and academic achievement and detect whether SES may interact with SEC. The PISA graph was a random subset of the first plausible value of scientific literacy of 1500 students from Australian schools. Students and schools were categorised as low, middle or high according to their respective relationship with mean SES and SEC. Low was more than one standard deviation below the mean. High was more than one standard deviation above the mean. The remaining students and schools were categorised as middle. The NAPLAN graph was the proportion of students achieving minimum national benchmarks of band 6 in all academic domains in grade 9. It similarly categorised students and schools as low, middle and high SES and SEC respectively.

A set of multilevel residualised-change models were constructed to answer research questions 1-6a with the NAPLAN dataset. Residualised-change models are two-occasion growth models where the prior score on a dependent variable is included as a covariate of the dependent variable (Gollwitzer et al., 2014). When academic achievement is the dependent variable, the resulting coefficients of the predictor variables provide estimates of their influence on achievement growth. *Equation 1* represents the residualised-change model this study used to answer research questions 1, 2 and 6a.

$$Y_{i2j2} = \beta_0 + \beta_1 Y_{i1j1} + \beta_2 (X_{ij} - \bar{X}_{..}) + \beta_3 \bar{X}_{.j} + \delta_{0j} + \varepsilon_{ij} \quad (1)$$

For student i in school j , Y_{i2j2} was academic achievement at either grades five or nine, β_0 was the intercept, β_1 was the coefficient for prior achievement at either grades three or seven, β_2 was the coefficient for grand-mean centred student SES, β_3 was the coefficient for grand-mean centered school-average SES, δ_{0j} was the school-level residual variance and ε_{ij} was the student-level residual variance. Academic growth was modelled over two years of primary school, grades three to five, and two years of secondary school, grades seven to nine.

Research question 1 (What is the relationship between SEC and academic achievement growth when accounting for measurement error?) was tested through the detection of positive statistically significant coefficients of β_3 in equation 1 after having addressed measurement error in SES through confirmatory factor analysis. Research question 2 (Does the effect of SEC on academic achievement growth differ in size between primary and secondary schools?) was

answered through comparing the effect sizes of β_3 in the same academic domain in grades 3-5 with grades 7-9 in equation 1. Research question 6a (How much of the variance between schools in academic achievement growth is due to SEC?) was answered through calculating the reduction in school-level variance in equation 1 compared to *Equation 2*, being the null model for academic growth.

$$Y_{i2j2} = \beta_0 + \beta_1 Y_{i1j1} + \delta_{0j} + \varepsilon_{ij} \quad (2)$$

Research question 3 (Do school selection effects bias relationships between SEC and achievement growth?) was tested through a modification of equation 1 where a level-2 dichotomous control variable for school sector, being public or private, was added as a covariate to the model to detect its impact on SEC. The effect sizes of β_3 were compared in the models with, and without, the school sector variable to detect a reduction in the SEC effect with the inclusion of school sector. A reduction in the coefficient for SEC with the addition of the school-sector control variable would indicate potential selection bias due to school choice. The school sector variable would capture much of the variance in SEC due to school choice in Australia as 34.4% of school students were enrolled in private schools in 2017 (ABS, 2017).

Research question 4 (To what extent is the relationship between SEC and academic achievement growth non-linear?) was answered through a modification of equation 1 where a quadratic (squared) term for SEC was added to the model to test for non-linearity of SEC. A statistically significant quadratic coefficient would indicate non-linearity of the SEC effect whereas a non-significant coefficient would indicate a non-quadratic relationship. A positive quadratic coefficient would represent a convex non-linear effect whereas a negative quadratic coefficient would indicate a concave non-linear effect.

Research question 5 (Does sensitivity to SEC vary by SES?) was tested through the detection of a statistically significant interaction term in *Equation 3*. β_4 was the coefficient for the cross-level interaction between SES and SEC. The group-mean centring of student SES avoided the potential confounding of the interaction between SES and SEC by the between-level variance captured by raw level-one measures (Enders & Tofighi, 2007; Snijders & Bosker, 2012, pp. 83-84). A significant interaction effect of β_4 would indicate that the effect of SEC varies by SES, that is, some students are more sensitive to SEC than others.

$$Y_{i2j2} = \beta_0 + \beta_1 Y_{i1j1} + \beta_2 (X_{ij} - \bar{X}_{.j}) + \beta_3 \bar{X}_{.j} + \beta_4 (X_{ij} - \bar{X}_{.j}) \cdot \bar{X}_{.j} + \delta_{0j} + \varepsilon_{ij} \quad (3)$$

Research question 6b (How does SEC effect compare to other effects associated with academic achievement?) was answered through a multilevel model of the PISA dataset as represented in *Equation 4*. For student i in school j , β_0 was the intercept, β_1 was the coefficient for group-mean centred student SES, β_2 was the coefficient for grand-mean centred latently-aggregated SES (Lüdtke et al., 2008), P was a vector of coefficients of student-level comparative variables, S was a vector of coefficients of school-level comparative predictors, δ_{0j} was the school-level residual variance and ε_{ij} was the student-level residual variance. The SEC effect was derived from *Equation 5* (Raudenbush & Bryk, 2002, p. 139) where β_c was the coefficient for SEC. Research question 6b was answered through a comparison of the effect size of SEC with the other effects in equation 4.

$$Y_{ij} = \beta_0 + \beta_1 (X_{ij} - \bar{X}_{.j}) + \beta_2 \bar{X}_{.j} + P(X_{ij}) + S(\bar{X}_{.j}) + \delta_{0j} + \varepsilon_{ij} \quad (4)$$

$$\beta_c = \beta_{\bar{X}_{.j}} - \beta_{(X_{ij} - \bar{X}_{.j})} \quad (5)$$

All statistical analyses were performed with *MPlus* Version 8.2. Missing data were handled with the full information maximum likelihood (FIML) method with robust standard errors. It was assumed data were missing at random (MAR). FIML has advantages over listwise deletion in that it produces unbiased parameter estimates if data was MAR and has greater power (Enders, 2010, p. 87). An additional advantage of the FIML estimator implemented in MPlus is that it calculates standard errors that are robust to deviations from normality (Muthén & Muthén, 2017, p. 668).

Results

We found a positive statistically significant relationship between SEC and achievement growth in each academic domain in primary and secondary schools that was larger in secondary schools. We found little evidence of a selection effect biasing the SEC effect. We found a concave non-linear SEC effect in several academic domains. We found that sensitivity to SEC decreased with increasing levels of student SES in secondary schools. We found that SEC accounted for a substantial proportion of variance in achievement growth due to schools and that SEC was one of the largest comparative predictors of academic achievement.

Figure 1 graphs a random subset of the Australian PISA dataset. The line of best fit indicated a positive relationship between SES and scientific literacy. A positive SEC effect was observed in the graph. Most students attending high SEC schools scored above average achievement, including those from low SES families. Inversely, most students attending low SEC schools scored below average achievement, including those from high SES families. There was an even spread of achievement by students attending middle SEC schools across family SES. This was consistent with the OECD's (2016) finding of SEC effects predicting academic achievement.

Figure 1. Effects of SES and SEC on scientific literacy (first plausible value) in Australia: PISA 2015.

Figure 2 graphs the grade 9 2017 NAPLAN dataset. It was observed that both SES and SEC predicted the likelihood of students' achievement of national minimum benchmarks across all academic domains. For each student SES category, SEC demonstrated a positive relationship with academic achievement. Low SES students in high SEC schools were 1.8 times more likely to achieve minimum benchmarks than low SES students in low SEC schools. The same relationship for middle and high SES students were 1.55 and 1.24 respectively. This indicates that low SES students may be more sensitive to the effects of SEC than high SES students.

Figure 2. Proportions of grade 9 students achieving minimum NAPLAN standards in 2017.

A positive, statistically significant relationship was found between SEC and each domain of achievement growth from the equation 1 residualised-change models which addressed measurement error. *Figure 3* shows the effect sizes in standard deviations for SES and SEC in each academic domain with the full results in *Table S1* in the technical appendix. Effect sizes were standardised on the total model variance to make them comparable across levels. Two patterns were observed in the results. Firstly, SEC had a larger effect on achievement growth in secondary schools compared to primary schools in each domain. Secondly, except for writing and spelling, SES was larger than SEC in primary schools, whereas SEC is larger than SES in secondary schools. In writing, SEC was always larger than SES while in spelling, SES equalled SEC in secondary schools.

Figure 3. Effect sizes of SES and SEC on achievement growth, NAPLAN 2017.

Note. Grey bars represent SES effects and black bars represent SEC effects.

We found little evidence that selection effects biased the relationship between SEC and achievement growth. Table S1 in the technical appendix shows the effect sizes for SES and SEC from equation 1 with, and without, the level-2 school sector variable. Controlling for school choice made very little difference to SEC effect sizes. With the addition of the school sector covariate, SEC was smaller by 0.01 SD in 7-9 writing, larger by 0.01 SD in 3-5 grammar, and smaller by 0.01 SD in 7-9 numeracy.

A statistically significant quadratic SEC coefficient was found across most academic growth domains in primary and secondary school levels with results summarised in *Tables 1* and 2. In each statistically significant case, the quadratic effect was negative, indicating a concave slope. That is, the SEC effect size diminished as school SEC increased.

Table 1

Results of modified equation 1 residualised change models grades 3-5 with quadratic term

	Reading	Writing	Spelling	Grammar	Numeracy
Intercept	239.52	247.59	214.01	236.37	247.75
Prior Ach.	0.73	0.58	0.85	0.67	0.76
SES	0.11	0.11	0.05	0.10	0.11
SEC	0.07	0.13	0.03	0.08	0.08

SEC ²	-0.02	-0.03	-0.01*	-0.02*	-0.01**
ε_{ij}	2629.41	2186.93	1244.72	4008.21	1590.40
δ_{0j}	138.05	222.58	104.76	236.37	152.82

Note. All $p < .001$ unless * $p < .01$ or ** $p < .05$ or ^{NS} (not significant). Fixed effect coefficients were standardized except for the intercept.

Table 2

Results of modified equation 1 residualised change models grades 7-9 with quadratic term

	Reading	Writing	Spelling	Grammar	Numeracy
Intercept	203.25	200.12	107.15	219.67	195.94
Prior Ach.	0.75	0.58	0.87	0.68	0.85
SES	0.08	0.10	0.04	0.07	0.06
SEC	0.11	0.16	0.04	0.09	0.09
SEC ²	-0.05	-0.03	0.00 ^{NS}	-0.03	-0.01 ^{NS}
ε_{ij}	1747.19	4977.90	1228.29	3203.20	5186.48
δ_{0j}	78.58	293.54	47.07	110.23	87.43

Note. All $p < .001$ unless * $p < .01$ or ** $p < .05$ or ^{NS} (not significant). Fixed effect coefficients were standardized except for the intercept.

A statistically significant interaction between SEC and SES was found in secondary schools with the SEC effect weakening as SES increased. *Tables 3 and 4* summarise the results for equation 3. No consistent pattern was found for the interaction between SES and SEC in primary schools. In most domains there was no significant interaction effect. In writing there was a significant negative interaction effect while in numeracy there was a significant positive effect. A consistent negative pattern was found for the interaction between SES and SEC in secondary schools.

Table 3

Results of equation 3 residualised change models grades 3-5 with interaction term

	Reading	Writing	Spelling	Grammar	Numeracy
Intercept	241.83	255.50	216.79	242.08	248.70
Prior Ach.	0.62	0.52	0.69	0.60	0.61
SES	6.77	5.50	2.47	9.17	3.71
SEC	11.97	13.54	4.39	14.32	10.73
SES x SEC	-0.08 ^{NS}	-0.55	-0.09 ^{NS}	-0.01 ^{NS}	0.31**
ε_{ij}	2594.48	2104.05	1229.24	3925.64	1575.14
δ_{0j}	120.19	185.25	101.83	242.08	144.11

Note. All $p < .001$ unless * $p < .01$ or ** $p < .05$ or ^{NS} (not significant). Fixed effect coefficients were unstandardized.

Table 4

Results of equation 3 residualised change models grades 7-9 with interaction term

	Reading	Writing	Spelling	Grammar	Numeracy
Intercept	209.16	210.93	109.37	223.90	195.83
Prior Ach.	0.68	0.66	0.86	0.64	0.73
SES	4.49	7.50	2.40	4.67	2.99
SEC	11.57	22.19	5.48	12.13	8.47
SES x SEC	-1.98	-1.87	-0.26**	-0.90	-0.45
ε_{ij}	1717.50	4840.12	1218.33	3152.58	904.05
δ_{0j}	64.69	247.40	45.47	100.58	83.98

Note. All $p < .001$ unless * $p < .01$ or ** $p < .05$ or ^{NS} (not significant). Fixed effect coefficients were unstandardized.

We found that SEC accounted for a substantial proportion of the school-level variance in academic growth. Table S1 in the technical appendix compares the null model for academic growth in each domain with the model of SEC in equation 1. In each case SEC substantially reduced the between-school variance in achievement growth, being the δ_{0j} term. On average in grades 3-5, SEC accounted for 36% of school-level variance in achievement growth, ranging between 30% and 50%. On average in grades 7-9, SEC accounted for 46% of school-level variance in achievement growth, ranging from 27% to 54%.

The SEC effect was found to be one of the largest predictors of scientific literacy in the PISA 2015 Australian sample. *Figure 4* graphs the effect sizes of all statistically significant predictors in equation 4. Most available indices in the dataset were included in the model, while allowing for error-free regression on all plausible values. A total of 54 variables including SEC were regressed onto scientific literacy. *Figure 4* shows that SEC was third in strength out of 30 statistically significant effects. Effect sizes were standardised on the total model variance to make them comparable across levels. It should be noted that many of the other predictors including *epistemology* and *perceived feedback*, were based on questions related to the science curriculum. Thus, the comparative strength of SEC in relation to many of the other measures in the PISA sample is substantial as it is a more distal construct from scientific literacy compared to the other more proximal concepts (Kraft, 2019). *Table S6* in the technical appendix provides the full results of the analysis of equation 4.

Figure 4. Comparative effect sizes of statistically significant predictors of scientific literacy in PISA 2015.

Note. Black bars represent positive effects and white bars represent negative effects.

Discussion

Our finding of a positive relationship between SEC and achievement growth is consistent with prior research (Palardy, 2008; Rumberger & Palardy, 2005). We believe that it is very unlikely that our findings are substantially biased as our measures of SES addressed measurement error through factor analysis and we found little evidence of bias due to selection effects from school choice. By showing that SEC has a predictive relationship with achievement growth in a population assessment of literacy and numeracy, we argue that it should be considered in school effectiveness research and future policy reform. SEC not only predicts achievement growth, but has been found to differentially predict the effectiveness of school practices (Palardy, 2008). Policy reforms aimed at lifting student achievement need to account for contextual differences as the impact of reforms may vary with school composition.

Our findings of SEC effects being larger in secondary schools than primary schools may be due to Australian secondary schools having a greater degree of socio-economic segregation than primary schools. In the 2017 NAPLAN dataset, 30% of grade five students attended private schools whereas 43% of grade nine students attended private schools. The difference between the average raw scores of SEC of public and private schools was 56% larger in grade nine than it was in grade five. That is, the systemic socio-economic differences between public and private schools in Australia are greater in secondary schools than primary schools. Thus, the strengthening of socio-economic segregation from primary to secondary schools may explain our finding of a larger SEC effect in secondary schools.

Our findings that the SEC effect diminished as school SEC increased, and that lower SES students are more sensitive to secondary school SEC than higher SES students, have significant implications for policy reforms. Alongside the main effect of SEC, these findings suggest that reducing the socio-economic segregation of students would improve both equity and systemic performance in achievement growth. It may be assumed that school socio-economic desegregation is a zero-sum game where the achievement gains of students moving from low to middle SEC schools would be offset by the achievement losses of high SEC schools becoming middle SEC schools. Our findings instead suggest that the potential gains of desegregating schools would be greater than any potential losses. This would be consistent with the first goal of the Alice Springs (Mparntwe) Education Declaration (2019) that “the Australian education system promotes excellence and equity” (COAGEC, 2019, p. 4). Further research is needed to confirm these results with other international samples and differing measures of school outcomes.

Our finding that SEC is a substantive effect suggests that it should be a central focus in future school-level policy reforms. We found that on average across academic domains, SEC accounts for approximately a third of primary school and half of secondary school variation in achievement growth. School-level reforms that ignore SEC are unlikely to succeed, especially given that research (Palardy, 2008) has found differential effectiveness of school practices according to SEC. The policy import of SEC is reinforced by our analysis of PISA 2015 showing SEC being a stronger predictor of achievement than nearly all measures of school resources and structures, teaching practices, and student attitudes and beliefs.

Overall, our findings suggest that school compositional effects continue to warrant focus by educational researchers and future policy reforms. A limitation of our study was its sole consideration of socio-economic compositional effects. Other compositional effects that have been found include intellectual ability (Opdenakker & Damme, 2001), ethnicity (Caldas & Bankston, 1998) and migration status (Peetsma et al., 2006). Some of these compositional effects may overlap or explain some of the variance due to SEC. A second limitation of our study was our measure of SES in the NAPLAN dataset. Both parental occupation and education had limited categories of differentiation compared to other large-scale studies such as PISA while there was no measure of family wealth. As such, the predictive relationships of SES and

SEC with achievement growth may have been diminished. A third limitation of our study design was that it did not measure class-level effects. The NAPLAN dataset allows the measurement of cohort effects while the PISA dataset allows the measurement of school effects. A three-level model of achievement growth measuring variance at student, classroom and school levels may find that a proportion of the student-level variance found in our residualised-change models may be better explained by classroom composition (Lamb & Fullarton, 2002).

Many avenues of future research would advance our understanding of SEC effects. In particular, our understanding of the mediators of SEC and achievement growth is not fully developed (Marks, 2015). Such knowledge may be able to guide interventions aimed at ameliorating SEC effects. Palardy's (2008, p. 26) model of school effects may provide a framework for future studies of mediators of SEC effects. Structural equation modelling may provide researchers a flexible means to identify mediators of compositional effects (Preacher et al., 2011; Preacher et al., 2010).

Conclusion

The veracity of SEC effects has been debated since it emerged as a major educational research and policy construct with the Coleman Report. Methodological critiques have remained as to whether measurement error and selection effects have biased statistically significant findings of compositional effects. Taking compositional effects as a given, many gaps remain in our understanding of the relationship between SEC and achievement growth. The comparative differences of SEC effects in primary and secondary schools is poorly understood. Little evidence has been found for differential sensitivity to the SEC effect and research findings have been inconsistent regarding the linearity of the SEC effect. The relative importance of SEC effects remains an active and open question for researchers and policy makers.

We found that SEC has a statistically significant relationship with achievement growth in primary and secondary schools when addressing error in the measurement of SES. We found minimal evidence of selection bias in our achievement growth models when controlling for school choice. As such we conclude that SEC has a real predictive relationship with achievement growth.

We found that the SEC effects were stronger in secondary compared to primary schools. We suggest that a possible mechanism for this effect is the greater degree of socio-economic segregation in Australian secondary schools compared to primary schools. Across most academic domains in primary and secondary schools, we found a concave non-linear SEC effect. As school SEC increased, its relationship with achievement growth decreased. Within secondary schools, we found a statistically significant interaction between student SES and school SEC. As student SES increased the effect of SEC decreased. We argue that reducing the socio-economic segregation of schools may increase average systemic performance and reduce socio-economic inequality.

We contend that SEC is of key importance to school-level reforms as it explains a large proportion of variance between schools in achievement growth, is one of the largest comparative predictors of academic achievement, predicts the average performance and equity of schools, and research has shown that the effectiveness of school-level practices vary with school SEC. Further research is required to better understand the mechanisms by which SEC is predictive of schooling outcomes internationally.

Acknowledgements

We wish to acknowledge the Australian Curriculum, Assessment and Reporting Authority for allowing access to the 2017 deidentified student-level NAPLAN dataset.

References

- Australian Bureau of Statistics (2017). *4221.0 – Schools, Australia, 2017*. Retrieved from <https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4221.0main+features22017>
- Australian Curriculum, Assessment and Reporting Authority (2017). *Assessment framework: NAPLAN online 2017-2018*. Sydney, Australia: ACARA. Retrieved from <https://www.nap.edu.au/docs/default-source/default-document-library/naplan-assessment-framework.pdf>
- Australian Curriculum, Assessment and Reporting Authority (2018). *NAPLAN achievement in reading, writing, language conventions and numeracy: National report for 2018*. Sydney, Australia: ACARA. Retrieved from <https://www.nap.edu.au/results-and-reports/national-reports>
- Armor, D. J., Marks, G. N., & Malatinszky, A. (2018). The Impact of school SES on student achievement: evidence from U.S. statewide achievement data. *Educational Evaluation and Policy Analysis*, 40(4), 613-630. doi:10.3102/0162373718787917
- Benito, Ricard, Miquel Àngel Alegre, & Isaac González-Balletbò. 2014. School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems. *Comparative Education Review* 58(1):104-34. doi: 10.1086/672011.
- Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In *Handbook of structural equation modeling*. (pp. 361-379). New York, NY, US: The Guilford Press.
- Caldas, S. J., & Bankston, C. (1998). The inequality of separation: Racial composition of schools and academic achievement. *Educational Administration Quarterly*, 34(4), 533-557. doi:10.1177/0013161x98034004005
- Chesters, J., & Daly, A. (2017). Do peer effects mediate the association between family socioeconomic status and educational achievement? *Australian Journal of Social Issues*, 52(1), 63-77. doi:10.1002/ajs4.3
- Chiu, M. M., & Khoo, L. (2005). Effects of resources, inequality, and privilege bias on achievement: Country, school, and student level analyses. *American Educational Research Journal*, 42(4), 575-603. doi:10.3102/00028312042004575
- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity* (Report No. OE-38001). Washington, DC: Government Printing Office.
- Council of Australian Governments Education Council. (2019). *Alice Springs (Mparntwe) Education Declaration*. Carlton South, Victoria: Education Services Australia.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138. doi:10.1037/1082-989X.12.2.121
- Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.
- Gollwitzer, M., Christ, O., & Lemmer, G. (2014). Individual differences make a difference: On the use and the psychometric properties of difference scores in social psychology. *European Journal of Social Psychology*, 44(7), 673-682. doi:10.1002/ejsp.2042
- Gonzalez, E. (2012). *Rescaling sampling weights and selecting mini-samples from large-scale assessment databases*. D. Hastedt, & M. von Davier (Edits.), IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 5, 117-134.
- Hallberg, K., Cook, T. D., Steiner, P. M., & Clark, M. H. (2018). Pretest measures of the study outcome and the elimination of selection bias: Evidence from three within study comparisons. *Prevention Science*, 19(3), 274-283. doi:10.1007/s11121-016-0732-6
- Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement*, 15(2), 177-199. doi:10.1076/sesi.15.2.177.30432

- Kraft, M. (2019). *Interpreting effect sizes of education interventions*. (EdWorkingPaper: 19-10). Retrieved from Annenberg Institute at Brown University: <http://www.edworkingpapers.com/ai19-10>
- Lamb, S., & Fullarton, S. (2002). Classroom and school factors affecting mathematics achievement: A comparative study of Australia and the United States using TIMSS. *Australian Journal of Education*, 46(2), 154-171. doi:10.1177/000494410204600205
- Lauen, D. L., & Gaddis, S. M. (2013). Exposure to classroom poverty and test score achievement: Contextual effects or selection? *American Journal of Sociology*, 118(4), 943-979. doi:10.1086/668408
- Laukaityte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics - Theory and Methods*, 46(22), 11341-11357. doi:10.1080/03610926.2016.1267764
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13(3), 203-229. doi:10.1037/a0012869
- Marks, G. N. (2015). Are school-SES effects statistical artefacts? Evidence from longitudinal population data. *Oxford Review of Education*, 41(1), 122-144. doi:10.1080/03054985.2015.1006613
- Marks, G. N. (2020). Should value-add school effects models include student- and school-level covariates? Evidence from Australian population assessment data. *British Educational Research Journal*. <https://doi.org/10.1002/berj.3684>
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44(6), 764-802. doi:10.1080/00273170903333665
- McPherson, A., & Willms, J. D. (1987). Equalisation and improvement: Some effects of comprehensive reorganisation in Scotland. *Sociology*, 21(4), 509-539. doi:10.1177/0038038587021004003
- Muthén, L.K. and Muthén, B.O. (2017). *Mplus user's guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.
- Opdenakker, M.-C., & Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal*, 27(4), 407-432. doi:10.1080/01411920120071434
- Organisation for Economic Co-operation and Development. (2003). *Literacy skills for the world of tomorrow: Further results from Pisa 2000*. Paris, France: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris, France: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Science competencies for tomorrow's world: Volume 1: Analysis*. Paris, France: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2010). *PISA 2009 results: Overcoming social background: Equity in learning opportunities and outcomes (Volume II)*. Paris, France: OECD Publishing.
- Organisation for Economic Co-operation and Development (2013). *PISA 2012 results: Excellence through equity (Volume II): Giving every student the chance to succeed*. Paris, France: OECD Publishing.
- Organisation for Economic Co-operation and Development (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris, France: OECD Publishing.
- Organisation for Economic Co-operation and Development (2017), *PISA 2015 technical report*. Paris, France: OECD Publishing.

- Organisation for Economic Co-operation and Development (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris, France: OECD Publishing. <https://doi.org/10.1787/b5fd1b8f-en>.
- Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: a multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, 19(1), 21-49. doi:10.1080/09243450801936845
- Palardy, G. J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal*, 50(4), 714-754. doi:10.3102/0002831213481240
- Palardy, G. J. (2015). High school socioeconomic composition and college choice: multilevel mediation via organizational habitus, school practices, peer and staff attitudes. *School Effectiveness and School Improvement*, 26(3), 329-353. doi:10.1080/09243453.2014.965182
- Peetsma, T., Van Der Veen, I., Koopman, P., & Van Schooten, E. (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement*, 17(3), 275-302. doi:10.1080/13803610500480114
- Perry, L., & McConney, A. (2010). School socio-economic composition and student outcomes in Australia: Implications for educational policy. *Australian Journal of Education*, 54(1), 72-85. <https://doi.org/10.1177/000494411005400106>
- Pokropek, A. (2015). Phantom effects in multilevel compositional analysis: problems and solutions. *Sociological Methods & Research*, 44(4), 677-705. doi:10.1177/0049124114553801
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209-233. doi:10.1037/a0020141
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: the advantages of multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 161-182. doi:10.1080/10705511.2011.557329
- Rangvid, B. S. (2007). School composition effects in Denmark: quantile regression evidence from PISA 2000. *Empirical Economics*, 33(2), 359-388. doi:10.1007/s00181-007-0133-6
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. 2nd ed (Vol. 1). Thousand Oaks: Sage.
- Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? the impact of student composition on academic achievement in high school. *Teachers College Record*, 107(9), 1999.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453. doi:10.3102/00346543075003417
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. (2nd ed.). Los Angeles: Sage.
- Sofroniou, N., Archer, P., & Weir, S. (2004). An analysis of the association between socioeconomic context, gender, and achievement. *The Irish Journal of Education / Iris Eireannach an Oideachais*, 35, 58-72.
- Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J., & Malmberg, L.-E. (2015). Phantom effects in school composition research: consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement*, 26(1), 75-101. doi:10.1080/09243453.2013.871302
- Thomson, S., De Bortoli, L., & Underwood, C. (2017). *PISA 2015: Reporting Australia's results*. Camberwell, Victoria: Australian Council for Educational Research.
- Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5(2), 134-150. doi:10.1016/j.edurev.2010.02.001
- Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). *What are plausible values and why are they useful*. IERI monograph series, 2(1), 9-36.
- Willms, J. (1986). Social Class Segregation and Its Relationship to Pupils' Examination Results in Scotland. *American Sociological Review*, 51(2), 224-241. <https://doi.org/10.2307/2095518>

- Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record*, 112(4), 1008-1037.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114-128. doi:10.1016/j.stueduc.2005.05.005

Technical Appendix

Table S1

Results of equation 1 residualised change models of NAPLAN 2017

	Null Model	Equation 1	Equation 1 with School Sector
Reading Grades 3-5			
Intercept	227.99	239.70	240.28
Prior Ach.	0.75	0.73	0.73
SES		0.11	0.11
SEC		0.07	0.07
Sector			-0.01
ε_{ij}	2672.10	2629.76	2629.89
δ_{0j}	220.31	137.89	136.63
Writing Grades 3-5			
Intercept	236.92	247.66	247.33
Prior Ach.	0.60	0.58	0.58
SES		0.11	0.11
SEC		0.13	0.13
Sector			0.01**
ε_{ij}	2213.96	2187.10	2187.14
δ_{0j}	338.41	224.06	223.70
Spelling Grades 3-5			
Intercept	210.61	214.05	213.79
Prior Ach.	0.86	0.85	0.85
SES		0.05	0.05
SEC		0.03	0.03
Sector			0.01**
ε_{ij}	1250.45	1244.69	1244.68
δ_{0j}	210.61	105.04	104.91
Grammar Grades 3-5			
Intercept	225.56	236.53	236.84
Prior Ach.	0.69	0.67	0.67
SES		0.10	0.10
SEC		0.08	0.09
Sector			-0.01 ^{NS}
ε_{ij}	4050.95	4008.47	4008.62
δ_{0j}	396.22	277.17	277.20
Numeracy Grades 3-5			
Intercept	238.40	247.83	248.22
Prior Ach.	0.77	0.76	0.76
SES		0.11	0.11
SEC		0.08	0.08
Sector			-0.01*
ε_{ij}	1618.60	1590.44	1590.47
δ_{0j}	218.36	152.69	152.23
Reading Grades 7-9			
Intercept	193.84	203.96	203.42
Prior Ach.	0.76	0.75	0.75
SES		0.08	0.08

	SEC		0.11	0.11
	Sector			0.01**
	ε_{ij}	1761.74	1747.98	1748.02
	δ_{0j}	176.15	82.06	81.61
Writing Grades 7-9				
	Intercept	189.53	200.58	198.73
	Prior Ach.	0.59	0.58	0.59
	SES		0.10	0.10
	SEC		0.16	0.15
	Sector			0.02
	ε_{ij}	5023.49	4978.35	4978.78
	δ_{0j}	620.64	294.50	290.18
Spelling Grades 7-9				
	Intercept	102.77	107.13	106.68
	Prior Ach.	0.87	0.87	0.87
	SES		0.04	0.04
	SEC		0.04	0.04
	Sector			0.01**
	ε_{ij}	1233.63	1228.29	1228.28
	δ_{0j}	64.02	47.04	46.90
Grammar Grades 7-9				
	Intercept	210.47	220.03	220.28
	Prior Ach.	0.69	0.68	0.68
	SES		0.07	0.07
	SEC		0.10	0.10
	Sector			-0.00 ^{NS}
	ε_{ij}	3222.83	3203.76	3203.78
	δ_{0j}	195.47	110.00	109.74
Numeracy Grades 7-9				
	Intercept	189.24	196.02	195.50
	Prior Ach.	0.86	0.85	0.85
	SES		0.06	0.06
	SEC		0.09	0.08
	Sector			0.01 ^{NS}
	ε_{ij}	916.93	908.86	908.87
	δ_{0j}	189.24	87.27	87.05

Note. All $p < .001$ unless * $p < .01$ or ** $p < .05$ or ^{NS} (not significant).

¹Fixed effect coefficients were standardized on the total model variance.

²Fixed effect coefficients were unstandardized.

Table S2
NAPLAN descriptive statistics

Grade	Variable	<i>M</i>	<i>SD</i>	Missing %
Three: 2015				
	Reading	427.48	93.74	.09
	Writing	418.14	68.21	.10
	Spelling	410.52	86.81	.09
	Grammar	435.08	101.46	.09
	Numeracy	399.96	81.75	.10
	SES	0.02	1.25	.03
	SEC	0.02	0.74	.00
Five: 2017				
	Reading	506.82	83.61	.03
	Writing	473.44	65.98	.03
	Spelling	501.03	73.12	.03
	Grammar	500.44	95.34	.03
	Numeracy	495.44	69.85	.04
	Sector	0.30		.00
Seven: 2015				
	Reading	548.48	71.96	.09
	Writing	512.68	79.48	.09
	Spelling	548.05	73.40	.09
	Grammar	543.79	83.84	.09
	Numeracy	545.66	71.89	.10
	SES	0.02	1.25	.03
Nine: 2017				
	Reading	583.12	70.33	.07
	Writing	553.00	98.22	.07
	Spelling	583.68	75.14	.06
	Grammar	575.89	83.83	.06
	Numeracy	595.60	65.03	.08
	Sector	0.43		.00

Table S3

PISA descriptive statistics (Australia 2015)

Variable	<i>M</i>	<i>SD</i>	Missing
Scientific Literacy	509.99	102.30	.00
Age	15.78	0.29	.00
First Generation	0.12		.00
Second Generation	0.30		.00
Gender	1.50		.00
Grade	0.03		.00
Indigenous	0.04		.00
Language	0.12		.03
Provincial	0.25		.00
Remote	0.01		.00
Catholic School	0.23		.00
Independent School	0.19		.00
Change Schools	0.92		.32
Repeat	0.07		.04
Adaptive Instruction	0.20	0.95	.25
Being Bullied	11.68	5.06	.06
Cooperation	0.09	0.96	.05
Emotional Support	0.11	0.96	.04
Enjoy Science	0.12	1.17	.13
Environmental Awareness	0.10	1.16	.13
Epistemology	0.26	0.99	.17
Inquiry-based Teaching	0.18	0.84	.22
Instrumental Motivation	0.16	1.06	.16
Interest in Science	0.04	1.04	.16
Motivation	0.33	0.96	.05
Perceived Feedback	0.07	0.98	.24
Science Activities	-0.30	1.12	.17
Science Disciplinary Climate	-0.19	1.02	.21
Science Learning Time	210.55	113.70	.09
Scientific Efficacy	0.07	1.27	.17
Sense of Belonging	-0.13	0.95	.09
Teacher Supportiveness	0.23	0.98	.23
Teacher-directed Instruction	0.27	0.99	.23
Test Anxiety	0.19	0.97	.04
Unfair Teacher	10.18	4.11	.06
Autonomy	0.76	0.21	.10
Certified Teachers	0.97	0.12	.14
Class Size	24.02	4.33	.09
Computer/Student Ratio	1.63	1.50	.13
Creative Extra-curriculum	2.20	0.91	.12
Educational Shortages	-0.35	0.84	.14
Leadership	0.56	1.02	.14
Negative Student Behaviors	-0.12	1.07	.14
Negative Teacher Behaviors	0.29	0.84	.14
Responsibility for Curriculum	0.00	0.87	.10
Responsibility for Resources	0.06	0.91	.10
School Size	839.01	442.51	.08

Science Resources	6.10	1.46	.12
Staffing Shortages	-0.20	1.01	.14
Student/Teacher Ratio	12.85	4.37	.15
Teacher Higher Qualifications	0.13	0.11	.20
Teacher Participation	4.27	2.19	.10
Total Science Teacher Number	8.79	6.87	.13

Table S4

Factor loadings and item descriptors of SES in NAPLAN 2017

Factor/Item Label	Item Description	Item Loading
SES (N = 1134724)		
Mother's Highest Education	1 = Year 9 or below, 2 = Year 10, 3 = Year 11, 4 = Year 12, 5 = Certificate I-IV, 6 = Diploma, 7 = Bachelor or above	-0.78
Father's Highest Education	As above	-0.79
Mother's Occupational Group	1 = senior management, 2 = other business manager, 3 = tradesperson, clerk, sales or service staff, 4 = machine operator, 5 = not in paid work	0.62
Father's Occupational Group	As above	0.71
RMSEA, CFI, TLI, SRMR Within, SRMR Between		NA

Table S5

Factor loadings and item descriptors of SES and SEC in PISA 2015

Factor/Item Label	Item Description	Item Loading
SES (N = 14148)		
Highest Parental Education in Years (PARED)	Pre-primary = 3, Primary = 6, Lower secondary = 10, Upper secondary for direct access to labor market = 11, Upper secondary for access to tertiary education = 12, Non-university tertiary = 14, University = 15	.65
Highest Parental Occupation (hisei)	International socio-economic index of occupational status.	.58
Home Possessions (HOMEPOS)	Item response theory score for household items (see OECD, 2017 p. 300).	.31
SEC (N = 754)		
Latent Average Highest Parental Education (PARED)	Latent school average score for highest parental education.	.96
Latent Average Highest Parental Occupation (hisei)	Latent school average score for highest parental occupation.	.98
Latent Average Home Possessions (HOMEPOS)	Latent school average score for home possessions.	.92
RMSEA, CFI, TLI, SRMR Within, SRMR Between		0.000, 1.000, 1.000, 0.000, 0.000

Table S6

Results of equation 4 of comparative predictors of scientific literacy in PISA 2015

Variable	Estimate	Notes
Intercept	475.89	
ε_{ij}	5443.27	
δ_{0j}	359.39	
SES	11.00	
SEC	32.75	
Age	0.15 ^{NS}	
First Generation	6.89 ^{NS}	
Second Generation	1.28 ^{NS}	
Gender	6.66*	
Grade	22.04	
Indigenous	-34.24	
Language	-28.36	
Provincial	8.86*	
Remote	2.64 ^{NS}	
Catholic School	-14.16*	
Independent School	-1.20 ^{NS}	
Change Schools	-5.33	
Repeat	-22.35	
Adaptive Instruction	6.34*	
Being Bullied	-0.08 ^{NS}	
Cooperation	1.34 ^{NS}	
Emotional Support	-2.60**	
Enjoy Science	12.01	
Environmental Awareness	5.34	
Epistemology	16.17	
Inquiry-based Teaching	-10.53	
Instrumental Motivation	-2.87**	
Interest in Science	7.60	
Motivation	5.90	
Perceived Feedback	-15.57	
Science Activities	-2.25 ^{NS}	
Science Disciplinary Climate	3.67*	
Science Learning Time	0.05	
Scientific Efficacy	7.45	
Sense of Belonging	-2.99**	
Teacher Supportiveness	-3.10 ^{NS}	
Teacher-directed Instruction	5.99	
Test Anxiety	-8.01	
Unfair Teacher	-3.28	
Autonomy	3.58 ^{NS}	
Certified Teachers	7.57 ^{NS}	
Class Size	0.70 ^{NS}	
Computer/Student Ratio	1.58 ^{NS}	
Creative Extra-curriculum	1.98 ^{NS}	
Educational Shortages	-3.32 ^{NS}	
Leadership	-1.76 ^{NS}	
Negative Student Behaviors	-9.47	

Negative Teacher Behaviors	0.14 ^{NS}	
Responsibility for Curriculum	-0.08 ^{NS}	
Responsibility for Resources	2.81 ^{NS}	
School Size	0.00 ^{NS}	
Science Resources	1.31 ^{NS}	
Staffing Shortages	3.89**	
Student/Teacher Ratio	0.33 ^{NS}	
Teacher Higher Qualifications	27.25**	Sum of the proportion of teachers with masters and doctoral qualifications.
Teacher Participation	-1.34 ^{NS}	
Total Science Teacher Number	0.14 ^{NS}	

Note. All $p < .001$ unless * $p < .01$ or ** $p < .05$ or ^{NS} (not significant).