

Grattan Institute Report
Measuring What Matters: Student Progress by Dr Ben Jensen
January 26th 2010



Author: Chris J Lloyd
Professor of Business Statistics, Melbourne Business School
Date: 09 February, 2010

On January 26, 2010 the [Grattan Institute](#) released a report ([HERE](#)) on measuring school performance. The main recommendation of the report is to replace measurement of average school performance with so-called *value-added* indices. The idea is very simple – to measure student *progress* as the primary outcome – and by employing an appropriate statistical model to extract that component of the improvement which can be attributed to the school.

The report was particularly well timed. The federal government launched the [My school website](#) on January 28 which publishes average student performance by school, reported within groups of “similar schools”. Moreover, results for the 2010 National Assessment Program – Literacy and Numeracy ([Naplan](#)) are to be published in May 2010 and will, for the first time in Australia, mean that value-added measures can be calculated.

There is much to like about the report and it is difficult to argue against its main conclusion – that measuring student outcomes at one time point and averaging over each school does not provide a valid measure of school performance. Obviously this depends on what one means by a school’s performance. If one means the ability of a school to attract and retain smart students while deterring less smart students then the average student outcome is probably the only meaningful measure. But if by school performance you have in mind the effect of the school on the student’s learning outcome, then school averages will be hopelessly biased. The reason is that students are not randomly allocated to schools. Rather, gifted students tend to concentrate in some schools while disadvantaged students concentrate in others.

MySchool tries to correct for this by measuring the level of disadvantage of the school. They use a single measure of disadvantage (ICSEA) largely based on the student’s ABS census district (not the post code as claimed in the Grattan report). This is then averaged over the school. So ICSEA really measures the disadvantage of the geographical catchment region of the school – not of the particular students who happen to attend the school. According to the [government fact sheet](#):

The Index of Community Socio-Educational Advantage (ICSEA) is a special measure that enables meaningful and fair comparisons to be made across schools.

This “fact” is wrong in several ways. First, using a student’s community status as a proxy for the student’s status causes bias because the actual socio-economic differences between schools are diluted by the crudeness – a bit like regression to the mean. Second, even if the advantage of each student was correctly measured (rather than by census district), to use this data correctly it must be linked directly to the student’s score, not averaged over the school. Thirdly, the most pertinent and easily measured index of the student’s aptitude is surely their result on the previous Naplan test. This is now available and to not use it is both scientifically invalid and wasteful of the expensive data already collected.

There are some interesting features of the Australian educational landscape that are pertinent. Based on an international standard test (PISA) the variation in scientific literacy outcomes in Australia is larger than the OECD average (by 11%). But this higher inequality is not explained by the popular notions of disadvantaged schools failing to compete with elite schools. In fact, 81% of the variation in outcome is *within* schools. This suggests that measuring individual student outcomes over time will be useful in targeting problem students as well as assessing intervention programs.

Naplan tests are administered at years 3, 5, 7 and 9 every second year. With publication of the 2010 results in May, there will no longer be a practical excuse to not consider more meaningful measures, assuming that Naplan can match the students in their 2008 and 2010 exercises. To not use the student level identifiers (the main value of which is to track student level changes) is statistically indefensible. No trained statistician would publish the school level averages when student identifiers were available.

An appropriate statistical model would try to deconstruct each student’s outcomes into different components. For instance one might say that the 2010 year 9 result depends primarily on (1) that student’s 2008 year 7 result, (b) the student’s ability to learn, (c) the student’s non-school circumstances and (4) the school environment. It is the last of these components that we think of as measuring school performance. The role of the statistical model is to measure this adjusting for the first three components which are out of the school’s control. The non-school circumstances might include things like parents’ education, occupation, marital and employment status, family size, parity, gender, ethnicity, migration status and language preference. In the Australian context, MySchool already measures socio-economic disadvantage of the student based on their postcode.

There are some misleading statements in the report but nothing that undermines their main contention. For instance, at various points it is claimed that naïve school averages have “consistently been shown to produce biased estimates of school performance compared to value added modeling.” As pointed out above, it is really a matter of how one defines school performance and the matter is not resolved by a mathematical or empirical study. Nevertheless, most reasonable people would agree with the author’s view that average school performance is not a good measure.

At another point, the report notes that overseas experience shows that the volatility of value-added measures is greater for smaller schools. They then argue that small school results should not hold implications for those schools. This is surely incorrect since each school’s value-added measure will come with a measure of statistical significance which takes into account the fact that measures for smaller schools are less reliable. Even for the smallest school, if results were sufficiently bad then some action would be indicated.

The report is (probably deliberately) vague about exactly how the value-added measures are to be calculated. The measure will depend on a statistical model which will require some expert statistical modeling. They do mention however that academic research has shown that the final measures are quite insensitive to the exact details of the model employed. Nevertheless, it would have been nice to have an indication of the kind of statistical model that one might employ, even if relegated to an appendix. For those of a more mathematical background, one might start with a fixed effects regression model such as

$$y_{it} = \beta x_{it} + \alpha_t y_{i,t-1} + \theta_j 1_{\text{stud } i \text{ in school } j} + \varepsilon_{it}$$

where y is the Naplan result, “ i ” is the index of the individual student, t is the time, j is the school and x are the individual level covariates of non-school environment. A better alternative would be a so-called “random effects” model where the historical achievement term $\alpha_t y_{i,t-1}$ is replaced by a more complicated but flexible term.

This report is timely and persuasive. While those opposed to school performance measurement will find reasons to question the validity even of value-added measures, the reality is that performance data will continue to be generally available and we would all be better served if the current school averages were replaced by something imperfect but better.