

SAVE OUR SCHOOLS

Education Policy Brief

**Issues in Using Enrolment Data to
Measure the Socio-Economic Status of
Schools**

Trevor Cobbold

October 2010

<http://www.saveourschools.com.au>

Summary

The current system of comparing the literacy and numeracy test results of so-called “like schools” on the *My School* website is systematically biased against government schools because the results of higher socio-economic status (SES) private schools are compared with lower SES government schools.

This bias occurs because the measure of school SES is based on the average of geographical areas in which students live rather than on their family SES. Some high SES families live in low SES areas, but their children are more likely to attend private schools than government schools. They carry their low area SES rating with them and this causes the SES of private schools they attend to be under-estimated. It also causes the SES of government schools to be over-estimated because it is based on the area average which includes high SES families who do not attend government schools. So, the SES of private schools is systematically under-estimated and that of government schools is systematically over-estimated.

My School thus compares the results of dissimilar schools. As higher student achievement is strongly associated with higher SES, comparing the results of higher SES private schools with lower SES government schools makes private schools look better.

In light of this criticism, the Australian Curriculum, Assessment and Reporting Authority (ACARA) is investigating using student level data on parent education and occupation obtained from school enrolment forms to provide a more accurate measure of school SES and avoid classifying dissimilar schools as “like schools”.

In principle, it is preferable to use individual family/student data rather than area-based data to measure school SES. However, there are practical problems in using information from enrolment forms which are likely to also create systematic bias against government schools in “like school” comparisons.

There are high non-response rates on the parent education and occupation questions on enrolment forms. They vary considerably from state to state and by school sector according to NAPLAN reports. They have also been highly volatile from year-to-year. Non-responses increased from 30–37 per cent of all school parents in 2007 to 40–47 per cent in 2008 and then declined to 17–25 per cent in 2009.

These high non-response rates and their volatility raise serious questions about the validity of using this data to measure school SES. It is difficult to understand how non-response rates can increase significantly in one year and then decline by so much in the following year.

The non-responses appear to be highly concentrated amongst lower education and occupation groups. This is demonstrated by comparing the NAPLAN results for students whose parent education and occupation is not stated with those for whom it is stated. For example, the average mean scores in the 2009 NAPLAN tests for students in the non-response group were below those for students whose parents only completed Year 12 or its equivalent. Their mean scores were also similar to students whose parents were in low skilled occupations.

As the missing data appears to be concentrated amongst low SES families, using school enrolment data to measure the SES of schools is likely to significantly over-estimate the SES of government schools because low SES students comprise a much larger proportion of their

enrolments than in private schools. Many government schools will be incorrectly classified as “like schools” with higher SES private schools and their average results then unfavourably compared with these higher SES private schools.

Thus, exactly the same biases as now exist in the comparison of so-called “like” government and private schools will continue.

It is also possible that the biases could even be worse than at present in some common circumstances. For example, school SES could be well over-estimated in government schools with a very large proportion of students from low SES families and many of whom do not respond to the questions on education and occupation. There can be no assurance that using student enrolment data will improve accuracy in the measurement of school SES.

There are also other issues arising from the use of enrolment form data which may bias or distort “like school” comparisons. One issue is that students from low SES families frequently change school but it will not be reflected in changes in the SES of the exited and receiving schools where there are high non-response rates from low SES families. School SES will remain unchanged even if a large proportion of low SES students move from one school to another.

“Like school” comparisons on *My School* will remain fatally flawed and misleading for as long as non-responses to enrolment form questions remain substantial, either generally or for a significant proportion of schools. “Like school” comparisons should be discontinued while the data remains so inadequate.

At the very least, if school enrolment data is used to construct the SES measure for some or all schools for *My School 2010*, the non-response rates should be reported for each school for each Year level. In addition, a caveat should be entered in each school report that the “like school” comparisons may involve comparisons of dissimilar schools because of high non-response rates where they occur.

1. Background

1.1 “Like school” comparisons are systematically biased against government schools

My School purports to compare national literacy and numeracy test results of schools with similar socio-economic status (SES) student populations. However, its comparisons are systematically biased in favour of private schools. It makes private school results look better than their “like” government schools by comparing higher SES private schools with lower SES government schools.¹

The bias occurs because the Index of Community Socio-Educational Advantage (ICSEA) used to measure school SES is based on the average SES of geographical areas in which students live rather than their family SES. Studies by the Australian Bureau of Statistics show that some high income families live in low SES areas and vice versa.

Higher income families are more likely to send their children to private schools than low income families. For example, 55% of higher income families in Australia choose private secondary schools compared to 26% of low income families. However, when high income families in low SES areas attend private schools they carry their low area SES rating with them. This artificially lowers the measured SES of the private schools they attend. The predilection of high income families to attend private schools causes the ICSEA rating of private schools to generally under-estimate their actual SES because these students are classified by their (lower) area SES measure rather than by their (higher) family SES.

The enrolment of students from high income families living in lower SES areas in private schools has the opposite effect on the measurement of the SES of government schools. Their SES is systematically over-estimated because it is based on area which includes high income residents whose children do not attend government schools. Government schools take a greater proportion of low SES students, but many are classified at a (higher) area SES rating rather than by their family SES.

Thus, ICSEA systematically under-estimates the SES of private schools that draw enrolments from high SES families living in lower SES areas and over-estimates the SES of government schools because high SES families resident in their area tend to choose private schools. As a result, *My School* compares the results of dissimilar schools rather than “like schools”. These comparisons favour private schools because, on average, students from higher SES families have higher results than lower SES students.

1.2 Proposal to use student level data

The Ministerial Council for Education, Early Childhood Development and Youth Affairs has requested the Australian Curriculum, Assessment and Reporting Authority (ACARA) to investigate the feasibility and appropriateness of making use of student-level SES data, such as information on parent education and occupation from school enrolment forms, to measure school SES.

ACARA reported in June that for *My School 2010*, two approaches are being considered. One is to obtain the family information in all jurisdictions. The other is to use it for those

¹ This and other flaws in the “like school” comparisons on *My School* are discussed in more detail in Cobbold 2010.

jurisdictions that have it and to continue to use home addresses and Census collection district data for the others.

2. Missing school enrolment form data

A major problem in using information on parent education and occupations obtained from school enrolment forms to construct measures of school SES is the high non-response rates on these questions. Response rates are also highly variable over time, between jurisdictions and between schools. The non-responses also appear to be highly concentrated amongst low SES families.

2.1 Extent of missing data

There are high non-response rates on parent education and occupation data obtained from school enrolment forms. The report on the 2009 NAPLAN results states that the non-response rate on providing parent education information varied between Year levels from 17 to 25 per cent (Table 1). The non-response rate for parent occupation information varied from 20 to 25 per cent. While these non-response rates are high, they are much lower than in the previous years.

The exclusion of such high proportions of families from the data used to construct measures of school SES and groups of “like schools” may lead to inaccurate classification of “like schools” and misleading comparisons of results. Schools classified as “like schools” may be quite dissimilar if particular groups of families are less likely to provide the information than others and these non-responding families are distributed differently between school sectors and schools.

Table 1: Non-Response Rates to Parent Education and Occupation Questions on School Enrolment Forms (%)

Year	Parent Education			Parent Occupation		
	2007	2008	2009	2007	2008	2009
Year 3	37	45	17	40	47	20
Year 5	33	47	25	37	49	25
Year 7	30	40	19	35	42	22
Year 9	na	44	21	na	46	24

Source:

1. 2007: National Report on Schooling in Australia
2. 2008 & 2009: NAPLAN reports

2.2 Variable response rates

The response rates to enrolment form questions have been highly variable in recent years. For example, the rate for the parent education question for Year 3 enrolments in 2007 was 37%; it increased significantly to 45% in 2008 and in 2009 it declined remarkably to 17% (Table 1). Similar variation occurred for other Year levels and for the parent occupation question.

Such volatility in response rates may cause the measured SES of many schools to be highly unstable from year-to-year even though the actual SES of the schools is largely unchanged. This could result in significant changes in “like school” classifications from year-to-year and misleading comparisons of “like school” performance.

It is difficult to understand why these non-response rates vary so much between years. The NAPLAN reports provide no explanation for the increase in the non-response rate of 10 percentage points in 2008 or for the decline of 20-30 percentage points in the following year.

Anecdotal information indicates that many schools do not regularly update information on parent education and occupation for existing enrolments. In many schools, new information is only supplied for new enrolments. Some schools never update their information while others do so only every three or four years.

Response rates are not reported by jurisdiction, but there is apparently significant variation between states and territories. The 2008 NAPLAN report stated that response rates are only reported at the national level because of low response rates in some school sectors in states and territories. The 2009 report stated that response rates are only reported nationally because the rates are variable between states and territories.

“Like school” groups include schools from across Australia and not just those in one state or territory. Large variation in response rates between jurisdictions may invalidate such comparisons of “like schools” and result in comparisons of quite dissimilar schools if the response rates vary between parents with different educational qualifications and occupations.

Furthermore, it appears that response rates are also highly variable between school sectors as witnessed by statement in the 2008 NAPLAN report. This could also invalidate comparisons of school results between government and private schools measured as having similar SES.

2.3 Non-response rates appear to be concentrated in low SES families

There is strong evidence that families who do not provide education and occupation information on school enrolment forms are concentrated in the lowest education and occupational groups.

The 2009 NAPLAN results for students whose parent education is not stated are generally below those for parents who only completed Year 12 and did not have further training or educational qualifications (Table 2). For example, 10 per cent of Year 3 students whose parents’ education was not stated did not achieve the national reading standard compared to 6 per cent of students whose parents completed Year 12. The average mean score of the non-response group was below that of students of parents who only completed Year 12.

Eleven per cent of Year 3 students whose parental occupations were not stated did not achieve the reading standard, which was higher than for students whose parents are machine operators, hospitality staff, office/sales assistants and labourers (8 per cent). The mean scores for the two groups were similar.

Similar results occurred for Year 9. Twelve per cent of students whose parent education was not stated did not achieve the Year 9 reading benchmark compared to 7 per cent of students whose parents completed Year 12. The average mean score for the non-response group was below those of students whose parents completed Year 12.

In addition, 13 per cent of students whose parent occupation was not stated were below the national reading benchmark compared to 11 per cent of students whose parents were machine operators, hospitality staff, office/sales assistants and labourers. The average mean score for students in the non-response group was slightly higher than that for machine operators, etc, but lower than for trades, clerks, office and sales staff.

Table 2: Achievement in Reading by Parent Education and Occupation, Years 3 & 9, 2009

	Year 3		Year 9	
	Mean Score	% Below Standard	Mean Score	% Below Standard
Parent Education				
Bachelor Degree or above	455.5	2.3	619.1	2.0
Year 12 or equivalent	401.0	6.0	576.3	7.2
Year 11 or below	368.1	12.5	547.8	14.8
Not stated	391.9	10.1	568.2	12.3
Parent Occupation				
Senior Manager or Professional	453.3	2.1	616.8	2.0
Trades, clerks, office and sales	404.0	5.0	576.0	6.0
Machine operators, hospitality, labourer	385.8	8.3	558.6	10.7
Not in paid work	374.1	13.5	545.0	18.0
Not stated	386.8	10.8	564.5	13.1

Source: 2009 NAPLAN Report

Note: Some higher education and occupation categories have been excluded.

3. Missing data may cause school SES to be mis-measured

The missing parent education and occupation data is likely to lead to incorrect measurement of school SES by over-estimating the SES of many schools with high proportions of students from low SES families. This could bias comparisons between schools by wrongly classifying some schools with a high proportion of students from low SES families as “like schools” with other schools which have significantly lower proportions of low SES students and lower non-response rates to enrolment form questions.

This will involve misclassification and comparisons between lower and higher SES government schools, lower and higher SES private schools and lower SES government and higher SES private schools.

However, the missing data is likely to cause a systematic bias in favour of private schools in “like school” comparisons because government schools enrol much higher proportions of low SES students than private schools. As the missing data appears to be concentrated amongst low SES families, the SES of government schools with high proportions of low SES students is likely to be significantly over-estimated. The average results for these schools will then be unfavourably compared with the results of higher SES private schools.

Some simple examples illustrate the potential effects.

First, consider a government school where low SES students comprise 50 per cent of enrolments, medium SES students are 40% of enrolments and high SES students are 10 per cent of enrolments. If parents of 40 per cent of students do not state their education or occupation and they are largely, or all, from low educational and occupational backgrounds, their exclusion from the data will cause the school to be assessed as a medium SES school rather than a school with a much lower SES. Only 10 per cent of enrolments will be considered as low SES while 50 per cent will be considered as medium or high SES students.

In any “like school” comparisons, this low SES school will be compared with medium SES schools and because its low SES score is likely to be reflected in low literacy and numeracy results it will compare unfavourably with the performance of its so-called “like schools”.

As a second example, take a private school where low SES students comprise 20 per cent of enrolments, medium SES students 50 per cent and high SES students 30 per cent. If all the low SES parents fail to provide their education and occupation details, the school SES will be over-estimated but not to the same extent as the low SES government schools. It is still likely to be assessed as a medium SES school because low SES students are a minority of the school population.

Thus, while the SES measure of all schools is likely to be artificially pushed upwards by the non-reporting of parent education and occupation information on school enrolment forms, this will be much more significant for government schools where the proportion of students from low SES families is, on average, much higher than in private schools.

Consequently, the current systematic bias against government schools in “like school” comparisons on *My School* will remain even when individual family information from school enrolment forms is substituted for area-based Census data in the measurement of school SES.

Indeed, it is possible that the existing bias could be compounded rather than reduced for some schools. Whether it is greater or less than under the current system of measuring school SES will depend on the extent to which high income families in low SES areas attend private schools and the extent to which non-response rates to enrolment form questions are concentrated amongst low SES families. These factors will obviously vary from school to school.

For example, in using ICSEA to measure school SES it is conceivable that the measured SES of some government schools in low SES areas may not be significantly over-estimated by the leakage of the small number of high SES students in the area to private schools. Using family information from enrolment forms to measure school SES could increase the accuracy of the measure because the high SES students incorporated in the ICSEA measure would be no longer included. However, it is possible that a high non-response rate from the high proportion of low SES families in the district could more than off-set this improvement in accuracy and lead to greater over-estimation of the school SES than under the ICSEA measure.

Thus, there can be no assurance that substituting family education and occupation data obtained from school enrolment forms will lead to more accurate measurement of school SES than under the current arrangements. In some cases, it may but in others it may not. The only way to ensure greater accuracy is to get the current high non-response rates down to a very low proportion of families in each and every school.

4. Other issues

4.1 Missing data and student mobility between schools

The missing data may also mean that actual changes in school SES induced by student mobility between schools are not consistently reflected in the measured SES of the affected schools.

There is evidence of high mobility amongst low SES students, at least in some parts of Australia. A recent Queensland study found that nearly 30 per cent of low SES primary school students change school 2 or more times in five years [Simons et.al. 2007]. Another Queensland study also shows very high student mobility in some schools [Hill et.al. 2009]. Other data shows that nearly 40 per cent of students in the Northern Territory, where average SES is low, change schools in any one year, apart from those who go on to a higher stage of schooling [Dunn 2009].

If the parents of such a large proportion of students changing schools do not report their education and occupation, this data cannot be included in the measured SES of the relevant schools. Thus, changing schools will not affect the measured SES of either the exited or the receiving schools. Yet, their actual SES will have changed. The SES of the exited schools will increase as a high proportion of low SES students have left while the SES of the receiving schools will decline.

Such changes not reflected in the measured SES of schools will cause some schools to be misclassified as “like schools” and lead to misleading comparisons of school performance.

4.2 Failure to update enrolment data may lead to misleading comparisons of schools

Another problem with using school enrolment form data is that it does not appear to be regularly updated for all students in many schools. Often, it is only updated for new enrolments. Consequently, no account is taken of occupational mobility over time. Students spend up to 7 years in primary school and 6 years in high school. Many parents may change occupational groups in those periods and this could lead to significant movement between occupational groups. As a result, actual school SES could change over time and not be reflected in measured school SES, thus leading to comparisons of dissimilar schools which are wrongly classified as “like schools”.

4.3 Scope to manipulate the measurement of school SES

A further problem in using enrolment form data is that it adds to the scope for schools to manipulate comparisons of school results on *My School*. Already, there are extensive opportunities available to schools to manipulate their results under the pressure of public comparisons of school results and the need to protect or improve school rankings [see Cobbold 2010b].

Enrolment forms and the information they contain are under the control of schools. Where this data is used in ways which may affect the reputation and standing of schools and the careers of teachers and principals, there will be a temptation to manipulate the data to show the school in the best light. Schools will be able to tamper with the data to manipulate the measurement of their SES so they are compared with lower SES schools.

5. Conclusions

Use of information on parent education and occupations obtained from school enrolment forms will lead to inaccurate measures of school SES. There are high non-response rates to these questions on school enrolment forms and they appear to be concentrated amongst low SES families. This invalidates the use of this data to measure school SES for the purpose of comparing “like schools”.

The basic flaw in the current system of measuring school SES from area-based Census data will remain. Government school SES scores will be systematically over-estimated with the

result that government schools will be wrongly, and unfavourably, compared with higher SES private schools in “like school” comparisons.

There can be no assurance that resorting to individual family data from school enrolment forms to measure school SES will result in more accurate measures. It is possible in some circumstances that the measured SES of many government schools will be even more over-estimated than at present.

This suggests that “like school” comparisons on My School should be discontinued until such time as the current high non-response rates on the parent education and occupation questions on enrolment forms can be reduced to a very low proportion of parents in all schools.

At the very least, if school enrolment data is used to construct the SES measure for some or all schools for *My School 2010*, the non-response rates for the data used should also be reported for each school for each Year level. In addition, a caveat should be entered in each school report that the “like school” comparisons may involve comparisons of dissimilar schools because of high non-response rates where they occur.

There are also other issues arising from the use of enrolment form data which may bias or distort “like school” comparisons. One issue is that students from low SES families frequently change school and this will not be reflected in changes in the SES of the exited and receiving schools where there are high non-response rates from low SES families.

Using enrolment form data also creates further scope for schools to manipulate school comparisons. It creates the potential for schools to tamper with the data to manipulate the measurement of their SES so they are compared with lower SES schools.

References

Cobbold, Trevor 2010a. Like School Comparisons Do Not Measure Up. Research Paper, Save Our Schools, Canberra, February. Available at: <http://www.saveourschools.com.au>

Cobbold, Trevor 2010b. Rorting and Cheating of School Results is the Future Under My School. Education Policy Brief, Save Our Schools, Canberra, May. Available at: <http://www.saveourschools.com.au>

Dunn, Bruce 2009. Student Movement 2007 and 2008: NT Government Schools. School for Social and Policy Research, Charles Darwin University, July. Available at: http://www.cdu.edu.au/sspr/documents/DET_Student_Movement_2007_and_2008.pdf

Hill, Angela; Navin, Fiona and Lynch, Andrea 2009. Coming to Grips with Student Mobility and Policy Implications: A Case Study from Regional Queensland. Paper presented to the Australian Association for Education Research Annual Conference, 29 November- 3 December, Canberra.

Ministerial Council on Education, Employment, Training and Youth Affairs 2007. National Benchmark Results Reading, Writing and Numeracy Years 3, 5 and 7. National Report on Schooling in Australia, Preliminary Paper. Available at: http://www.mceecdya.edu.au/verve/_resources/ANR2007Bmrks-Layout_FINAL.pdf

National Assessment Program for Literacy and Numeracy 2008. Achievement in Reading, Writing, Language Conventions and Numeracy. Available at: http://www.naplan.edu.au/reports/national_report.html

National Assessment Program for Literacy and Numeracy 2009. Achievement in Reading, Writing, Language Conventions and Numeracy. Available at: http://www.naplan.edu.au/reports/national_report.html

Simons, Roland; Bampton, Margo; Findlay, Andrea and Dempster, Alistair 2007. Student Mobility, Attendance, and Student Achievement: The Power of Implementing a Unique Student Identifier. Paper for the Annual Conference of the Australian Association for Research in Education. Available at: <http://www.aare.edu.au/07pap/sim07568.pdf>