SUBMISSION TO NAPLAN SENATE ENQUIRY

Inadequacies of NAPLAN results for measuring school performance

Margaret Wu
Associate Professor
University of Melbourne

INADEQUACIES OF NAPLAN RESULTS FOR MEASURING SCHOOL PERFORMANCE

INTRODUCTION

The government's key justification for the MySchool website is that there is a need to be transparent about school performance. The assumption is that student performance, as measured by the NAPLAN tests, can be used to infer school performance, after student background factors are accounted for. Currently, a socio-economic index (ICSEA) is used to adjust for student background factors. Consequently, when schools with similar ICSEA index are compared, differences between schools are attributed to school effect. Further, for future NAPLAN results, there is a plan to use student growth measures as a proxy for school effect measures, since growth measures for individual students already control for the variations in student scores, so student background factors will not need to be controlled for.

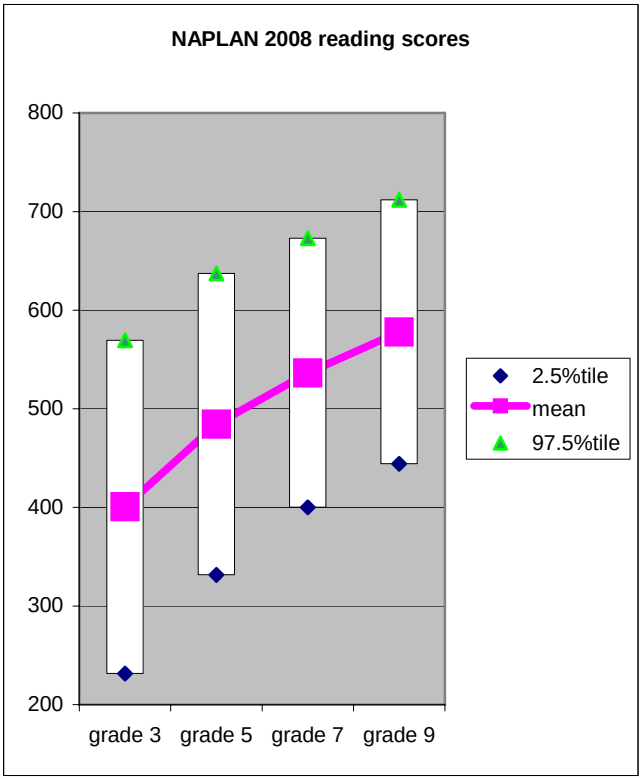
This submission demonstrates that neither of the two methods (i.e., using ICSEA or using growth measures to control for student background factors) provides sufficiently accurate measures of school effect. In fact, the publication of NAPLAN results on the MySchool website should be deemed as providing false information to the public, as the red and green bars do not in any way show school performance as claimed by the government.

MARGINS OF ERROR IN MEASURING *STUDENT* PERFORMANCE

The main problem with measuring school effect through measures of student test results is that test scores have large margins of error. For example, if David was a Grade 5 student and he obtained 25 out of 40 on the 2008 reading test, we would expect David's test scores to range between 20 and 30 should similar tests of 40 questions be given (e.g., the 2009 and 2010 tests are given). This variability in a student's test scores when similar tests are given is called Measurement Error. On the NAPLAN scale, David's numeracy score is 488 ± 54. That is, should similar reading tests be administered, we would expect David's NAPLAN reading score to vary between 434 and 542. This is a very wide range. It reflects the uncertainty with which David's reading ability is measured by the NAPLAN test. To put the magnitude of this range in perspective, Figure 1 shows the 2008 average reading scores of Years 3, 5, 7 and 9, and the spread of student scores.

Figure 1 shows student scores distributions for Grades 3, 5, 7 and 9 for the NAPLAN 2008 reading assessment. The white rectangular box for each grade shows the range *mean*±2×*standard deviation* (i.e., 95% of the student scores are expected within this range).

Figure 1 NAPLAN reading scores distributions in 2008



We see that the possible range of David's scores (434 to 542) is so wide that it covers about half of the range of Grade 5 student scores, and that the range (434 to 542) covers Year 4 to the Year 7 average reading scores.

When David takes the Year 7 reading test in 2010, a similarly inaccurate estimate of David's reading score is obtained. A growth measure for David over the two-year period based on the two NAPLAN tests David has taken has a margin of error of ±76 NAPLAN score points. Compare this magnitude of the error with the actual expected growth of around 52 NAPLAN score points, it is clear that a growth measure based on two time points does not provide much information on how much a child has really grown.

MARGINS OF ERROR IN MEASURING *SCHOOL/TEACHER* PERFORMANCE

Research studies on measuring teacher effect have found that a high performing teacher can raise student standards by one more year of growth as compared to a low performing teacher. That is, if you try your best in teaching Grade 3 students, you may raise your students to Grade 5 level in one year, at the most. This is reasonable, as one would not expect that a teacher can make Grade 3 students grow to Grade 6 and beyond in one year. This means that if teacher effect is measured on the NAPLAN scale, students taught by the highest performing teacher will be 50 score points higher than if the same students were taught by a lowest performing teacher, after one year of learning. So if individual teacher

effects are to be measured, we need a precision to within a few NAPLAN score points, given that the range of teacher effects across all teachers is about 50 points. However, research studies have also shown that, based on student growth measures on two testing occasions, the estimates of individual teacher effects have a margin of error of around ±20 points. That is, the uncertainty associated with the estimate of a teacher's performance is about as large as the difference between the best and the worst teacher. Growth measures obtained from NAPLAN tests simply cannot provide information about individual teacher performance.

SUMMARY

With the current NAPLAN design where there is only one annual test of 40 questions per subject area, student scores contain large margins of error. NAPLAN results do not provide sufficiently accurate information on student performance, student progress, or school performance. It is educationally unsound to publish the results and to call on parents to judge schools based on these results. We owe the public the real transparency of the NAPLAN results.

NOTE

I would be happy to provide further technical justifications for this submission.