The impact of school academic quality on low socioeconomic status students

PATRICK LIM

SINAN GEMICI

TOM KARMEL
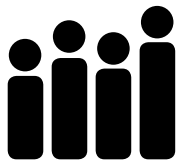
Longitudinal Surveys of Australian Youth

The impact of school academic quality on low socioeconomic status students

Patrick Lim
Sinan Gemici
Tom Karmel

National Centre for Vocational Education Research

LONGITUDINAL SURVEYS OF
AUSTRALIAN YOUTH

RESEARCH REPORT 63

About the research

The impact of school academic quality on low socioeconomic status students

Patrick Lim, Sinan Gemici and Tom Karmel, NCVER

This paper uses data from the Longitudinal Surveys of Australian Youth (LSAY) to investigate the impact of academic school quality on student outcomes. A companion paper by Gemici, Lim and Karmel (2013) describes the measure of school quality used in this paper.

In particular, this paper examines the interactions between students' individual socioeconomic status (SES), their academic achievement at age 15 years and the academic quality of the school they attend and school completion, tertiary entrance rank (TER) and university participation. This paper explores whether students from low socioeconomic backgrounds benefit to a greater or lesser extent from attending high-quality schools when compared with their more advantaged peers.

Key messages

- Academic school quality has a considerable differential effect on school completion for those who come from the lowest socioeconomic band. It also has a differential effect for those with low academic achievement at age 15 years.

- A differential effect is also seen in relation to the impact of academic school quality on tertiary entrance rank and the probability of going to university.

- Coming from a high socioeconomic background insulates students from early school leaving, even if they are weak performers and attend a non-academic school.

The conclusion is that the quality of the school matters and that students from a low socioeconomic background benefit even more from attending a school of high academic quality.

Tom Karmel
Managing Director, NCVER

Contents

Tables and figures

Tables

Figures

Abstract

One of the enduring goals of Australian social policy is to improve the educational outcomes of students from lower socioeconomic status (SES) backgrounds. A 2013 paper by Gemici, Lim and Karmel derived a measure of academic school quality. This paper is a follow-on study and examines the interactions between students' individual socioeconomic status, their academic achievement at age 15 years and the academic quality of the school they attend. The primary focus of this paper is how these interactions affect Year 12 completion. Tertiary entrance rank (TER) and university enrolment are also included as ancillary outcomes.

The findings show that academic school quality has a differential impact on school completion for those from a low socioeconomic background and those with low academic achievement at age 15 years; that is, those who come from a low socioeconomic background benefit more from academic school quality than those who have higher socioeconomic backgrounds. The effect is more pronounced for those from a low socioeconomic background who also had a low achievement score at age 15.

Introduction

One of the enduring goals of Australian social policy is to improve the educational outcomes of students from lower socioeconomic status (SES) backgrounds.[1]

The relationship between individual socioeconomic disadvantage and academic outcomes is well established, and current data confirm that underprivileged students have lower rates of Year 12 completion and university uptake (Department of Education, Employment and Workplace Relations 2011; Steering Committee for the Review of Government Service Provision 2011). There is also substantial evidence that the quality and socioeconomic profile of schools matter with respect to academic outcomes (Gonski et al. 2011; Organisation for Economic Co-operation and Development [OECD] 2010; Perry & McConney 2010; Watson & Ryan 2010; Gemici, Lim & Karmel 2013).

The concept of a 'high quality' school is inherently contentious and partly depends on a person's perspective on the objectives of schooling. However, facilitating key academic outcomes such as Year 12 completion and the transition to higher education are undoubtedly critical aims of schooling. Based on this perspective, Gemici, Lim and Karmel (2013) used data from the Longitudinal Surveys of Australian Youth (LSAY) to explore the impact of numerous school characteristics on tertiary entrance rank (TER) and the probability of commencing university by age 19 years. As part of the analysis, the authors used their statistical models to derive 'academic quality' scores for each of the schools in their sample. This present paper uses these academic quality scores to explore whether there is a differential effect of academic school quality on students from low socioeconomic backgrounds. In other words, the primary focus of this analysis is on whether low socioeconomic status students benefit to a greater or lesser extent from attending high-quality academic schools when compared with their more advantaged peers. It also looks at the differential effect on another disadvantaged group: those with relatively low academic achievement at age 15 years.

This paper uses data from the 2006 LSAY cohort to examine the interactions between students' individual socioeconomic status, their individual academic achievement (their academic achievement at age 15 years as measured by the Programme for International Student Assessment [PISA]) and the academic quality of the school they attend. The primary focus is on how these interactions affect Year 12 completion, noting that completing school is an explicit goal stated by the overwhelming majority of students[2] and a clear policy objective of government. Tertiary entrance rank and university enrolment at age 19 years are included in the analysis as ancillary outcomes. These latter outcomes are subject to a selection process that favours students from medium and high socioeconomic backgrounds. Not all students who complete Year 12 obtain a tertiary entrance rank and enrol at university. Further, a disproportionate number of those have medium and high socioeconomic backgrounds. Nonetheless, tertiary entrance rank and university are included in addition to Year 12 completion, as they also represent key outcomes from schooling.

---

[1] For example through the Smarter Schools National Partnership for Low Socio-economic Status School Communities, the Australian Government is providing $1.5 billion over seven years (2008—09 to 2014—15) to support education reform activities in approximately 1700 low socioeconomic status schools around the country. This funding is matched by co-investment from the states and territories (Council of Australian Governments 2008).

[2] In the first wave of the 2006 LSAY cohort (modal age of 15), only 5.5% of respondents explicitly stated that they were not planning to complete Year 12.

The remainder of the paper is structured as follows. The first section describes the data and modelling approach. The next section presents the results from statistical modelling, while the last section provides a conclusion.

Method

Data and sample

This study uses data from the 2006 cohort of the Longitudinal Surveys of Australian Youth. LSAY tracks a nationally representative sample of 15-year-olds over a period of ten years to capture young people's transition from school to tertiary education and work. The 2006 base year of LSAY is linked to the 2006 Programme for International Student Assessment, which provides a rich set of individual and school-level measures.

A total of 14 170 students participated in the 2006 base year. Attrition in longitudinal surveys reduces initial samples over time, as some students drop out for a variety of reasons (see Rothman 2009). The present analysis includes all students who were still part of the LSAY sample in 2010 (6316 students in 356 schools). An appropriate analysis weight is used to account for the effects of complex sampling and response bias. Details on the creation of this weight are provided in appendix A.

Outcome measures

The primary outcome measures for this study are Year 12 completion by 2010 (modal age of 19 years), tertiary entrance rank and university enrolment by age 19 years. Descriptive data for the outcome measures are provided in table 1.

Table 1   Descriptive statistics

Outcome measures	n	% (unweighted)	% (weighted)[1]
Year 12 completion			
Completed	5426	84.1	85.9
Not completed	890	15.9	14.1
Total	**6316**	**100.0**	**100.0**
TER groupings[2]			
Q1 (lowest)	966	17.6	15.3
Q2	880	12.9	13.9
Q3	966	13.2	15.3
Q4 (highest)	976	12.2	15.5
Completed Y12 but not awarded TER (or unknown)	1638	28.2	25.9
Early school leaver	890	15.9	14.1
Total	**6316**	**100.0**	**100.0**
Higher education study status (at age 19)			
Commenced bachelor or higher degree	3276	45.7	51.9
Not commenced bachelor or higher degree	3039	54.3	48.1
Total[3]	**6315**	**100.0**	**100.0**

Notes: 1 Weighted for both sample design and attrition.
       2 TER quartiles are determined for those that received a TER score only.
       3 One individual has unknown information and is coded as missing.

Explanatory measures

The primary focus of this study is to examine whether a school's academic quality affects a student's likelihood of completing Year 12, their tertiary entrance rank score and university enrolment differently, based on individual socioeconomic background. Such differential effects can be captured via the interplay of students' socioeconomic status, individual academic achievement and academic school quality.

Individual socioeconomic status

In LSAY, individual socioeconomic status is commonly measured using the Index of Economic, Social and Cultural Status (ESCS). This index, which represents a mixture of parental occupation, parental education and home possessions, measures students' socioeconomic status across all 57 countries that participated in the Programme for International Student Assessment in 2006. The problem associated with the Index of Economic, Social and Cultural Status is that the need for multi-country usability renders the measure less relevant in the Australian context. The present paper addresses this issue by creating a custom measure from the 2006 Programme for International Student Assessment variables similar to the Index of Economic, Social and Cultural Status, yet more accurately captures the variation in students' socioeconomic status in Australia. Details on the creation of this measure are provided in Lim and Gemici (2011). An individual is of low socioeconomic status if they are in the lowest quartile of this measure.

Individual academic achievement

The Programme for International Student Assessment assesses the literacy of 15-year-olds in three major domains: reading, mathematics and science. These literacy scores are often used as proxies for academic achievement. In this study, a composite academic achievement measure is created by averaging literacy scores across the three domains for each individual.[3] Every Programme for International Student Assessment survey tests these three domains in terms of general understanding. The Programme for International Assessment does not test how well a student understands the specific curriculum.

Academic school quality

The academic school quality variable is derived from a companion paper (see Gemici, Lim & Karmel 2013), in which the authors identified the school attributes that influence young people's transition to university *over and above* their individual background characteristics. As part of the companion paper a cluster analysis was performed to identify groups of high, medium and low-performing schools, based on students' predicted tertiary entrance rank and the probability of university enrolment by age 19 years. In the present paper, a single continuous measure of academic school quality is created by combining the predicted tertiary entrance ranks and probabilities of university enrolment for each student (net of their individual background characteristics) and then aggregating student scores up to the school average.

The advantages of using predicted tertiary entrance rank and probabilities of university enrolment to create an academic school quality measure for Year 12 completion are twofold. Conceptually, a school's emphasis on academic success is most strongly reflected in the predicted tertiary entrance

---

[3] The Programme for International Student Assessment calculates five plausible value scores per domain and student (see OECD 2009 for details). The academic ability measure used in this study is calculated from the first of the five plausible value scores.

rank and university enrolment probability of its student body. From an analytical perspective, tertiary entrance rank and university enrolment by age 19 years offer more variation in the data and thus lend themselves to the construction of a more robust measure of academic school quality.

It is important to emphasise that differences in relevant school characteristics (notably, sector, gender mix, average socioeconomic status of the student body, academic pressure from parents, school-level variables) are accounted for in the academic school quality measure. This measure also accounts for idiosyncratic school differences; that is, unmeasured characteristics that reflect issues such as school culture or ethos. For further details on the academic school quality measure, readers are referred to the companion paper.

Other relevant individual background characteristics

The focus of this paper is on individuals who come from low-socioeconomic backgrounds. Of course, individuals have diverse backgrounds and therefore to narrow the focus to socioeconomic status, it is necessary to reduce the impact of any potential biases in the sample. For example, if an experiment were conducted to investigate the impact of coming from a low socioeconomic background on completing Year 12, low socioeconomic status could be thought of as a 'treatment condition' and young people could be randomly allocated to either a low socioeconomic status group or an appropriate control group. The process of randomisation assists in ensuring that the background characteristics of individuals are spread 'evenly' between the low socioeconomic status and the control groups. Given that this is neither practical nor possible, a methodology based on propensity scores (Rosenbaum & Rubin 1983) is used to help remove potential selection bias and, at the same time, account for relevant individual background characteristics. In the present analysis, these characteristics comprise gender, Indigenous status, parental education, regionality, achievement scores in mathematics, reading and science, immigration status and language spoken at home. The propensity score estimates the probability that an individual comes from a low socioeconomic background. By weighting these propensities, any differences in the outcome variable can be more confidently attributed to coming from a low socioeconomic background, net of other time-invariant influences. Details on the calculation and use of propensity scores in creating the weights are provided in appendix A.

Modelling approach

The modelling approach entails regressing the outcomes of interest — Year 12 completion, tertiary entrance rank and university enrolment — on individual socioeconomic status, individual academic achievement at age 15 and school quality, together with their corresponding interactions. Including these interactions is the basis for the paper: we are interested in whether the impact of school quality varies by socioeconomic and academic achievement. The remaining background characteristics for individuals and schools are not of interest here and are accounted for via the use of propensity scores at the student level (see also appendix A), and via the use of the academic school quality variable at the school level. Our particular interest is in school quality interactions with student socioeconomic status and academic achievement. For completeness we also include a student socioeconomic status by academic achievement interaction (table 2).[4]

---

[4] Note that the school quality variable is constructed using both the TER score and the probability of university enrolment. Thus, the results for these two outcomes need to be interpreted cautiously; that is, by considering only the interactions between school quality and individual SES and achievement. That is, the overall relationship (main

Table 2   Predictors used in regression analysis

Individual predictors	Interaction terms
Student SES	Student SES by student academic achievement
Student academic achievement	Student SES by school quality
School quality	Student academic achievement by school quality

The sampling strategy of LSAY is to sample schools before randomly selecting individuals from within each school. This strategy means that individuals who attended the same school are more similar on relevant background characteristics (and they also share the same school quality). To account for this structure, multi-level (mixed or nested) models are the most appropriate technique; that is, the effects of individuals are modelled within schools. Mixed models allow a covariance structure to be fitted at the school level. Details on the multi-level model are provided in appendix B.

The multi-level model includes a combined weight variable. This combined weight variable comprises the multi-level weight, sampling weight for schools and individuals (that is, the probability of selection of individuals), longitudinal weight (to help reduce the impact of attrition from wave 1), and the propensity scores weight. (Propensity scoring is a technique that allows the creation of a weight to help reduce the impact of the non-random assignment of respondents into groups, in our case, into the low socioeconomic group.) Details on the construction of this weight variable are provided in appendix A.

---

effects) between school quality and TER and university enrolment will be important simply through the construction of the school quality variable.

Results

The emphasis of this analysis is on differential effects for students from different socioeconomic backgrounds. This means that the primary focus of the present analysis is on whether low socioeconomic status students benefit to a greater or lesser extent from attending high academic quality schools when compared with their higher socioeconomic background peers (given similar individual academic achievement).

The differential effects are best illustrated using a series of figures. The figures comprise the interaction effects of school quality for students from the range of socioeconomic and academic achievement backgrounds on the three outcomes investigated (Year 12 completion, tertiary entrance rank score and university enrolment). The figures are split into three separate panels which differentiate students on individual academic achievement. The panel on the left captures students with low academic achievement (around the tenth percentile); the panel in the middle is for students whose academic achievement is around the median; the panel on the right features students with high academic achievement (around the 90th percentile). The respective panel labels are placed horizontally across the top of the figure.

Each of the three panels has a separate x-axis, which represents academic school quality scores ranging from low to high. The y-axis represents the outcome of interest jointly for all three panels. The body of each panel contains three lines that represent individual student socioeconomic status (low socioeconomic status around the tenth percentile; medium socioeconomic status around the median; high socioeconomic status around the 90th percentile). The legend with the respective line labels is placed horizontally across the bottom of the figure.

Each line features 95% confidence intervals (so-called 'error bars'), which capture the range of prediction error. The size of prediction error decreases as the number of students per school in the sample increases and vice versa. The significance of the factors included in the models appears in the tables presented in each section below, and it is important that the regression coefficients are read in conjunction with the figures presented.

Year 12 completion

Table 3 presents the regression results for the probability of completing Year 12. Statistically significant predictors are highlighted.

Table 3   Regression results for the probability of completing Year 12

Effect	β	SE	df	t	Pr > |t|
Intercept	**2.383**	**0.099**	**325**	**24.01**	**<0.0001**
Student SES	0.082	0.070	4381	1.17	0.2418
Student academic achievement	**1.220**	**0.077**	**4381**	**15.83**	**<0.0001**
School quality	**0.470**	**0.091**	**325**	**5.14**	**<0.0001**
Student SES by student academic achievement	-0.045	0.056	4381	-0.81	0.4159
Student SES by school quality	**-0.150**	**0.053**	**4381**	**-2.84**	**0.0046**
Student academic achievement by school quality	**-0.132**	**0.063**	**4381**	**-2.09**	**0.0366**

Figure 1   Differential effects of academic school quality on Year 12 completion
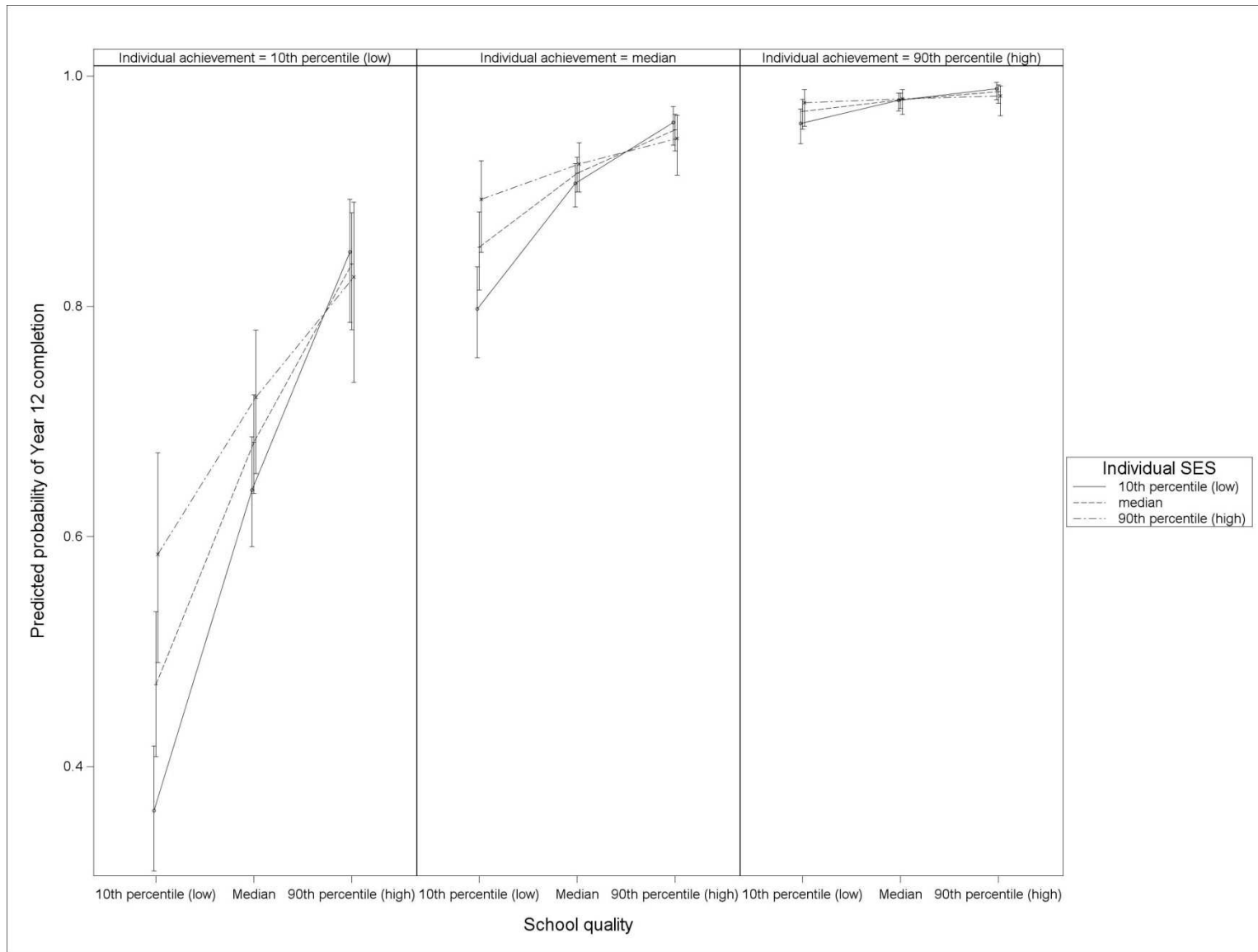
The low academic achievement (individual achievement = 10th percentile [low]) panel on the left of figure 1 shows that academic school quality has a strong differential effect for students from distinct socioeconomic backgrounds. For low-ability students from low socioeconomic backgrounds attending low-quality schools, the probability of completing Year 12 is less than 0.4, compared with more than 0.5 for their high socioeconomic peers. The difference is statistically significant, as indicated by the non-overlapping error bars; that is, a school with low academic quality has a particularly negative effect on Year 12 completion for low socioeconomic background students of low academic achievement.

When looking at low ability, for low socioeconomic students who attend a low-quality to even a medium-quality school, their chance of completing Year 12 is raised from less than 40% to above 60%. And if these students attended a high-quality school their chance of completing Year 12 would be in excess of 80%. Therefore, for this most vulnerable group of all students, increasing quality has an extremely large impact on a key outcome from schooling.

As academic school quality increases, individual socioeconomic background becomes less relevant in relation to the probability of completing Year 12. Low-ability students in high-quality academic schools have a predicted probability of Year 12 completion of around 0.8, regardless of their individual socioeconomic background.

These patterns are seen to a much more moderate extent within students of average academic achievement. For students of high academic achievement, school quality has almost no effect — the predicted probability of completing Year 12 lies above 0.95.

Tertiary entrance rank

The first point to note is that the significant positive relationship between tertiary entrance rank score and school quality is present by design. The school quality variable has been constructed by considering the tertiary entrance rank score of individuals. For this reason, there will be a strong positive relationship between tertiary entrance rank and school quality. However, the interest in this analysis is on the interaction relationships between individual socioeconomic status and school quality and academic achievement.

Figure 2 Differential effects of academic school quality on tertiary entrance rank
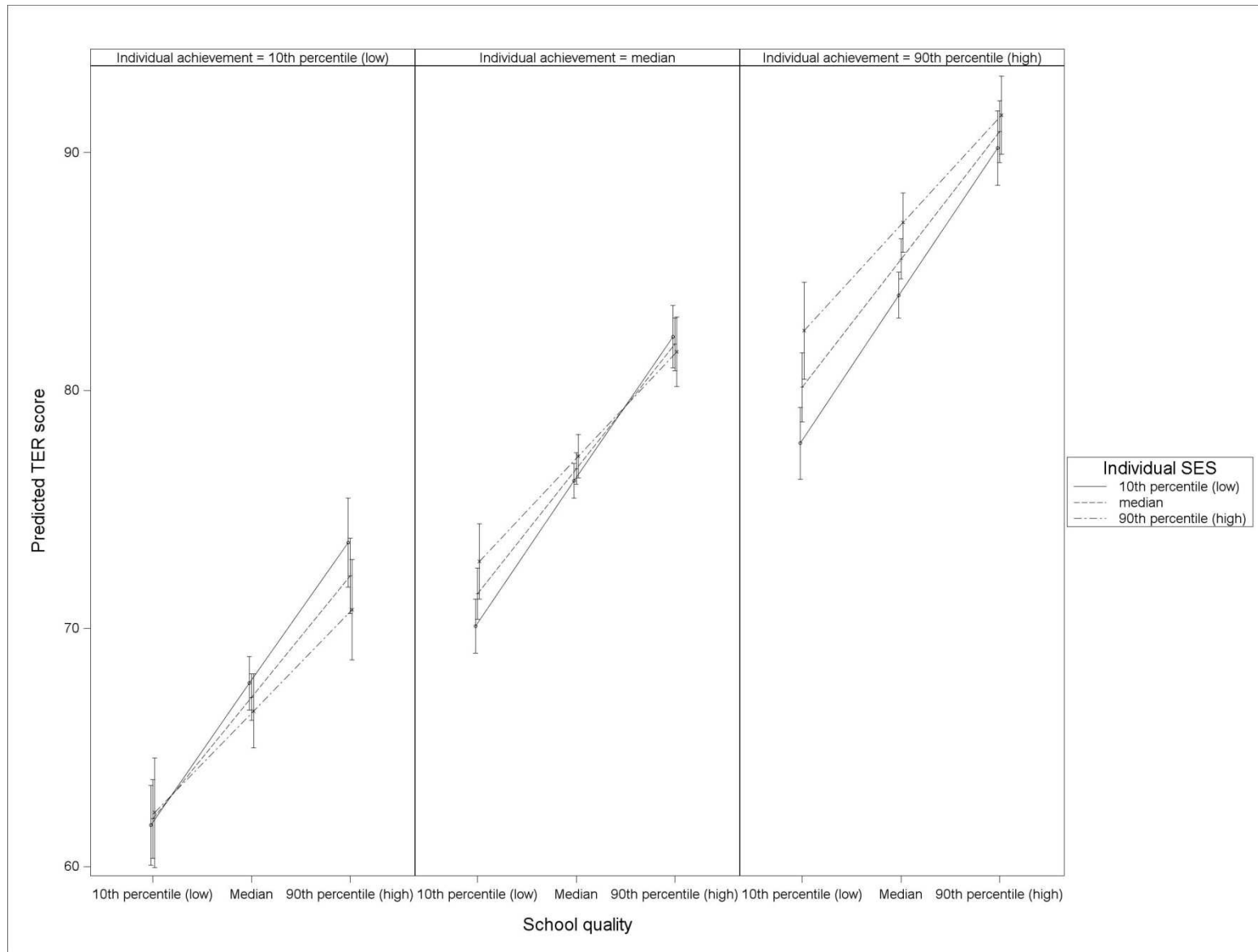
From table 4, an important point is the negative interaction effect between student socioeconomic status and school quality. This is the same effect observed for Year 12 completion and indicates that low socioeconomic background students benefit more from attending a high academic quality school than do high socioeconomic background students.

From table 4, we also see that the student academic achievement by academic school quality interaction is not statistically significant; that is, academic school quality does not impact differentially on those with low or high academic achievement after controlling for socioeconomic status.

Table 4   Regression results for tertiary entrance rank

Effect	β	SE	df	t	Pr > |t|
Intercept	**76.72***	**0.333**	**324**	**230.65**	**<0.0001**
Student SES	**0.39***	**0.197**	**3375**	**2.00**	**0.0461**
Student academic achievement	**7.22***	**0.250**	**3404**	**28.91**	**<0.0001**
School quality	**3.87***	**0.327**	**314**	**11.85**	**<0.0001**
Student SES by student academic achievement	**0.63***	**0.208**	**3462**	**3.02**	**0.0026**
Student SES by school quality	**-0.47***	**0.193**	**3072**	**-2.43**	**0.0150**
Student academic achievement by school quality	0.08	0.243	3142	0.31	0.7535

Note: Statistically significant predicators are bolded.

University enrolment at age 19 years

Table 5   Regression results for university enrolment by age 19 years

Effect	β	SE	df	t	Pr > |t|
Intercept	**0.790***	**0.065**	**325**	**12.11**	**<0.0001**
Student SES	**0.162***	**0.045**	**4381**	**3.59**	**0.0003**
Student academic achievement	**1.264***	**0.060**	**4381**	**21.16**	**<0.0001**
School quality	**0.765***	**0.063**	**325**	**12.08**	**<0.0001**
Student SES by student academic achievement	0.017	0.048	4381	0.36	0.7223
Student SES by school quality	**-0.099***	**0.043**	**4381**	**-2.30**	**0.0212**
Student academic achievement by school quality	-0.008	0.054	4381	-0.15	0.8790

Note: Statistically significant predictors are bolded.

The results of interest for university enrolment are shown in figure 3. Again, the interest in this figure lies in looking at the interactions of individual socioeconomic background with school quality and achievement. The fact that the probability of higher education is increasing with school quality is a given by the design of the school quality variable.

As with the two previous outcomes, the sign on the student socioeconomic status by academic school quality interaction is negative and statistically significant. Again, we find that those with a low socioeconomic background benefit more from attending a high academic quality school than their high socioeconomic background counterparts. The sign of the student academic achievement by academic school quality is also negative, but not of any significance (substantially and statistically).

Figure 3   Differential effects of academic school quality on university enrolment by age 19 years
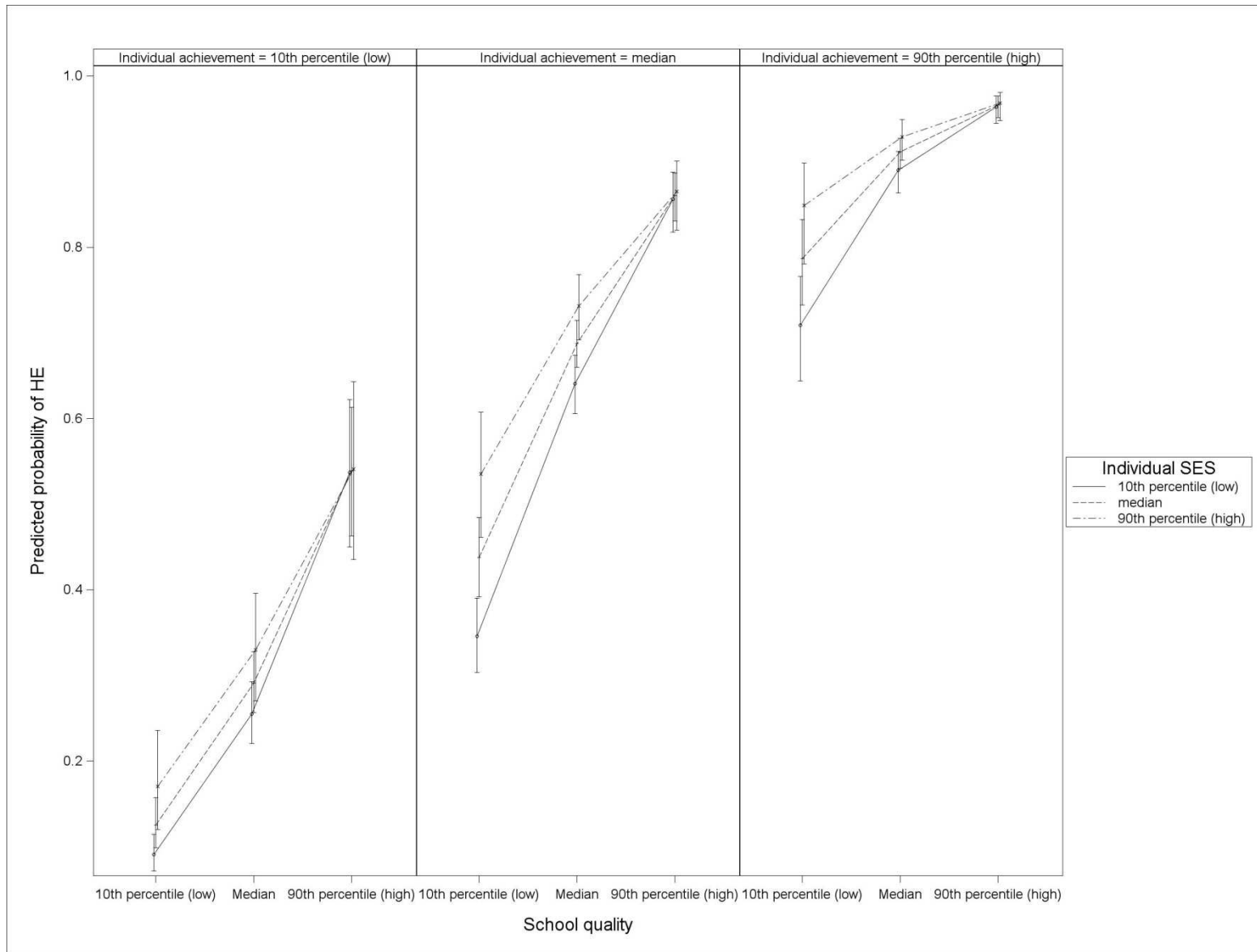
Conclusion

There is ample evidence suggesting that schools play an important role with respect to key student outcomes such as Year 12 completion and the transition to tertiary education (Gemici, Lim & Karmel 2013; OECD 2010; Rothman & McMillan 2003). The present paper has explored whether students from a low socioeconomic background benefit to a greater or lesser extent from attending high-quality academic schools when compared with their more advantaged peers.

The results indicate that academic school quality has a considerable differential effect on school completion for the most vulnerable of students: those who come from the lowest socioeconomic stratum and who are in the lowest academic achievement decile. The differential effect is that school academic quality is more important for the most vulnerable students.

What is notable is the strength of this effect. For students suffering from a 'double dose' of disadvantage, academic school quality may indeed have a critical impact on their completing school. In contrast, coming from an advantaged background can insulate students from early school leaving.

In addition, there is a significant interaction between academic school quality and socioeconomic background in relation to both tertiary entrance rank and the probability of going to university. Individuals from a low socioeconomic background benefit differentially from attending a school with a high academic quality.

In conclusion, this paper shows that the quality of the school matters, even more so for individuals from a low socioeconomic background.

References

Chantala, K, Blanchette, D & Suchindran, CM 2011, *Software to compute sampling weights for multilevel analysis*, viewed September 2012, Carolina Population Center, University of North Carolina, <www.cpc.unc.edu/research/tools/data_analysis/ml_sampling_ weights/Compute%20Weights%20for%20Multilevel%20Analysis.pdf>.

Council of Australian Governments 2008, *National Partnership Agreement on Low Socio-Economic Status School Communities,* COAG, Canberra, viewed September 2012, <www.smarterschools.gov.au/low-socio-economic-status-school-communities>.

Department of Education, Employment and Workplace Relations 2011, *Student 2010 full year: selected higher education statistics publication*, DEEWR, Canberra.

Gemici, S, Lim, P & Karmel, T 2013, *The impact of schools on young people's transition to university*, NCVER, Adelaide.

Gonski, D, Boston, K, Greiner, K, Lawrence, C, Scales, B & Tannock, P 2011, *Review of funding for schooling — final report*, Department of Education, Employment and Workplace Relations, Canberra.

Lim, P 2011, *Weighting the LSAY Programme of International Student Assessment Cohorts*, NCVER, Adelaide.

Lim, P & Gemici, S 2011, *Measuring the socioeconomic status of Australian youth*, NCVER, Adelaide.

Lumsden, M & Stanwick, J 2012, *Who takes a gap year and why?,* Briefing paper 28, NCVER, Adelaide.

OECD (Organisation for Economic Co-operation and Development) 2009, *PISA 2006 technical report*, OECD, Paris.

——2010, *PISA 2009 results: what makes a school successful? Resources, policies and practices,* OECD, Paris.

——2012, *PISA 2009 technical report*, OECD, Paris.

Perry, L & McConney, A 2010, 'School socioeconomic composition and student outcomes in Australia: implications for educational policy', *Australian Journal of Education*, vol.54, pp.72—85.

Rosenbaum, PR & Rubin, DB 1983, 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, vol.70, pp.41—55.

Rothman, S 2009, *Estimating attrition bias in the Year 9 cohorts of the Longitudinal Surveys of Australian Youth*, Technical report 48, Australian Council for Educational Research, Melbourne.

Rothman, S & McMillan, J 2003, *Influences on achievement in literacy and numeracy*, LSAY research report 36, Australian Council for Educational Research, Melbourne.

Steering Committee for the Review of Government Service Provision 2011, *Report on government services 2011*, Productivity Commission, Canberra.

Watson, L & Ryan, C 2010, 'Choosers and losers: the impact of government subsidies on Australian secondary schools', *Australian Journal of Education*, vol.54, pp.86—107.

Appendix A

Creation of the analysis weight

In this paper, four different weights are created and combined to help ensure that the results are representative of the population under consideration. The four weights entail the sampling weight, attrition weight, propensity score weight and multi-level weight.

Sampling weight

Given the nature of the sampling design of LSAY, the probability of an individual being selected to participate is dependent first on the probability of selection of their school, and then on the probability of their being selected from that school. A sample weight is calculated to ensure that the final sample is representative of the population being investigated. There are two weights of interest in the case of LSAY: the school weight (probability of a school being selected) and the individual weight (probability of an individual being selected). Details on the creation and use of these weights are provided in OECD (2012).

Attrition weight

Given that LSAY is a longitudinal survey, there is a chance that different groups of respondents drop out of the survey at different rates (that is, differential attrition). For each wave of the survey a weight is calculated from the key variables to account for the effect of differential attrition. These attrition weights adjust the sample to ensure that the distribution matches that observed in the first wave. Details on the creation of attrition weights for LSAY are provided in Lim (2011).

Propensity score weight

Non-random selection may cause the results from any analysis to be biased; that is, reported effects may in fact be higher (or lower) than otherwise observed, had an individual been randomly assigned to the 'treatment' group of interest. (This effect is known as selection bias.) The use of propensity scores (Rosenbaum & Rubin 1983) provides a way to mitigate selection bias. The propensity scores model serves as a mechanism for determining the probability of an individual coming from a low socioeconomic background, based on relevant individual student characteristics.

The propensity score model used in this paper determines the probability of an individual coming from a low socioeconomic background. (Low socioeconomic status is defined as being in the bottom quartile of the socioeconomic status distribution.) This probability is determined using a logistic regression model that includes a range of individual-level characteristics commonly associated with an individual's socioeconomic background, such as parental education, Indigenous status or regionality.

The propensity score model used in this paper is

$$logit(\boldsymbol{y_i}) = \alpha + \boldsymbol{X\tau} + \boldsymbol{\epsilon}$$

where $\boldsymbol{y_i}$ is a vector of 0's and 1's indicating whether an individual is classified as being from a low socioeconomic background, $\boldsymbol{X}$ is the design matrix of fixed explanatory effects, $\boldsymbol{\tau}$ is the vector of fixed effects relating to Indigenous status, gender, parental education background, regional status, mathematics, reading and science academic achievement scores, immigration status and language

spoken at home, and ϵ is the vector of residuals, which are assumed to be normally distributed with a mean of 0 and a variance of σ^2.

The probability of being from a low socioeconomic background is converted to a weight (1/probability of being low socioeconomic status) to ensure an equal distribution of low socioeconomic status individuals across the range of characteristics included in the model. (This is similar to randomly distributing individuals to the low socioeconomic background group in an experimental design setting.) These weights are then included in any subsequent statistical analysis. Tables A1 and A2 outline the impact of the propensity score weights on the predictors used in the regression model.

Table A1 Impact of propensity score weights on categorical variables

Categorical predictors	Percentage of individuals classified as low SES	
	Raw data	PS-weighted data
Indigenous status		
Indigenous	51.5	49.6
Not Indigenous	22.9	51.1
Geographic location		
Metropolitan	21.7	49.6
Provincial	31.4	49.9
Remote	38.9	50.1
Gender		
Female	23.6	49.2
Male	26.4	50.2
Immigration status		
Aust.-born students	25.6	49.7
First-gen. students	23.1	49.5
Foreign-born students	25.4	50.3
Missing	33.0	49.1
Home language		
English	24.1	49.6
Language other than English	33.5	50.8
Missing	39.3	52.6
Parents highest education level		
None	51.9	50.9
ISCED 1	61.0	51.2
ISCED 2	44.9	50.0
ISCED 3B, C	36.8	50.4
ISCED 3A, ISCED 4	30.0	50.2
ISCED 5B	24.0	49.6
ISCED 5A, 6	12.8	49.0
Missing	47.8	50.2

Note: ISCED = International Standard Classification of Education.

Table A2 Impact of propensity score weights on continuous variables

Continuous predictors	Raw means		PS-weighted means	
	Low-SES	Not Low-SES	Low-SES	Not Low-SES
PISA maths score	477.74	529.88	516.41	516.90
PISA reading score	463.28	524.27	507.75	509.04
PISA science score	478.17	538.99	522.94	523.80
Highest parental occupational status score	49.22	56.18	54.27	54.42

As can be seen from tables A1 and A2, the propensity score weights are effective in balancing out the proportion of low socioeconomic status individuals in each group. This balancing effect is particularly strong for continuous variables.

Multi-level weight

Given that multi-level models are used for data analysis (that is, a component for individual effects and a component for school effects), it is important that the separate weights for individuals and schools are combined into an overall multi-level weight. The procedure for creating this multi-level weight is based on the methodology outlined by Chantala, Blanchette and Suchindran (2011), in particular, the population weighted iterative generalised least squares (PWIGLS) method A, in which the individual and school weights are multiplied and then divided by the average of the individual weights within a school:

$$mlweight_{i,j} = \frac{fsu\_wt_{i|j} \times psu\_wt_j}{\left(\frac{\sum_{i=1}^{n_j} fsu\_wt_{i|j}}{n_j}\right)}$$

where $fsu\_wt_{i|j}$ is the weight for individual i in school j, psu_wt_j is the weight for school j and n_j is the number of individuals in school j.

This weight is then used as a survey weight. Note that due to issues in the implementation of the mixed procedure within the SAS software, the weight was normalised to ensure estimates are produced with correct standard errors.

Combining the different weights

The four individual weights are combined by multiplication to generate the overall analysis weight.

Appendix B

Multi-level models and results

This appendix briefly outlines the multi-level model used to model the outcomes of interest. The multi-level modelling fitted is a two-level school-effects model; that is, there are data at two levels; in this case, students sit within schools. The responses of interest, Year 12 completion and university enrolment, are binary variables. Tertiary entrance rank scores is a continuous response. The response variables are modelled as a function of both student and school-level variables. For binary outcomes, a logistic mixed model is used (C.1), and for the continuous outcome an ordinary mixed model is utilised (C.2).

The model fitted allows a random intercept for schools and fixed effects for all other school and student-level characteristics:

$$logit(\boldsymbol{y}_{ij}) = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{\epsilon} \qquad \text{(C.1)}$$

where $\boldsymbol{y}_{ij}$ is an binary indicator vector indicating whether student i in school j has completed Year 12 (or enrolled in university), $\boldsymbol{X}$ is the design matrix of fixed effects (both school and individual level), $\boldsymbol{u}$ is the vector of regression coefficients obtained for the corresponding fixed effects, $\boldsymbol{Z}$ is the design matrix of random school effects, $\boldsymbol{u}$ represents the variation in intercepts between schools and $\boldsymbol{\epsilon}$ is the between student (within school) variation. Further, it is assumed that $\boldsymbol{u} \sim N(0, \sigma^2_{sch})$ and $\boldsymbol{\epsilon} \sim N(0, \sigma^2_e)$. In the case of logistic models, σ^2_e is approximated by $\frac{\pi^2}{3} = 3.290$. Results from this model appear below in tables B1 to B3.

For tertiary entrance rank score, the model fitted allows a random intercept for schools and fixed effects for all other school and student level characteristics:

$$\boldsymbol{y}_{ij} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{\epsilon} \qquad \text{(C.2)}$$

where $\boldsymbol{y}_{ij}$ is the tertiary entrance rank scored for student i in school j, $\boldsymbol{X}$ is the design matrix of fixed effects (both school and individual level), $\boldsymbol{u}$ is the vector of regression coefficients obtained for the corresponding fixed effects, $\boldsymbol{Z}$ is the design matrix of random school effects, $\boldsymbol{u}$ represents the variation in intercepts between schools and $\boldsymbol{\epsilon}$ is the between-student (within school) variation. Further, it is assumed that $\boldsymbol{u} \sim N(0, \sigma^2_{sch})$ and $\boldsymbol{\epsilon} \sim N(0, \sigma^2_e)$.

Year 12 completion – model fit statistics

Table B1 Fit statistics for Year 12 completion

Statistic	Value
-2 Res Log Pseudo-Likelihood	29 226.07
Generalised Chi-Square	3 430.97
Generalised Chi-Square/DF	0.73

Table B2 Covariance parameter estimate for Year 12 completion

Variance parameter	Subject	Estimate	Standard error
School Intercepts (σ^2_{sch})	L2_id	0.7447	0.1222

Table B3 Parameter estimates for Year 12 completion

Effect	β	SE	df	t	Pr > |t|
Intercept	2.383	0.099	325	24.01	<0.0001
Student SES	0.082	0.070	4381	1.17	0.2418
Student academic achievement	1.220	0.077	4381	15.83	<0.0001
School quality	0.470	0.091	325	5.14	<0.0001
Student SES by student academic achievement	-0.045	0.056	4381	-0.81	0.4159
Student SES by school quality	-0.150	0.053	4381	-2.84	0.0046
Student academic achievement by school quality	-0.132	0.063	4381	-2.09	0.0366

Note: Year 12 completion is a binary outcome, which requires the use of a mixed model using a logistic link.

Tertiary entrance rank score – model fit statistics

Table B4 Fit statistics for tertiary entrance rank score

Statistics	Value
-2 Res Log Likelihood	28495.0
AIC (smaller is better)	28499.0
AICC (smaller is better)	28499.0
BIC (smaller is better)	28506.6

Table B5 Variance parameter estimates for tertiary entrance rank score

Variance parameter	Estimate
School intercepts (σ^2_{sch})	16.4144
Residual (σ^2_e)	85.5223

Table B6 Regression results for tertiary entrance rank

Effect	β	SE	df	t	Pr > |t|
Intercept	76.72*	0.333	324	230.65	<0.0001
Student SES	0.39*	0.197	3375	2.00	0.0461
Student academic achievement	7.22*	0.250	3404	28.91	<0.0001
School quality	3.87*	0.327	314	11.85	<0.0001
Student SES by student academic achievement	0.63*	0.208	3462	3.02	0.0026
Student SES by school quality	-0.47*	0.193	3072	-2.43	0.0150
Student academic achievement by school quality	0.08	0.243	3142	0.31	0.7535

University enrolment – model fit statistics

Table B7 Fit statistics for university enrolment by age 19 years

Statistic	Value
-2 Res Log Pseudo-Likelihood	25 974.08
Generalised Chi-Square	3 724.92
Generalised Chi-Square/DF	0.79

Table B8 Covariance parameter estimate for university enrolment by age 19 years

Variance parameter	Subject	Estimate	Standard error
School intercepts (σ_{sch}^2)	L2_id	0.3969	0.0696

Table B9 Regression results for university enrolment by age 19 years

Effect	β	SE	df	t	Pr > |t|
Intercept	0.790*	0.065	325	12.11	<0.0001
Student SES	0.162*	0.045	4381	3.59	0.0003
Student academic achievement	1.264*	0.060	4381	21.16	<0.0001
School quality	0.765*	0.063	325	12.08	<0.0001
Student SES by student academic achievement	0.017	0.048	4381	0.36	0.7223
Student SES by school quality	-0.099*	0.043	4381	-2.30	0.0212
Student academic achievement by school quality	-0.008	0.054	4381	-0.15	0.8790

Note: University enrolment status is a binary outcome, which requires the use of a mixed model using a logistic link.