Using Standardized Tests to Predict Access to Higher Education and Measure School Value-Added*

Brendan Houng[a] and Moshe Justman[a, b]

[a] Melbourne Institute of Applied Economic and Social Research, the University of Melbourne

[b] Department of Economics, Ben Gurion University of the Negev

Abstract

We demonstrate, out of sample, that Australia's nationally standardized grade-nine test scores, combined with demographic and socio-economic covariates, provide accurate probabilistic predictions of students' prospective access to higher education. While prior scores have the larger effect, family background has a substantial further impact on access, and our findings indicate where these effects are largest. Among larger schools, out-of-sample predictions based solely on student intake account for 87% to 89% of the variance in school-level success rates; and value-added indicators derived out-of-sample explain a further 5-6 percentage points—over 40% of the remaining variance.

JEL classification: I21, I24, I28

Keywords: Access to higher education, equal opportunity, standardized tests, longitudinal analysis, predicting educational achievement, school effects, NAPLAN, ATAR, VCE, Victoria, Australia

* This paper is based on research commissioned by the Victorian Department of Education and Early Childhood Development (DEECD), now the Department for Education and Training (DET). It uses unit record data from Australia's National Assessment Program—Literacy and Numeracy (NAPLAN) from an administrative data collection created by the Victorian Curriculum and Assessment Authority (VCAA), which delivered the Victoria Certificate of Education (VCE) assessments, matched student data and provided the data to DEECD. We thank Chris Ryan, Gary Marks, Paul Jensen and seminar participants at the University of Melbourne and DET for their helpful comments. The findings and views reported in this paper are those of the authors and should not be attributed to DET or any branch of the Victorian or Australian Commonwealth government.

Corresponding author: Moshe Justman (mjustman@unimelb.edu.au)

1. Introduction

Students, parents and teachers often find it difficult to form realistic prior assessments of a student's prospective access to higher education. The Longitudinal Survey of Australian Youth (LSAY) reports that almost 40% of the students surveyed who reached grade twelve but chose not to sit for tertiary admissions tests had reported earlier, in grade nine, that they planned to attend university (Cardak and Ryan, 2009, Table 3). Better information on their future chances of success, based on their prior academic performance to grade nine, might have helped these students form a more realistic assessment of the work needed to achieve their challenging goals or encouraged them to explore alternative educational and career paths.

In this paper we show that nationally standardized ninth-grade numeracy and literacy test scores from Australia's National Assessment Program—Literacy and Numeracy (NAPLAN), used in conjunction with student-level demographic and socio-economic covariates, can provide highly accurate probabilistic predictions of future access to higher education. Longitudinal data on two full cohorts of ninth grade students in Victoria, in 2008 and 2010, links their performance on NAPLAN reading and numeracy tests at the beginning of grade nine to the Australian Tertiary Admission Ranks (ATAR) they achieve (or fail to achieve) three and a half years later, when they sit for Victoria Certificate of Education (VCE) tests at the end of grade twelve. ATAR is the primary academic criterion that determines admission to university programs, and we use three binary outcome variables that allow us to distinguish between different degrees of access: achieving an ATAR of 50 or better ("ATAR50"), 70 or better ("ATAR70"), and 90 or better ("ATAR90").[1] Each of these indicators applies to the full ninth-grade cohort, including the many students who do not go on to complete a VCE and earn an ATAR.

We first use the earlier 2008 cohort to regress each of these indicators on students' standardized grade nine scores and on their socio-economic and demographic characteristics; and then use the coefficients from these regressions to form probabilistic predictions, out of sample, for students in the 2010 cohort. We then assess the accuracy of our out-of-sample predictions for each of our three ATAR levels by grouping together students in the 2010 cohort with the same predicted values of success at that level, and comparing the actual success rate of each such group, in 2013, with its predicted rate of success.[2] We find that these predicted values closely match the actual values, except for students with very high predicted success rates, where our predictions overstate the actual rates. Regressing

[1] ATAR ranks students in relation to the full cohort. Thus an ATAR of 50 places one at the median of the entire cohort, well below the median of students with an ATAR. Values below 50 are of little value for competitive admissions.

[2] For example, we group all 2010 students for whom the predicted probability of achieving an ATAR of 70 or better is 38%, and compare the share of this group who actually achieve an ATAR of 70 or better to their predicted rate of 38%.

actual on predicted rates over 100 percentiles, we obtain out-of-sample R^2 values of 0.99 for ATAR50, 0.98 for ATAR70 and 0.90 for ATAR90. This attests to the high level of year-to-year consistency in both NAPLAN scores and ATAR values as well as to the mutual consistency of NAPLAN and VCE assessments. It demonstrates the practical scope for using NAPLAN scores to inform students' decisions on their future education and career paths while still in the ninth grade.[3]

Our analysis also highlights the substantial impact of family background on student outcomes in senior secondary school, after controlling for ninth-grade achievement levels. This is consistent with earlier studies, among them Lamb et al. (2004), Le and Miller (2004, 2005) and Cardak and Ryan (2009) who similarly find that SES affects tertiary admissions scores after controlling for prior scores. We go beyond these studies in describing the distribution of these effects for students with varying prior abilities and different goals, which allows us to identify students for whom these affects are substantially larger than average. They are students in the middle of the distribution of NAPLAN scores, aiming for realistic goals that are within their reach. Students who did very well or very poorly in grade nine are less affected by their parents' socio-economic background, as are students who set their sights much too high or much too low. This information can help schools efficiently target their limited resources at students who are most likely to benefit from additional support.[4]

We then apply these predictions to analyse school-level success rates. We find that observable dimensions of student intake—academic, socio-economic and demographic—explain 78% to 80% of school-level success rates in achieving different levels of access to higher education, out of sample, over all 647 schools matched over the two cohorts (accounting for 97% of students). Among the 284 larger matched schools with at least 100 students in grade nine (75% of all students) we find that student intake explains 87% to 89% of the variance in school-level success rates for each of our binary indicators. This very large fraction highlights the importance of carefully controlling for student intake in using student outcomes to assess secondary school performance—all the more so if one allows that unobservable dimensions of student intake may further affect outcomes.[5]

[3] This follows on earlier longitudinal, mostly survey-based studies that demonstrated the power of grade nine test scores to predict outcomes in the transition from secondary to tertiary education. Among these are Marks and Fleming (1999), Marks, Fleming, Long and McMillan (2000), Marks, McMillan and Hillman (2001), McMillan and Marks (2003) Lamb, Rumberger, Jesson, and Teese (2004), Le and Miller (2004, 2005) and Cardak and Ryan (2009).

[4] Cardak and Ryan's (2009) further finding that SES does not directly affect enrolment after controlling for admissions scores, indicates the importance of intervening before ATAR is determined.

[5] Thus Homel, Mavisakalyan, Nguyen and Ryan's (2012) analysis of survey data finds that risky behaviors not indicated in our administrative data are significant predictors of secondary school completion.

Our school-level analysis also indicates that students attending schools with an academically strong student population have significantly higher average success rates than comparable students in weaker schools. Thus, for example, our out of sample analysis of schools with 100 students or more indicates that a school populated by students with a predicted 30% success rate of achieving an ATAR of 70 will have on average a 30% success rate but a school populated by students with a predicted 60% success rate will have on average a 70% success rate. The cause for this is not identified in our data—it may be direct peer effects, the preference that better teachers have for teaching better students, the ability of well-funded schools to attract better students, or other reasons,[6] but it does suggest that parents seeking schools that will help their children achieve better ATAR outcomes are right to prefer schools with higher (raw) success rates.

Finally, we add to these regressions school-specific value-added indicators calculated from the 2008 cohort as the difference between school-level predicted and actual success rates. They have highly significant coefficients throughout, ranging in magnitude from 0.60 to 0.81. Regressing over all 647 schools, we find that including these value-added measures raises the share of explained variance by 7-10 percentage points, to between 85% and 89%; and when we limit our attention to the 284 larger schools with at least 100 students in the cohort, the share of explained variance rises by 5-6 percentage points to between 92% and 95%. Thus our value-added indicators explain 31% to 47% of the residual variance for the full population of schools; and 41% to 53% of the residual variance among schools with 100 students or more in grade nine.

The significant, substantial effect of school-level value-added measures in predicting school performance out of sample supports their validity as stable indicators of schools' ability to promote access to higher education, while at the same time recognizing that these effects are an order of magnitude smaller than the effect of student intake. Thus our findings accord with previous studies, which found relatively little variation between schools in retention and tertiary entrance rates, after controlling for student characteristics (e.g., Marks et al., 2007; Marks, 2010, 2014, 2015; Le and Miller, 2004; Polidano et al., 2013; Cardak and Vecci, 2013), but go beyond these studies in showing that school value-added explains a substantial part of this residual variance out of sample.

As noted above, our analysis follows on previous longitudinal studies of Australian youth, most of which used LSAY data to analyse individual and school-level performance in the transition from secondary to tertiary education in Australia.[7] These studies similarly control for prior ability,

[6] Manski (1993) is the seminal contribution on the difficulty in separating these effects.

[7] Studies in this vein that used other data include the earlier work of Prior and Beggs (1989) and Vella (1999), which used the Australian Longitudinal Survey; and the recent work of Marks (2014, 2015a, 2015b), which uses the same linked data used

demographic characteristics and SES in analysing success measures such as retention to grade twelve, self-reported Equivalent National Tertiary Entrance Ranks (ENTER, a forerunner of ATAR, and participation in higher education. While all of these studies find, as we do, that ninth-grade scholastic achievement and socio-economic indicators predict student success in the transition from secondary to tertiary education, the present study has several methodological advantages that allow it to achieve more accurate predictions and draw sharper distinctions.

The first is our use of probabilistic outcome indicators that explicitly describe the conditional distribution of ATAR outcomes rather than provide point estimates. These indicators effectively combine retention with different levels of ATAR achievement thus taking full advantage of the comprehensive longitudinal data we analyse. This allows us to obtain a more nuanced indication of success than studies that focus on grade-twelve retention or on tertiary enrolment while avoiding the selection bias that arises in studies that exclude a disproportionate fraction of weaker students without ATAR or ENTER score values. In addition, these binary indicators are not sensitive to extreme values, where standardized test scores are less reliable in predicting performance.

A second key advantage is our use of administrative longitudinal data on two full cohorts, which substantially reduces attrition bias, sampling error and measurement error, compared to survey-base studies that use self-reported values, and allows precise statistical estimation within subgroups, resulting in more accurate and more detailed predictions. In particular, it allows us the greater flexibility of estimating separate regression equations within each SES quartile. Moreover, having data on two full cohorts allows us to demonstrate the accuracy of these predictions out of sample, and the year-to-year consistency and mutual compatibility of NAPLAN, VCE and ATAR, showing how NAPLAN scores can be usefully incorporated in planning students' education and career paths; and it enables us to demonstrate the validity of school-level measures of value-added as stable predictors of a school's future value-added.

The rest of the paper is organized as follows: Section 2 describes the data; Section 3 analyses the conditional distributions of our three twelfth-grade success indicators predicated on ninth-grade scores and student covariates; Section 4 demonstrates the accuracy of individual predictions out of sample; Section 5 forms and tests school-level predictions; and Section 6 concludes.

here to regress retention rates, study score aggregates (from which ATAR values are derived, for students with study scores) and tertiary enrolment rates on ninth-grade NAPLAN scores and student covariates.

2. The Data

Our population comprises two full cohorts of ninth-grade students in Victoria in 2008 and 2010, omitting a few hundred students in each cohort with reported ages younger than 14 or older than 16. For each student in the two cohorts, we have NAPLAN reading and numeracy scores, an ATAR outcome (which may be "no ATAR"), and individual demographic and socio-economic covariates. Ninth-grade scores and student characteristics were provided by DEECD; and ATAR outcomes were provided by the Victorian Curriculum and Assessment Authority (VCAA), which linked the two data sets. Summary statistics are presented in Table 1.

NAPLAN tests are administered annually by the Australian Curriculum, Assessment and Reporting Authority (ACARA) in May to all students in Grades 3, 5, 7 and 9. Students missing numeracy and/or reading scores are retained in the sample and the scores are marked as missing. In all our analyses we use scores that are standardized within each cohort. ATAR values in Victoria are derived by the Victorian Tertiary Admissions Centre (VTAC) from study scores, which are based in equal measure on teacher assessments and on student performance on state-wide VCE tests administered by VCAA at the end of twelfth grade. These scores are scaled for subject difficulty and combined to produce Tertiary Entrance Aggregates (TEA), which are then ranked nationwide to produce ATAR values. VCAA linked the 2011 VCE outcomes to the 2008 ninth-grade NAPLAN scores, and the 2013 VCE outcomes to the 2010 ninth-grade NAPLAN scores. As Table 1 shows, the two cohorts are similar in size and composition. There are slightly more students aged 16, and fewer aged 14 in the 2010 cohort; there are more students missing NAPLAN scores in the 2010 cohort, by 2 percentage points; and parents in the 2010 cohort are slightly better educated.

In calculating the shares achieving ATAR values of 50, 70 and 90 we count all students in the ninth-grade cohort not recorded as having achieved a VCE, as not having achieved an ATAR of 50 or better. Consequently, the rates reported in Table 1—respectively, 45.5%, 27.9% and 9.1%, below the target rates of 50%, 30% and 10%—are biased downwards slightly. Of students counted as not having achieved an ATAR of 50 or better, only 30% are positively recorded as such in our twelfth grade data; the remainder are students in the ninth-grade cohort who are absent from our twelfth-grade VCE data. These are mostly students who dropped out of school before sitting for their VCE exams—and therefore should be counted as not having achieved an ATAR of 50 or better—but they also include students leaving Victoria between ninth and twelfth grade, many of whom are likely to have achieved an ATAR of 50 or better elsewhere; students held back a year or skipping a year between grades nine and twelve, many of whom will have sat for their VCEs a year earlier or later; and a small number of students who opt for the equivalent International Baccalaureate Diploma (IBD).

Table 1. Descriptive statistics, ninth-grade students in Victoria, 2008 and 2010

	2008		2010	
Demographics				
Number of students		67,867		67,608
% male		51.3		51.2
% aged 14 / 15 / 16		20 / 74 / 5		18 / 76 / 6
% language background other than English (LBOTE)		25.2		25.3
% Aborigine and Torres Straits Islanders (ASTI)		1.1		1.2
NAPLAN scores				
Mean Numeracy Score (standard deviation)		592.8 (69.8)		595.1 (71.0)
Mean Reading Score (standard deviation)		586.2 (67.4)		583.5 (65.5)
% missing numeracy score		9.1		11.3
% missing reading score		9.3		11.2
% missing both scores		6.7		8.3
Parents' education,% in category	*Mother*	*Father*	*Mother*	*Father*
Not stated / unknown, %	13.7	28.2	11.5	26.9
9 years or less	7.4	5.7	6.8	5.5
10 years	10.2	7.0	9.4	6.9
11 years	11.3	6.4	10.6	6.2
12 years	10.5	7.6	11.0	7.8
Certificate I-IV	19.2	21.5	21.1	22.0
Diploma/ Advanced Diploma	11.3	8.3	11.8	8.7
Bachelor's degree or more	16.3	15.3	17.7	15.9
Parents' occupation, %in category	*Mother*	*Father*	*Mother*	*Father*
Senior manager or professional, %	11.4	12.5	11.7	12.2
Other business manager	15.8	17.9	16.0	17.3
Tradesmen / sales	18.2	17.4	19.1	17.8
Machine operator / hospitality worker	18.3	16.5	18.5	16.3
Has not worked in past 12 months	23.8	8.4	23.3	9.1
Not stated	12.4	27.3	11.5	27.1
ATAR outcomes, % successful of the ninth-grade cohort				
Achieved an ATAR of 50 or more		45.5		44.3
Achieved an ATAR of 70 or more		27.9		27.4
Achieved an ATAR of 90 or more		9.1		9.1
Mean ATAR for students with an ATAR $\geq$ 50		75.4		75.5

In Appendix A, we assess the number of missing students in each group, assign plausible success rates to the three groups, and add these to the observed rates. The revised values we obtain— 49.1%, 30.5% and 10.1%—are much closer to the target rates than those reported in Table 1. We use these adjusted rates to correct the estimated and predicted success rates in the following section, where indicated, multiplying observed rates by a factor of 1.079, 1.093 or 1.110, respectively for ATAR50, ATAR70 and ATAR90. We do not make these adjustments in testing our predictions out of sample, in the following sections, as the same correction would apply to both predicted and actual values, leaving the fit unchanged.

3. Distribution of Conditional ATAR Success Rates

We begin with a graphic description of ATAR success rates conditioned on ninth-grade NAPLAN scores and SES quartiles, for our in-sample cohort. This highlights both the strong link between ninth-grade NAPLAN scores and ATAR outcomes, and the substantial effect of SES after controlling for prior achievement. We then turn to a probit analysis of our three ATAR success indicators, regressing each on NAPLAN scores and on a set of student covariates within SES quartiles. This provides estimates of differences in achievement between demographic categories, controlling for ninth-grade scores and student covariates; and it yields individual predicted probabilities of success for each measure. In the following sections, we use these regression coefficients to predict individual and school-level success rates out of sample, on our 2010 cohort, and compare these predictions to actual success rates.

3.1 Graphic description of achievement levels conditioned on NAPLAN scores and SES

To describe the conditional distribution of ATAR success rates predicated on NAPLAN numeracy and reading scores and on SES, we define four SES categories based on parents' education and occupation (see Appendix B for details); as they are roughly equal in size, we refer to them as quartiles. Table 2 presents category frequencies in the population, NAPLAN averages and specific success rates for each of our three binary indicators. It highlights the strong positive link between SES, on the one hand, and NAPLAN scores and ATAR outcomes, on the other hand.

Table 2. NAPLAN standardized scores and ATAR success rates (%), by SES category, 2008 cohort

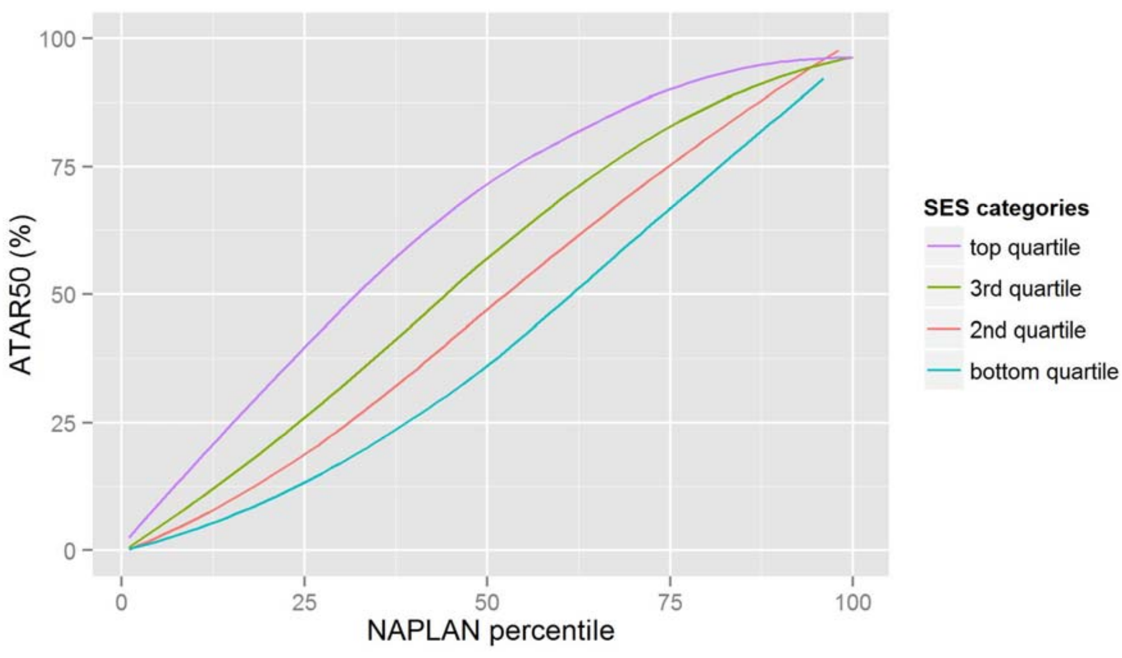
SES category	*N*	*NAPLAN numeracy*		*NAPLAN reading*		*ATAR ≥ 50, %*	*ATAR ≥ 70, %*	*ATAR ≥ 90, %*
		% missing	*Average score*	*% missing*	*Average score*			
Weakest	17,922	14.8	-0.43	15.6	-0.39	25.5	12.2	2.7
2	16,177	7.8	-0.19	8.0	-0.17	40.2	20.0	4.3
3	17,760	7.1	0.06	7.4	0.09	55.6	34.7	10.5
Strongest	16,008	6.0	0.50	5.8	0.55	77.3	57.1	23.9
Total	67,867		0		0	49.1	30.6	10.1

NAPLAN scores are standardized within the cohort. ATAR50, 70 and 90 raw success rates are adjusted upwards to account for students leaving Victoria, skipping or repeating a grade, or earning an equivalent IBD (Appendix A). See Appendix B for definitions of SES categories.

Figures 1 to 3 describe success rates for each of our three indicators, as a function of a student's NAPLAN rank in grade nine, separately for each SES quartile.[8] This shows the effect of SES on ATAR outcomes in senior secondary school, after controlling for grade-nine achievement.

Figure 1 describes success rates in achieving an ATAR of 50 or better. While the top curve is concave and the bottom curve is convex all four curves are nearly parallel between the 25th and 75th percentile of NAPLAN scores, with about the same difference in success rates—about 50 percentage points—between these points, for all four SES quartiles. The difference in success rates between SES categories is greatest at the median of NAPLAN scores, over 30 percentage points. Put differently, a student in the lowest SES quartile and the 65th NAPLAN percentile has about the same chance of achieving an ATAR of 50 or better as a student in the top SES quartile at the 30th NAPLAN percentile. This effect is greatly diminished at the extremes: students ranked very low in the ninth grade have little chance of success while the very strongest students have a high probability, whatever their SES.

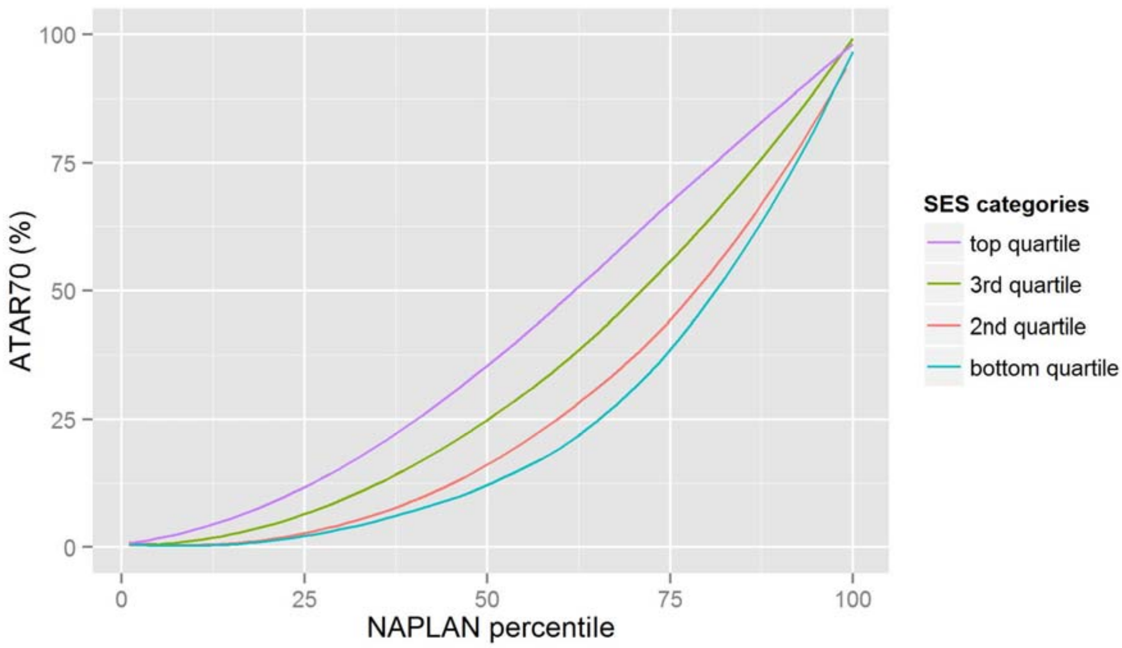
Figure 1. ATAR50 Success Rates by NAPLAN Percentile and SES



[8] We ranked students by the average of their standardized test scores in numeracy and reading and divided them into percentiles. (If either score was missing, we reduced the other score by 20%; the 6.7% of students with both scores missing are not included in these graphs.) Then, for each binary indicator, we computed the frequency of success for each SES category within each percentile. These frequencies were then adjusted upwards to account for students leaving Victoria, skipping or repeating a grade, or earning an equivalent IBD (Appendix A). The graphs were plotted with ggplot2 in R and the points were smoothed by local regression (LOESS).

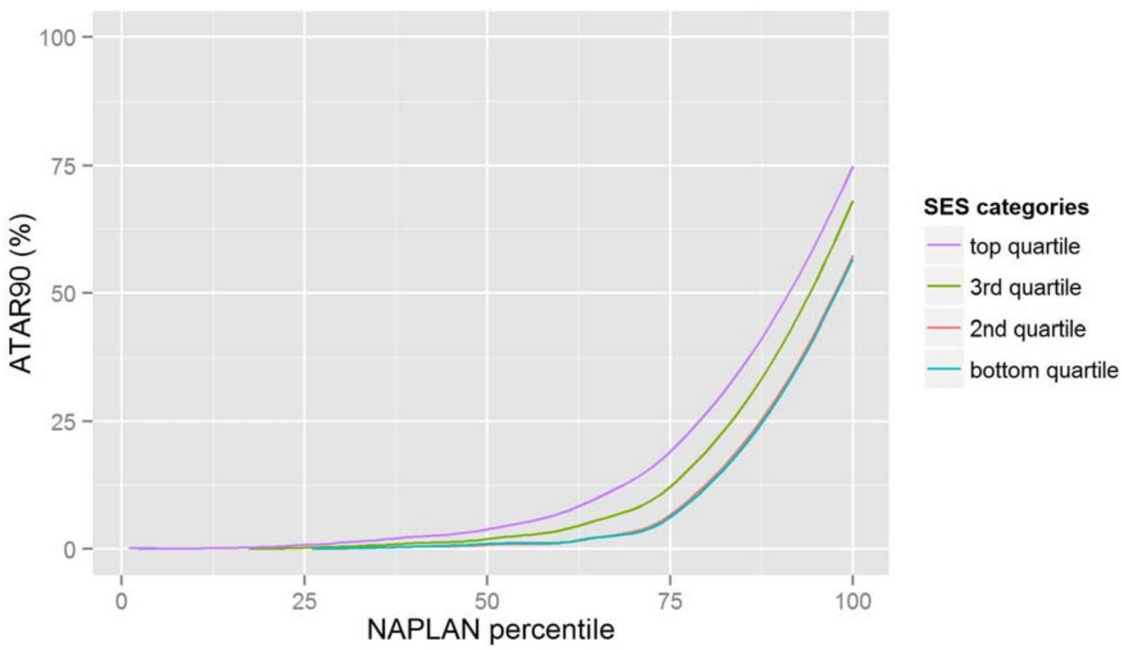
In Figure 2, which describes success rates in achieving an ATAR of 70 or better, all four curves are convex, but the degree of convexity varies markedly. Consequently, while the difference in success rates between the 25th and 75th percentile of NAPLAN scores remains about 50 percentage points for the top SES quartile, it is appreciably smaller, about 35 percentage points, for the bottom quartile. The difference in success rates between SES categories is greatest at the 75th percentile of NAPLAN scores, again over 30 percentage points, and again this difference vanishes at the extremes.

Figure 2. ATAR70 Success Rates by NAPLAN Percentile and SES



Finally, Figure 3 shows very different patterns for the more ambitious goal of achieving an ATAR of 90 or better. All four curves coincide with the horizontal axis for NAPLAN scores in the bottom 30% and there is hardly any increase below the median. At the 75th NAPLAN percentile, even students in the top SES quartile have less than a 20% chance of success. Most of the increase occurs past this point, but even at the very top, success is not assured: even among students in the top 1% of NAPLAN scores and in the top SES quartile, only 70% achieve an ATAR of 90. Doing well on ninth-grade NAPLAN tests is a necessary condition for achieving an ATAR of 90, and a stronger SES helps, but even when combined they do not ensure success.

Figure 3. ATAR90 Success Rates by NAPLAN Percentile and SES



Graphs such as these indicate the levels of effort and support needed to achieve different levels of access to higher education, and can thus help inform the decisions of secondary-school students, their parents and their schools in setting educational and career goals. In addition, the information they provide on the varying impact of socio-economic background, after controlling for achievement in ninth grade, suggests there is considerable scope for schools directing targeted efforts to compensate for socio-economic disadvantage, and provides a useful indication of where these efforts are likely to be most effective in helping disadvantaged students achieve broader access.

3.2 Regression analysis

We estimate probit regressions for each of our outcome variables separately within each SES quartile, regressing each of our binary outcome variables on six NAPLAN score variables (standardized ninth-grade numeracy and reading scores, these scores squared, and indicators for missing scores in each domain);[9] on a set of demographic indicators, for age, gender, English Speaking Background (ESB) or language background other than English (LBOTE), Aborigine and Torres Straits Islanders (ATSI), and interactions between LBOTE and gender and between ATSI and gender; and on a set of parents'

[9] Adding cubic and quartic terms in the NAPLAN scores had no effect. NAPLAN achievement enters these regressions as standardized scores, and so the shape of the regressions does not match Figures 1-3, where the axes are NAPLAN ranks.

education and occupation categories. Our main purpose in estimating these in-sample regressions is to use them to form individual, out-of-sample predictions of success for the 2010 cohort, and compare them to actual outcomes individually and at the school level. A second purpose is to use these regressions to derive in-sample value-added measures for each school and compare them with out-of-sample value-added across schools. In addition, the regressions provide indications of how other student covariates affect the predicted success probabilities, conditioned on ninth-grade test scores and SES.[10] We briefly highlight here selected effects for female students, LBOTE students, ATSI students and students aged 16.

We find a significant advantage for female ESB students over male ESB students; considering the full cohort (rather than just students with an ATAR) avoids the downward bias that results when higher male dropout rates are ignored. The gender effect is measured as the difference between female and male probabilities of success, and therefore generally increases with their level, hence with SES. However, for ATAR50, the greatest differences are observed in the weaker SES categories. We interpret this as reflecting the greater impact of socio-economic disadvantage on the academic achievement of young males (Goldin, Katz, and Kuziemko, 2006).

LBOTE students do better than ESB students, on average, after controlling for prior scores and SES. However, these averages range over diverse groups and may vary widely among them. We observe the largest effect in the lower SES quartiles, where parental education and occupation may not accurately reflect the cultural background of more recently arrived immigrants; in the higher SES quartiles, LBOTE outcomes are more similar to those of the general population. We also find a positive interactive term for male and LBOTE, which indicates that gender differences are smaller for LBOTE students than for ESB students.

The regressions highlight the substantial disadvantage of ATSI students, after controlling for their lower prior scores and lower SES. As the share of ATSI students in the population is small, few have a strong academic or an advantaged SES background, and few achieve strong ATAR outcomes. Consequently, the ATSI effect is estimated with less precision than other effects, and for some sub-categories could not be estimated at all. Finally, we note the disadvantage of students aged 16 in grade 9, again controlling for ninth-grade scores and parental background. To the extent that their older age difference is attributable to some of them having been held back a year this suggests that the difficulties causing this were not entirely resolved.

[10] A full set of estimated marginal effects from these regressions, for the entire 2008 cohort and within each SES quartile, are presented in Tables C1-C3 in the Appendix.

4. Goodness-of-fit out of sample, individual outcomes

We now use these regressions to predict out-of-sample success probabilities for the 2010 cohort and compare them to actual outcomes. We first apply the coefficients from the 2008 regressions to the 2010 cohort to form individual probabilistic predictions for each of our binary variables with regard to ATAR outcomes in 2013. We then aggregate these predictions by their values, for each success indicator, and compare them to actual success rates. Thus, for example, we group together all 2010 students with a predicted success probability in the interval between 53.00 and 53.99, and compare the share of these students actually achieving an ATAR of 70 or better to their average predicted probability (which will fall within that interval). We make no adjustments to the data for inessential attrition as we assume that both predicted and actual outcomes are affected by the same bias. The results are presented in Figures 4, 5 and 6. We added to each point prediction a 95% confidence interval.[11] The variation in the number of observations for each percentile is evident from the varying size of the confidence interval; it is especially large for ATAR90, where many more students have a low probability of achieving an ATAR of 90 or better, than have a high probability.

Figure 4 presents the results for ATAR50. Except for predictions in the top 10%, where predicted values are significantly above actual rates, the fit is clearly very close, with almost all actual values falling within the quite narrow confidence intervals. Figure 5 presents a very similar picture for an ATAR of 70. Again actual success rates for the top 10% are significantly below predicted rates, but for lower values, almost all the actual rates fall within their confidence intervals. In Figure 6, confidence intervals fan out at the top end of the distribution as the number of observations falls (few students have very high predicted rates of success in achieving an ATAR of 90 or better), and again, at the top end we find actual success rates falling below predicted rates, while in the middle range actual success rates slightly exceed predicted rates. Below an 80% predicted success rate almost all observed success rates fall within their confidence intervals, which are however considerably wider than for ATAR50 and ATAR70.

[11] Under the maintained hypothesis, the number of successes among n students with predicted success p distributes binomially with mean np and variance $np(1-p)$ so that the standard deviation of the success rate is $[p(1-p)/n]^{0.5}$.

Figure 4. Goodness of fit, out of sample, in achieving an ATAR of 50

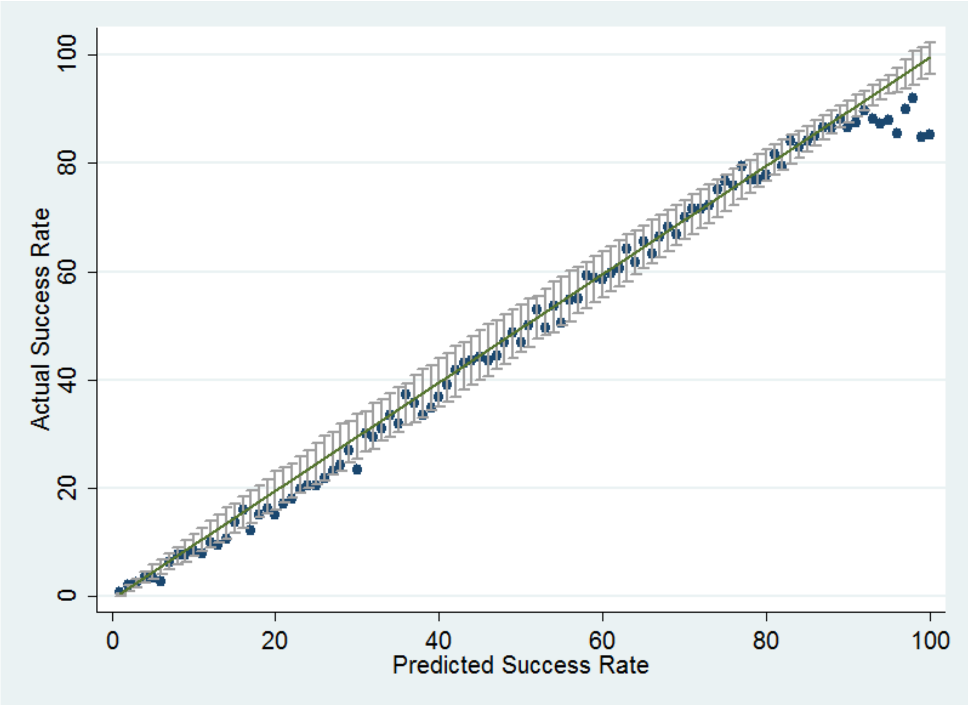


Figure 5. Goodness of fit, out of sample, in achieving an ATAR of 70

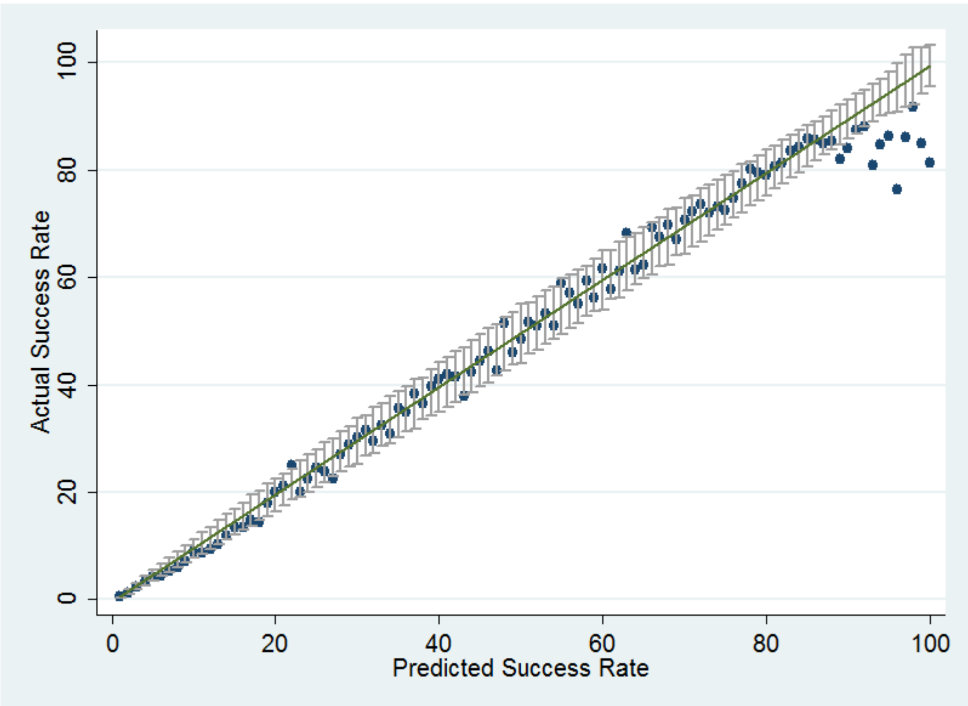
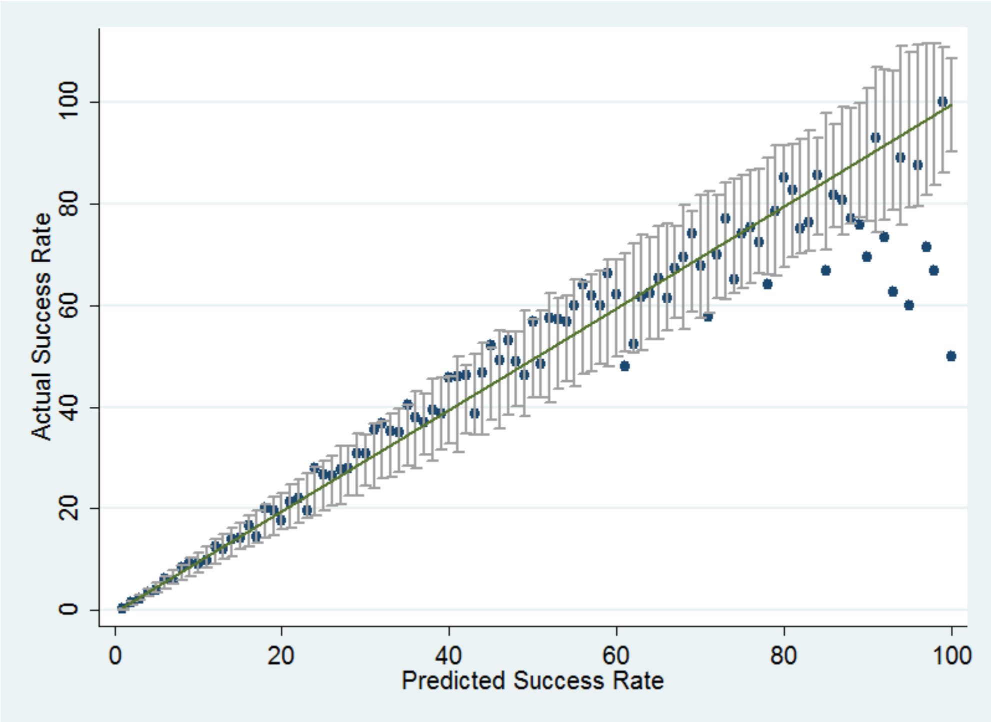


Figure 6. Goodness of fit, out of sample, in achieving an ATAR of 90



To quantify the goodness of fit evident from Figures 4-6, we regressed actual success rates on predicted probabilities across the 100 percentile points for each success indicator. The results are presented in Table 7. The R^2 values for ATAR50 and ATAR70 are very close to 1.00. The over-prediction of success rates at the high end of predicted rates pulls the slopes down, below one. [12]

Table 7. Out-of-sample goodness of fit, individual data

	Estimated slope	Standard error	R^2	*N*
ATAR50	0.983	.009	.99	100
ATAR70	0.952	.013	.98	100
ATAR90	0.837	.028	.90	100

Students in the 2010 cohort were grouped by predicted probability derived from regressions over the 2008 cohort into 100 percentile groups, and group-level actual success rates were regressed on predicted rates.

[12] When the regressions are limited to the bottom 90 percentiles for ATAR50 and ATAR70, and to the bottom 80 percentiles for ATAR90, the slopes range between 1.02 and 0.98 and the R^2 improve further—to 0.97 for ATAR90.

5. School-level outcomes

In this section, we use our individual out-of-sample predictions of ATAR success rates to form school-level predicted success rates for the 2010 cohort, by averaging the individually predicted rates over all students in each school. We then regress the actual school-level success rates on the predicted rates, and interpret the share of unexplained variance from these regressions as an upper limit on the extent to which the variance in ATAR performance can be attributed to schools, after controlling for student intake. We then add to the regressions school-specific value-added measures drawn from the 2008 cohort: each school's actual success rate in 2011 less its predicted rate. Two sets of results are presented for each binary outcome variable and for each of the two regression specifications: one set for all 647 schools that could be matched across cohorts (accounting for 97% of the total student population); and another for the 284 matched schools with at least 100 students in grade nine (accounting for 75% of the cohort).

5.1 Explaining school-level success rates by student intake

Results from the first set of six regressions, where school success rates are regressed only on predicted rates, are reported in Table 9; Figures 7 to 9 present scatterplots for the set of larger schools. They highlight two salient features. The first is the large share of variance in school success rates explained by the predicted values, out of sample, for both sets of schools (more for the larger schools); the second is the magnitude of the estimated slopes, which are all greater than one.

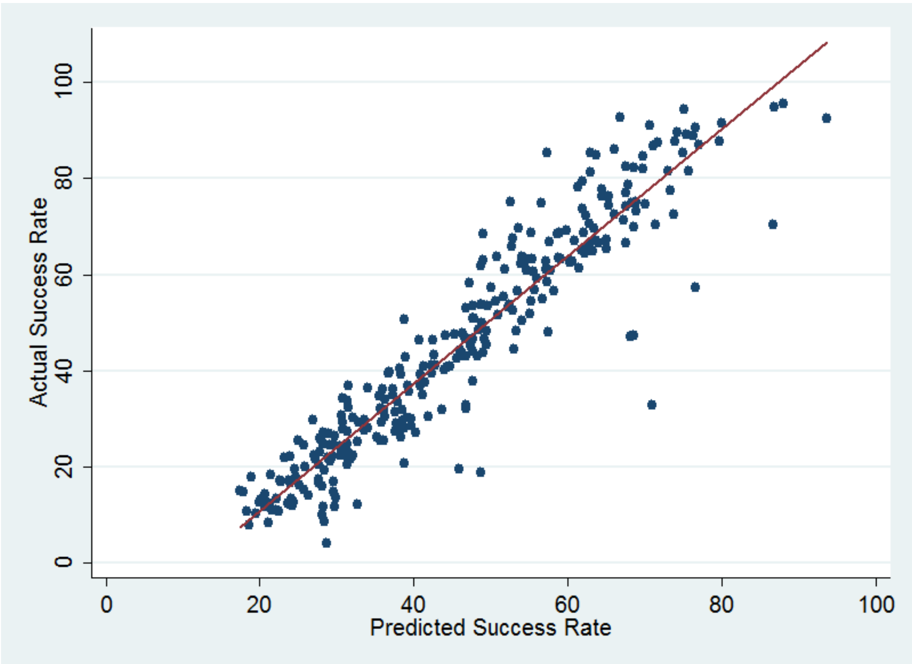
Table 9. School-level regressions of actual on predicted success rates, out of sample

	All matched schools (647 schools, 65,452 students)			Matched schools with cohort $\geq$ 100 (284 schools, 50,467 students)		
	Constant	Slope (standard error)	R^2	Constant	Slope (standard error)	R^2
ATAR50	-13.3	1.24 (.02)	.79	-15.4	1.32 (.03)	.88
ATAR70	-8.31	1.25 (.03)	.78	-9.3	1.32 (.03)	.89
ATAR90	-2.99	1.31 (.03)	.78	-2.7	1.32 (.03)	.87

School-level predicted success rates are averages of individual predicted success rates of all students in the school, taken from probit regressions estimated from within SES quartiles of the 2008 cohort. Actual shares are the success rates of the 2010 cohort.

The predicted values explain 78%-79% of the school-level variance for the full set of schools and 87%-89% for the set of larger schools. This highlights the large extent to which secondary-school performance is shaped by student intake; and the importance of carefully controlling for student intake in using test results to assess school performance. The magnitude of the estimated slopes, all precisely estimated and significantly greater than one, indicates substantial positive school-cohort effects. Taking, for example, the regression equation for ATAR70 for the larger schools, we find that the expected success rate for a school with weaker students, with a 30% predicted rate, is 30.4% while the expected rate for a school with a 60% predicted rate is 70.0%.[13] This may reflect various causes: direct peer effects on learning, a mutual attraction between high scoring students and schools that prepare students for VCE exams more effectively, a preference of better teachers for stronger students, or possibly other factors.[14] Whatever the reason, they suggest that parents seeking schools that will help their children achieve better ATAR outcomes are generally right to prefer schools with higher (raw) success rates.

Figure 7. Out of sample actual *versus* predicted school success rates, ATAR50, schools with 100 students or more in ninth grade



---

[13] Using the coefficients from Table 9: -9.2 + 1.32*30 = 30.4 and -9.2 + 1.32*60 = 70.0

[14] See Lamb et al. (2004) for an analysis of factors that help raise school performance.

Figure 8. Out of sample actual *versus* predicted school success rates, ATAR70

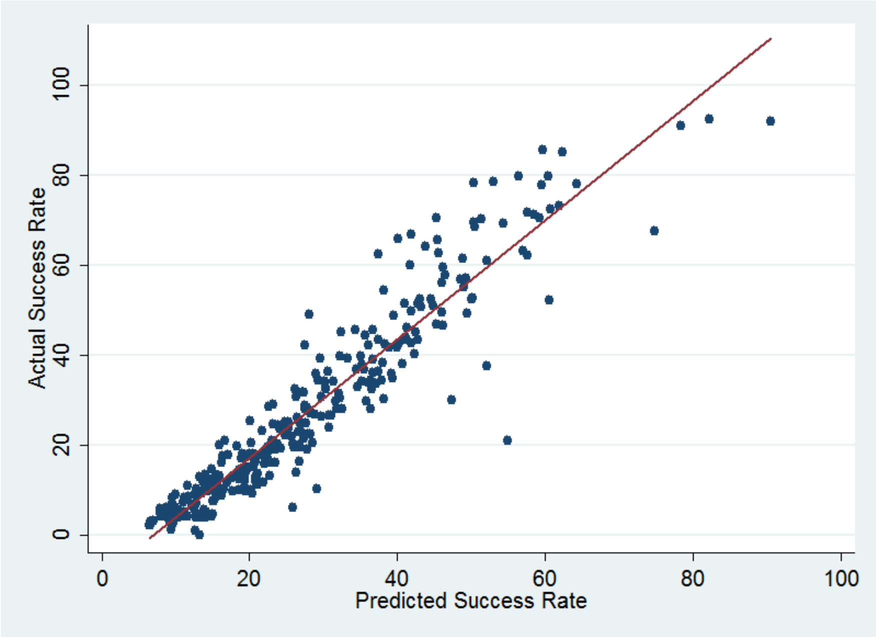
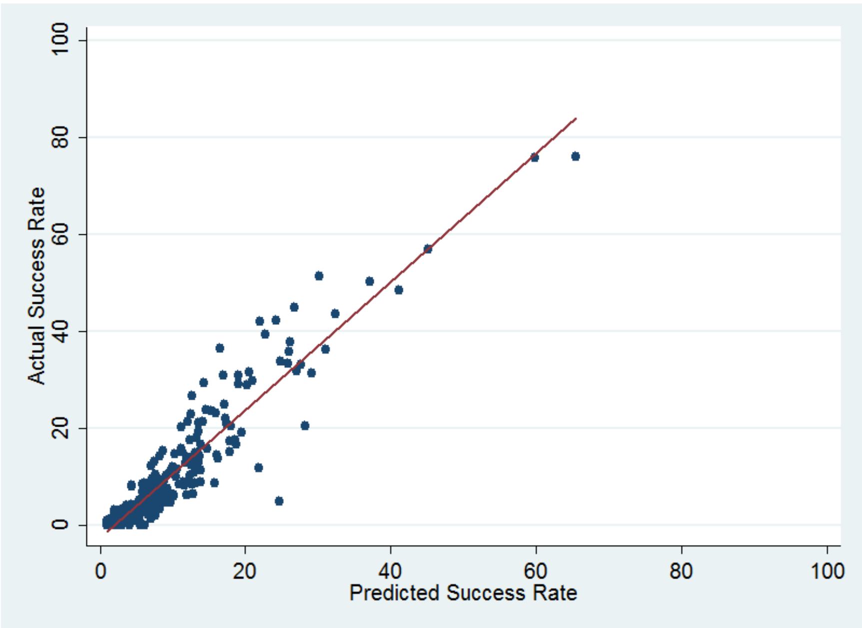


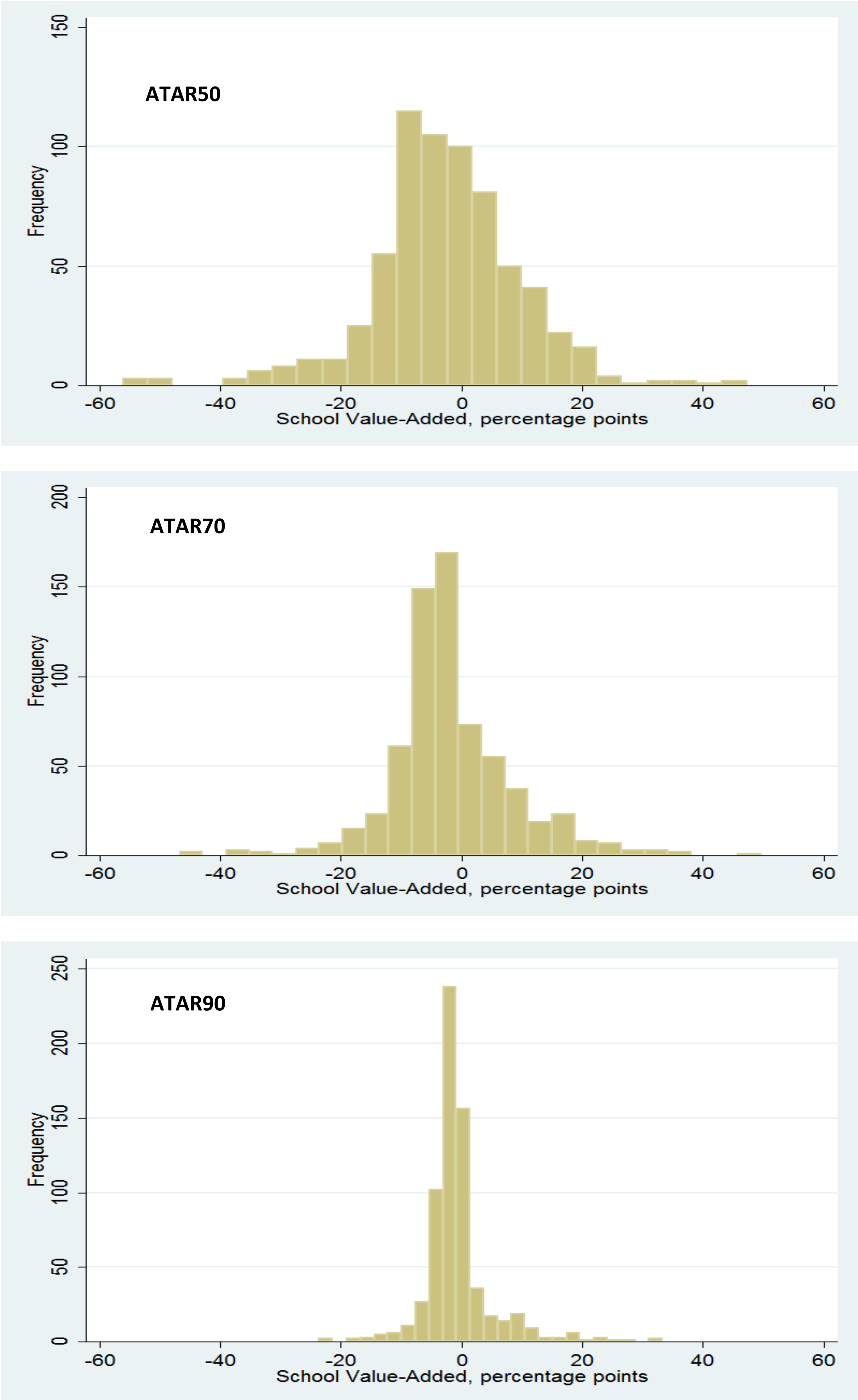
Figure 9. Out of sample actual *versus* predicted school success rates, ATAR90



5.2 Stability and size of school value-added

We first derived school-specific value-added measures for the 2008 cohort, calculated as the difference between actual school-level success rates and predicted rates in 2011, for each success indicator, where a school's predicted success rate is the average predicted success over its students.

Figure 10. School value-added, full sample of schools, 2008 cohort



Histograms of these value-added measures for the full set of schools are presented in Figure 10. These may reflect diverse factors: better teaching, better facilities, peer effects, sorting of students on unobserved variables, and so on.

We calculated correlations between these value-added measures for the 2008 cohort and the school-level residuals from the 2010 cohort regressions reported above, and obtained values of 0.73, 0.66 and 0.63 respectively for ATAR50, ATAR70 and ATAR90, indicating substantial cohort-to-cohort stability in estimated school-level value-added. We then included the 2008 value-added measures in our 2010 cohort regressions. Table 10 presents the regression results.

Table 10. School-level actual success rates regressed on predicted success rates and on school value-added, out of sample

	All matched schools (647 schools, 65,452 students)			Matched schools with cohort ≥ 100 (284 schools, 50,467 students)		
	Predicted rate	Value-added measure	R^2	Predicted rate	Value-added measure	R^2
ATAR50	1.05 (.02)	0.73 (.03)	.89	1.08 (.02)	0.81 (.05)	.94
ATAR70	1.07 (.02)	0.70 (.04)	.86	1.11 (.02)	0.74 (.04)	.95
ATAR90	1.13 (.03)	0.60 (.04)	.85	1.15 (.03)	0.63 (.05)	.92

As indicated by the correlations, all three coefficients of the value-added measure are highly significant, with *t*-values in excess of 10. All three coefficients are significantly less than 1.00, between 0.60 and 0.73 for the regression over all schools; between 0.63 and 0.81 when we limited the regressions to schools with at least 100 students in ninth grade. This suggests that while our value-added measure captures a stable element of school quality, it measures it with some error. The value-added measures also absorb some of the "quality effect", reducing the coefficients of the predicted success rate by about 0.20, compared to the first set of regressions.

Adding the 2008 value-added measures to the regressions increases the share of explained variance by 7-10 percentage points for the full set of 647 schools, to 85%-89%; and by 5-6 percentage points for the 284 larger schools with at least 100 students in ninth grade, to 92%-95%. The measures of value-added taken from the 2008 cohort thus explain between 31% and 47% of the residual variance out of sample, in the 2010 cohort, for the full population of schools; and 41% to 53% of the residual variance for the larger set of schools. This attests to the stability of school-level value-added measures derived (in this manner) from using NAPLAN scores to predict ATAR outcomes. These findings are consistent with previous efforts that found school level variation in outcomes to be small in relation to the impact of student intake, but depart from these efforts in identifying significant, substantially stable school-level effects.[15]

6. Summary

In this paper, we show that standardized ninth-grade test scores in reading and numeracy, from Australia's National Assessment Program—Literacy and Numeracy (NAPLAN), combined with students' demographic characteristics and their parents' education and occupation categories, provide a strong indication of the level of access to higher education these students are likely to achieve three and a half years later, as reflected in their ATAR outcomes. It shows how these scores can be used in conjunction with socio-economic and demographic background variables to help students in ninth grade assess the effort required to achieve different educational goals and reach informed decisions in setting these goals. Prior test scores and socio-economic background are stochastic indicators of access to higher education that indicate the level of effort and support necessary to attain different levels of access.

Longitudinal data on two full cohorts of ninth grade students in Victoria, in 2008 and 2010, links their performance on NAPLAN reading and numeracy tests in the beginning of grade nine to the ATAR values they achieve (or fail to achieve) three and a half years later, when they sit for Victoria Certificate of Education (VCE) tests at the end of grade twelve. We use the 2008 cohort to regress three binary ATAR success indicators on standardized grade-nine scores and on students' socio-economic and demographic characteristics; and then apply the coefficients from these regressions to form out-of-sample predictions for the 2010 cohort of ninth grade students. We then compare these predictions to their actual ATAR outcomes, achieved in 2013.

The three binary outcome variables we use are: achieving an ATAR of 50 or better, an ATAR of 70 or better, and an ATAR of 90 or better. After forming out-of-sample probabilistic predictions for each success indicator, for each student in the 2010 cohort, we group together at each percentile, all students in the 2010 cohort with that predicted probability of success, and compare the actual success rate of each group with its predicted probability.

[15] Correlations of school-level residuals between the two cohorts are 0.73, 0.66 and 0.63 for ATAR50, ATAR70 and ATAR90.

We find that these predicted values closely match the actual values, except for students with very high predicted success rates, where our predictions overstate the actual rates. Regressing actual on predicted rates over 100 percentiles, we obtain out-of-sample R^2 values of 0.99 for ATAR50, 0.98 for ATAR70 and 0.90 for ATAR90. These results attest to the high level of year-to-year consistency in both NAPLAN scores and ATAR values as well as to the mutual consistency of NAPLAN and VCE assessments. It illustrates how NAPLAN scores can be used in practice to inform students' decisions on their future education and career paths while in ninth grade.

Our analysis also highlights the substantial impact of family background on student outcomes in senior secondary school, after controlling for ninth-grade achievement levels; and it identifies those students who are most likely to benefit from such support. They are students in the middle of the distribution of NAPLAN scores, aiming for realistic goals that are within their reach. Students who did very well or very poorly in grade nine are less affected by their parents' socio-economic background, as are students who set their sights much too high or much too low. These findings can help schools direct limited resources efficiently to where these resources are likely to have the greatest impact.

Our school-level analysis of success rates shows that observable dimensions of student intake—academic, socio-economic and demographic—explain, out-of-sample, 78% to 80% of the variance in school-level success rates in achieving different level of access to higher education across all schools. When we limit our attention to larger schools with at least 100 students in ninth grade, student intake explains 87% to 89% of this variance. This large fraction highlights the importance of carefully controlling for student intake in using student outcomes to assess secondary school performance—all the more so if one allows that unobservable dimensions of student intake may further affect outcomes.

Our findings also confirm the well-known empirical observation that students attending schools with an academically strong student population have significantly higher average success rates than comparable students in weaker schools. The cause for this is not identified in our data—it may be direct peer effects, the preference that better teachers have for teaching better students, the ability of well-funded schools to attract better students, or other reasons. But it does suggest that parents seeking schools that will help their children achieve better ATAR outcomes are right to prefer schools with higher (raw) success rates.

Finally, we add to these regressions school-specific value-added measures derived from the 2008 cohort as the difference between school-level predicted and actual success rates in that cohort. The coefficients of these measures are all highly significant. Regressing over all 647 schools, we find that including value-added measures raises R^2 values by 7-10 percentage points, to between 0.85 and 0.89; and when we limit our attention to the 284 larger schools with at least 100 students in ninth grade, R^2 values rise by 5-6 percentage points to between 0.92 and 0.95. These value-added measures explain between 31% and 47% of the residual variance for the full population of schools; and 41% to 53% of the residual variance for schools with at least 100 ninth-grade students.

These findings support the validity of these measures as indicators of schools' success in promoting access to higher education, while at the same time recognizing that these effects are an order of magnitude smaller than the effect of student intake. Thus our findings accord with previous studies, which found little variation between schools in their retention and tertiary entrance rates, after controlling for student characteristics, but go beyond these studies in showing that school value-added is a stable school attribute that explains a substantial part of this residual variance. Again, we are not able to identify the source of these effects, and whether they reflect factors within the school's control, or peer effects, or sorting across schools on unobserved dimensions of student quality, or other factors.

In conclusion, we note that the testing methods underlying NAPLAN and VCE tests in Australia are widely used in many settings, suggesting that the empirical methodology described in this paper to similar data in other countries may reveal similar effects.

References

Cardak, BA, Vecci, J, 2013. Catholic school effectiveness in Australia: A reassessment using selection on observed and unobserved variables. *Economics of Education Review* 37:34–45.

Cardak, BA, Ryan, C, 2009. Participation in higher education: Equity and access. *The Economic Record* 85:443-448.

Goldin, C, Katz, LF and Kuziemko, I. 2006. The homecoming of American college women: The reversal of the college gender gap. *Journal of Economic Perspectives* 20(4):133-156.

Heckman, JJ, Stefano M, 2014. The Economics of Human Development and Social Mobility. *Annual Review of Economics*, *Annual Reviews* 6(1): 689-733.

Helal, M, 2012. Measuring education effectiveness: Differential school effects and heterogeneity in value-added. Working paper, Melbourne Institute of Applied Economic and Social Research. University of Melbourne.

Homel, J, Mavisakalyan, A, Nguyen, HT and Ryan, C, 2012. School completion: What we learn from different measures of family background. LSAY Research Report No. 59. Adelaide, Australia: National Centre for Vocational Education Research.

Houng, B, Justman, M, 2013. Comparing least-squares value-added analysis and student growth percentile analysis for evaluating student progress and estimating school effects. Melbourne Institute of Applied Economic and Social Research Working Paper 7/13. University of Melbourne.

Lamb, S, Rumberger, RW, Jesson, D, and Teese, R, 2004. School performance in Australia: Results from analyses of school effectiveness. Report for the Victorian Department of Premier and Cabinet. Centre for Postcompulsory Education and Lifelong Learning, University of Melbourne.

Le, AT, Miller, PW, 2004. High school graduation in Australia: Do schools matter? *Scottish Journal of Political Economy* 51(2):194-208.

Le, AT, Miller, PW, 2005. Participation in higher education: Equity and access? *The Economic Record* 81(253):152–165.

Manski, C, 1993. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies* 60:531-42.

Marks, GN, 2007. Do schools matter for early school leaving? Individual and school influences in Australia. *School Effectiveness and School Improvement* 18:429–450.

Marks, GN, 2010. What aspects of schooling are important? School effects on tertiary entrance performance. *School Effectiveness and School Improvement* 21:267–287.

Marks, GN, 2014. Reaching Year 12 in Victoria, Australia: Student and school influences, *Educational Research and Evaluation: An International Journal on Theory and Practice*, 20:5, 333-347.

Marks, GN, 2015a. Are School-SES Effects Statistical Artefacts? Evidence from Longitudinal Population Data. *Oxford Review of Education*, DOI: 10.1080/03054985.03052015.01006613

Marks, GN, 2015b. (Provisional Acceptance) School Sector Differences in Tertiary Entrance Performance in Victoria: A Longitudinal Analysis. *Australian Journal of Education*

Marks, GN, Fleming, N, 1999. Early school-leaving in Australia: Findings from the 1995 Year 9 LSAY cohort. LSAY Research Report No. 11. http://research.acer.edu.au/lsay_research/71

Marks, GN, Fleming, N, Long, M, McMillan, J, 2000. Patterns of Participation in Year 12 and Higher Education in Australia: Trends and Issues. LSAY Research Report No. 17, Australian Council for Educational Research, Victoria.

Marks, GN, McMillan, J, Hillman, K, 2001. Tertiary Entrance Performance: The Role of Student Background and School Factors. LSAY Research Report No. 22. Melbourne: Australian Council for Educational Research.

McMillan, J, Marks, GN, 2003. School leavers in Australia: Profiles and pathways. LSAY Research Report No. 31. http://research.acer.edu.au/lsay_research/35

Polidano, C, Hanel, B, Buddelmeyer, H, 2013. Explaining the socio-economic status school completion gap. *Education Economics*, 21, 230–247.

Prior, H, Beggs, JJ, 1989. Influence of Family Background on the Educational and Labour-Force Outcomes of Year 12 School Leavers. *The Australian Journal of Statistics*, 31A, 99–124.

Vella, F, 1999. Do catholic schools make a difference? Evidence from Australia. *Journal of Human Resources*, 34, pp. 208–24.

Appendix A: Students in the ninth-grade cohort missing from the twelfth-grade data.

As noted in Section 2, there are three categories of students in our ninth-grade cohort who do not appear in the twelfth-grade data and are therefore counted as not having achieved an ATAR but who may have achieved an ATAR of 50 or better elsewhere or at a different time: those leaving Victoria between ninth grade and twelfth grade; those held back a year or skipping a year in that time; and those earning the equivalent International Baccalaureate Diploma (IBD).

We first estimate the size of each group.

Students leaving Victoria between Year 9 and Year 12. Total annual departures from Victoria in 2009 amounted to just over 60,000; in Australia as a whole, just under 4% of internal migrants were between the ages of 15-19 (Australian Bureau of Statistics, 2009, *Migration Australia*, cat. no. 3412.0, http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3412.02009-10?OpenDocument viewed 6 November 2013; see tables A1 and A2). Assuming this applies to Victoria and distributes evenly among the five years, there were about 500 departures per cohort each year, or 1500 departures between Year 9 and Year 12, about 2.2% of our cohort. If their success rates are similar to the general population, we are understating the share of ATAR50 by 1.1 percentage points, the share of ATAR70 by 0.7 percentage points, and the share of ATAR90 by 0.2 percentage points.

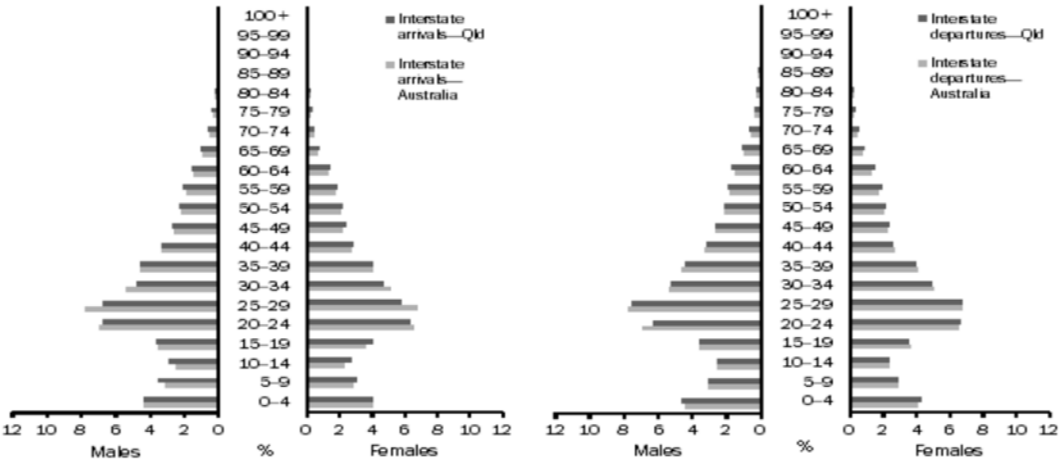
Table A1. Interstate migration flows, 2009-10

5.3 INTERSTATE MIGRATION FLOWS—2009–10(a)

	DEPARTURES FROM:								
	NSW	*Vic.*	*Qld*	*SA*	*WA*	*Tas.*	*NT*	*ACT*	*Total arrivals*(b)
Arrivals to:									
NSW	. .	20 088	35 355	4 989	7 525	2 118	2 969	9 938	82 982
Vic.	22 686	. .	17 594	6 544	8 109	3 185	2 451	2 527	63 096
Qld	42 044	18 605	. .	5 823	8 295	3 193	5 208	3 237	86 405
SA	4 628	5 552	4 548	. .	2 458	634	2 257	624	20 701
WA	8 297	8 422	8 696	2 849	. .	1 443	2 605	879	33 191
Tas.	2 405	2 731	3 341	691	1 660	. .	419	260	11 507
NT	3 320	2 783	4 482	2 007	2 284	315	. .	471	15 662
ACT	10 142	2 360	2 813	762	898	297	595	. .	17 867
Total departures(b)	93 522	60 541	76 829	23 665	31 229	11 185	16 504	17 936	331 411
Net	**–10 540**	**2 555**	**9 576**	**–2 964**	**1 962**	**322**	**–842**	**–69**	. .

Source: Australian Bureau of Statistics, 2009, *Migration Australia*, cat. no. 3412.0, Table 5.3

Table A2. Interstate movers for Queensland and Australia, by age and sex, 2009-10



Source: Australian Bureau of Statistics, 2009, *Migration Australia*, cat. no. 3412.0, Table 5.8

Students held back a year or skipping a year between ninth grade and twelfth grade. NAPLAN data from 2008 to 2011 indicates that on average roughly 0.8% of students in Victoria government schools repeat ninth grade annually (Table A3). The share of students repeating years 10 or 11 should be smaller; we assume it is 0.4% in each year, or 1.6% in total, and that the share of repeaters in private school is no larger.

Table A3. Students repeating ninth grade in Victoria government schools

Year	Number of students in cohort	Students repeating ninth grade	Share
2009	40,794	335	0.82%
2010	39,962	309	0.77%
2011	39,848	373	0.94%
2012	39,486	344	0.87%
total	160,090	1,361	0.85%

Source: Victoria DEECD NAPLAN data, authors' tabulation

The fraction of students with an ATAR of 50 or higher among repeaters is likely to be lower than in the general population, which suggests that adding repeaters who achieve an ATAR in later years would increase our average indicators of success by no more than 0.8, 0.48 and 0.24 percentage points, respectively. We do not have data on the share of students skipping a year between years 9 and 12.

Their share in the population at these ages is likely to be significantly smaller but their success rates are no doubt higher. We assume, arbitrarily, that these effects cancel each other out so that their impact on our success rates is the same as that of repeaters, implying that he total impact of students held back a year or skipping a year between Grades 9 and 12 is, respectively, 1.6, 1.0 and 0.5 percentage points.

Students earning an International Baccalaureate Diploma (IBD). The IBD is an alternative system of high school matriculation. Australia's University Admission Centre (UAC) publishes an equivalence scale equating IB Diploma scores to UAC ranks on the ATAR scale. About five hundred students earn an IB Diploma in Victoria annually, 0.9% of the cohort. Almost all of them achieve a ranking equivalent to an ATAR of 70 or better, and we assume that a third of these achieve an ATAR-equivalent of 90 or better (this is the ratio of ATAR90 to ATAR70 in the population). This implies further increases in our success rates of 0.9, 0.9 and 0.3 percentage points, respectively.

In sum, this implies that we are understating the share of ATAR50 by 1.1 + 1.6 + 0.9 = 3.6 percentage points; of ATAR70 by no more than 0.7 + 1.0 + 0.9 = 2.6 percentage points; and of ATAR90 by 0.2 + 0.5 + 0.3 = 1.0 percentage point. When these are added to the observed frequencies in Table 2, the total frequencies we obtain are 49.1%, 30.5% and 10.1%.

Note that VCAA estimates ATAR success rates as the ratio of successful 19-year olds to the estimated population of 17 year-olds two years earlier. Our numbers, though defined differently, are consistent with theirs.

Appendix B. Definitions of socio-economic variables

We define five categories of parental education from separate indicators of parents' education:

1. Least educated: neither parent above year 10
2. Partial HS-school: at least one parent year 11 or certificate
3. Both missing: both not stated or unknown
4. Full HS: at least one parent Year 12 or diploma
5. Higher education: at least one parent with bachelor degree or more

Similarly, we define a single, joint indicator of parental occupation, with five exclusive categories, from data on father's and mother's occupation types, classified as: senior manager or professional; other business manager; tradesman or sales; machine operator or hospitality worker; unemployed; and not stated or unknown. We combine them into a single parental indicator as follows:

1. "Both unemployed or father's occupation unknown": Both parents are unemployed, or the father's occupation is unknown and the mother is not a manager or professional. When only the father's occupation is unknown we interpret this as indicating an absent father.
2. "One parent not working": The father is unemployed and the mother is not a manager or professional, or the mother is unemployed or has unknown occupation and the father is employed but not as a manager or professional.
3. "Both employed": Father and mother both working but neither is a manager or professional
4. "Both unknown": Father's and mother's occupation category unknown; we interpret this as generally indicating that they are self-employed and do not fall comfortably in any category.
5. "Manager": Father or mother is a manager or professional

Finally, we define four levels of family socio-economic status (SES) based on these five categories of parental education and five categories of parental occupation, as follows:

(1) (lowest) Both education and occupation categories are in the lower range
(2) Both education and occupation are mid to low range and equivalents
(3) Both education and occupation are mid-range or one is high and one is low
(4) (highest) Both high range categories or one high and one mid-range

We set out the specific frequencies of each combination of family occupation and education in Table B2, the definition of family SES categories in Table B3, and the distribution of ninth-grade students by SES category in Table B4.

Table B2. Cross tabulation of occupation and education

	Family education categories					
Family occupation categories	Less	Some high school	Both missing	Full high school	Bachelor degree	Total
Father missing	3,893	4,337	982	2,768	850	12,830
One not working	1,481	3,919	542	3,130	1,053	10,125
Both working	1,547	9,816	851	8,209	3,508	23,931
Both missing	193	581	6,169	816	657	8,416
Manager	60	662	238	2,011	9,594	12,565
Total	7,174	19,315	8,782	16,934	15,662	67,867

Table B3. Definition of SES categories by family occupation and education

	Family education				
Family occupation	***Lowest 1***	***2***	***3***	***4***	***Highest 5***
Lowest 1	1	1	1	1	2
2	1	1	1	2	3
3	2	2	3	3	4
4	2	2	3	3	4
5 Highest	2	3	4	4	4

Table B4. Ninth-grade students by SES category

Lowest 1	2	3	Highest 4	Total
17,922	16,177	17,760	16,008	67,867
26.4%	23.8%	26.2%	23.6%	100%

Table C1. Probit regressions of ATAR50 (average marginal effects; standard errors in parentheses)

	Pooled	Lowest	SES2	SES3	Highest
Std. Numeracy	0.168	0.141	0.187	0.191	0.156
	(0.002)	(0.004)	(0.005)	(0.005)	(0.005)
Std. Reading	0.114	0.099	0.125	0.128	0.105
	(0.002)	(0.004)	(0.005)	(0.005)	(0.005)
Std. Numeracy, sq	-0.038	-0.014	-0.023	-0.042	-0.041
	(0.001)	(0.003)	(0.004)	(0.003)	(0.002)
Std. Reading, sq	-0.027	-0.008	-0.024	-0.026	-0.029
	(0.001)	(0.003)	(0.004)	(0.003)	(0.003)
Numeracy missing	-0.135	-0.144	-0.161	-0.128	-0.070
	(0.007)	(0.012)	(0.016)	(0.016)	(0.016)
Reading missing	-0.123	-0.106	-0.107	-0.124	-0.129
	(0.007)	(0.012)	(0.016)	(0.016)	(0.017)
Aged 14	0.001	0.014	-0.005	0.006	-0.007
	(0.004)	(0.006)	(0.008)	(0.008)	(0.008)
Aged 16	-0.089	-0.062	-0.105	-0.090	-0.099
	(0.008)	(0.013)	(0.018)	(0.016)	(0.017)
Male	-0.134	-0.114	-0.181	-0.138	-0.109
	(0.003)	(0.006)	(0.007)	(0.007)	(0.007)
LBOTE	0.057	0.095	0.066	0.041	-0.007
	(0.005)	(0.008)	(0.010)	(0.010)	(0.011)
LBOTE & male	0.030	0.022	0.052	0.014	0.058
	(0.007)	(0.011)	(0.014)	(0.014)	(0.015)
ATSI	-0.176	-0.117	-0.236	-0.155	-0.193
	(0.029)	(0.036)	(0.065)	(0.063)	(0.085)
ATSI & male	0.005	-0.045	0.174	-0.095	0.013
	(0.043)	(0.063)	(0.087)	(0.101)	(0.126)
Partial secondary	0.043	0.034	0.068	-0.061	
	(0.006)	(0.006)	(0.011)	(0.021)	
Both missing	0.067	0.047		-0.047	-0.037
	(0.008)	(0.010)		(0.014)	(0.025)
Full secondary	0.116	0.081	-0.011	0.019	-0.055
	(0.006)	(0.008)	(0.050)	(0.017)	(0.009)
Higher education	0.168	0.000	0.027	0.000	0.000
	(0.007)	(.)	(0.051)	(.)	(.)
One not working	0.037	0.026			
	(0.005)	(0.006)			
Employed	0.075	0.000	-0.103	-0.045	-0.013
	(0.005)	(.)	(0.051)	(0.011)	(0.008)
Both missing	0.102	0.000	-0.091	0.000	0.007
	(0.007)	(.)	(0.052)	(.)	(0.016)
Manager	0.101	0.000	0.000	0.000	0.000
	(0.006)	(.)	(.)	(.)	(.)
N	67,867	17,922	16,177	17,760	16,008

Table C2. Probit regressions of ATAR70 (average marginal effects; standard errors in parentheses)

	Pooled	Lowest	SES2	SES3	Highest
Std. Numeracy	0.166	0.099	0.139	0.192	0.223
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Std. Reading	0.107	0.060	0.098	0.125	0.144
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)
Std. Numeracy, sq	-0.025	-0.005	-0.009	-0.025	-0.041
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
Std. Reading, sq	-0.018	-0.005	-0.016	-0.018	-0.027
	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
Numeracy missing	-0.062	-0.076	-0.061	-0.048	-0.026
	(0.007)	(0.010)	(0.014)	(0.015)	(0.018)
Reading missing	-0.047	-0.017	-0.037	-0.061	-0.058
	(0.007)	(0.009)	(0.014)	(0.016)	(0.019)
Aged 14	0.006	0.014	0.001	0.007	0.000
	(0.003)	(0.004)	(0.006)	(0.007)	(0.008)
Aged 16	-0.052	-0.026	-0.039	-0.058	-0.081
	(0.008)	(0.011)	(0.016)	(0.016)	(0.021)
Male	-0.103	-0.066	-0.100	-0.118	-0.137
	(0.003)	(0.005)	(0.006)	(0.007)	(0.008)
LBOTE	0.044	0.050	0.060	0.035	0.006
	(0.004)	(0.005)	(0.008)	(0.009)	(0.011)
LBOTE & male	0.020	0.018	0.017	0.013	0.051
	(0.006)	(0.008)	(0.011)	(0.013)	(0.015)
ATSI	-0.106	-0.054	-0.048	-0.118	-0.233
	(0.030)	(0.031)	(0.048)	(0.068)	(0.104)
ATSI & male	0.006	0.029	0.025	-0.086	0.009
	(0.045)	(0.045)	(0.072)	(0.111)	(0.174)
Partial secondary	0.020	0.016	0.031	-0.080	
	(0.006)	(0.005)	(0.009)	(0.019)	
Both missing	0.058	0.036		-0.021	-0.040
	(0.008)	(0.007)		(0.012)	(0.028)
Full secondary	0.082	0.044	-0.035	0.029	-0.076
	(0.006)	(0.006)	(0.039)	(0.015)	(0.010)
Higher education	0.135	0.000	-0.007	0.000	0.000
	(0.007)	(.)	(0.039)	(.)	(.)
One not working	0.026	0.012			
	(0.005)	(0.004)			
Employed	0.046	0.000	-0.089	-0.074	-0.016
	(0.004)	(.)	(0.039)	(0.010)	(0.008)
Both missing	0.091	0.000	-0.063	0.000	0.026
	(0.006)	(.)	(0.040)	(.)	(0.017)
Manager	0.070	0.000	0.000	0.000	0.000
	(0.005)	(.)	(.)	(.)	(.)
N	67,867	17,922	16,177	17,760	16,008

Table C3. Probit regressions of ATAR90 (average marginal effects; standard errors in parentheses)

	Pooled	Lowest	SES2	SES3	Highest
Std. Numeracy	0.081	0.026	0.048	0.082	0.168
	(0.002)	(0.003)	(0.004)	(0.005)	(0.007)
Std. Reading	0.053	0.018	0.027	0.059	0.110
	(0.002)	(0.002)	(0.003)	(0.004)	(0.007)
Std. Numeracy, sq	-0.005	0.001	-0.004	-0.001	-0.015
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Std. Reading, sq	-0.005	-0.001	-0.002	-0.005	-0.013
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Numeracy missing	0.013	-0.017	0.006	0.007	0.075
	(0.006)	(0.006)	(0.008)	(0.012)	(0.017)
Reading missing	0.006	0.010	0.004	0.013	0.010
	(0.006)	(0.005)	(0.008)	(0.011)	(0.018)
Aged 14	0.000	-0.001	0.001	0.003	0.000
	(0.002)	(0.002)	(0.003)	(0.004)	(0.007)
Aged 16	-0.031	-0.006	-0.006	-0.036	-0.078
	(0.007)	(0.007)	(0.010)	(0.013)	(0.023)
Male	-0.041	-0.015	-0.021	-0.047	-0.084
	(0.002)	(0.003)	(0.003)	(0.004)	(0.006)
LBOTE	0.015	0.013	0.020	0.005	0.015
	(0.003)	(0.003)	(0.004)	(0.006)	(0.009)
LBOTE & male	0.018	0.008	0.010	0.013	0.045
	(0.004)	(0.004)	(0.005)	(0.008)	(0.012)
ATSI	-0.059	-0.003		-0.039	
	(0.030)	(0.016)		(0.051)	
ATSI & male					
Partial secondary	-0.005	0.002	-0.000	-0.060	
	(0.005)	(0.003)	(0.005)	(0.014)	
Both missing	0.024	0.010		-0.008	-0.015
	(0.006)	(0.004)		(0.007)	(0.024)
Full secondary	0.022	0.009	0.032	0.003	-0.079
	(0.005)	(0.003)	(0.034)	(0.009)	(0.010)
Higher education	0.056	0.000	0.034	0.000	0.000
	(0.005)	(.)	(0.034)	(.)	(.)
One not working	0.015	0.003			
	(0.004)	(0.002)			
Employed	0.015	0.000	0.022	-0.043	-0.017
	(0.004)	(.)	(0.034)	(0.006)	(0.007)
Both missing	0.042	0.000	0.029	0.000	0.025
	(0.004)	(.)	(0.035)	(.)	(0.013)
Manager	0.028	0.000	0.000	0.000	0.000
	(0.004)	(.)	(.)	(.)	(.)
N	67,445	17,668	16,036	17,689	15,968