

More Questions Raised About the New York School Reporting Model

Trevor Cobbold

More questions are being raised about the validity and reliability of the New York system of reporting school performance, which Julia Gillard wants to introduce in Australian schools.

Last week, report cards for New York City elementary and middle schools were released for 2007-08. They show huge variations from the previous year and significant inconsistencies with other test results.

First, there was a dramatic increase in the number of schools receiving an A or B grade. The number receiving an A grade jumped from 226 to 394, an increase of 74%, 71% of schools that got a C or D in 2006-07, received an A or B in 2007-08. Overall, 38% of schools received an A compared to 23% in 2006-07.

Second, several schools that were failed last year ('F' grade) received a top grade this year. Of the 34 schools that got an F in 2006-07, 80% scored an A or B in 2007-08. Nine schools leapt from Fs last year to As this year, 10 jumped from Fs to Bs, and another 17 rose from Ds to As.

Third, there was a large decline in the number of schools receiving a low grade. Just 50 out of 1,043 schools received a D for the 2007-08 year - down from 86 schools in 2006-07. The number of F-rated schools dropped by nearly half, from 35 to 18.

Jennifer Jennings and Tom Pallas from Columbia University have called these changes "magical transformations" which are more the "fabled stuff of Hollywood movies" because real school change takes time and doesn't happen overnight.

[<http://blogs.edweek.org/edweek/eduwonkette/> 22 September]

As was the case in 2006-07, the report cards reveal major inconsistencies between city, state and federal assessments of the same schools. For example, two schools that received an A rating were added to the state's list of failing schools in February. In more than 60 of the 394 A-rated schools, more than half the students failed to reach proficiency on the state's reading test.

Again there were major discrepancies between federal and city assessments. Thirty per cent of the schools deemed failures under the *No Child Left Behind Act* earned A's from the city, while 16 of the city's 18 failures are in good standing under the federal guidelines.

The *New York Times* reported that such inconsistencies had left "many who know the school system well somewhat befuddled" [16 September].

The wild fluctuations in year-to-year results seem to point to significant statistical flaws in the New York reporting model. The principal of one school that increased its grade from a D last year to an A this year said that they had not done anything differently.

"A school doesn't move from a D to an A in one year unless there is a flaw in the measurement or the standardized test itself" she said. "We have not done anything differently, certainly not in response to the progress report." [*New York Times*, 16 September]

According to one of the leading US experts in educational testing, Daniel Koretz, Professor of Education at Harvard University, the tremendous variation in schools' grades from last year to this year probably has less to do with school improvement than sampling and measurement error. [<http://blogs.edweek.org/edweek/eduwonkette/> 17 September]

Koretz says that the methodology behind the New York City reporting system is "baroque". He points out two major problems of using the report cards as a measure of school performance. The tests are not vertically linked from one Year level to the next and the measurement errors are not reported.

Tests for following grades are not measured on a continuing scale so it is not possible to compare a student's performance on this year's test to performance in the previous grade. Therefore, it is not possible to determine whether, for example, a student who gets the same score in grades 4 and 5 has improved, lost ground, or just not progressed. Neither is it possible to compare the progress made by different students.

This and other problems with the tests mean that equally effective schools, one serving higher achieving students than another, may not receive similar Progress Report grades. This invalidates comparisons between schools.

A further problem is that measurement error should be taken into account in assessing year-to-year changes. Koretz notes that experts working on value-added reporting models have warned for years that the results from a single year are highly error-prone, particularly for small groups. He says this seems to be exactly what the New York City results show in that there is far more instability from one year to the next than could credibly reflect true changes in performance.

"It strains credulity to believe that if these schools were really 'failing' last year, three-fourths of them improved so markedly in a mere 12 months that they deserve grades of A or B...This instability is sampling error and measurement error at work. It does not make sense for parents to choose schools, or for policymakers to praise or berate schools, for a rating that is so strongly influenced by error."

Analysis of the new data by Jennifer Jennings and Tom Pallas shows that there is almost no relationship between the progress scores in 2007 and 2008 [<http://blogs.edweek.org/edweek/eduwonkette/> 22 September]. They find that the changes in progress scores are largely due to measurement error.

"The progress measure, it appears, is a fruitless exercise in measuring error rather than the value that schools themselves add to students.... It's impossible to know what your A or your F means, because these grades are dominated by random error."

Another possible factor is test score inflation. Koretz has warned that when educators are under intense pressure to raise scores, high scores and big increases in scores become suspect. Scores can become seriously inflated—that is, they can increase substantially more than actual student learning.

One source of test score inflation is cheating and it seems that this could already be a problem with the New York system. For example, the *New York Sun* [30 June 2008] has reported that a South Bronx elementary school that adopted the motto "The Best School in the Universe" on the strength of soaring tests scores is being investigated for allegations that teachers helped students cheat on state tests. Several students said that teachers would examine their answers during official test administration periods and point out mistakes and how to correct them.

The *Sun* reports that the Department of Education is also investigating cheating allegations at another New York school. A group of teachers had written to the Department saying that school's principal had asked several teachers to help students during the state tests. One teacher provided the following statement:

"He basically said during the exam that I should go over close to them, and for example if they mark 'D' and 'D' is not the right answer, tell them, you know, 'That's not the right answer, try something else,' and just keep guiding them until they get the right answer."

A spokesman for the city teachers' union, said that the union has observed other cases similar to those in the South Bronx, with teachers saying their principals are pressuring them to cheat.

Gillard says that she is "inspired" and "impressed" by the New York model. Yet, it is subject to scathing criticism by test experts, teachers and parents alike in New York. Like many other models of reporting school results, it is plagued by statistical and other problems which make it an unreliable and inaccurate measure of school performance. It should be decisively rejected as a way of measuring school performance in Australia.

Trevor Cobbold

Convenor, Save Our Schools

Save Our Schools is a Canberra-based public education advocacy group.

<http://www.soscanberra.com/>

22 September 2008