

# **Why Ranking Schools Would do More Harm than Good**

Paper Presented to the Australian  
Education Union Federal Executive  
Meeting, Melbourne,  
12-13 September 2007

Bob Harris, Education International  
and TUAC (OECD)

# Summary

1. Aspiring to quality - and equity – in education
2. Evaluation, accountability and indicators
3. Testing and ranking
  - Where the pressure comes from
4. Examples elsewhere
  - (UK, USA, Poland, Norway)
5. Lessons to be drawn
6. Quality, equity and evaluation - getting it right
7. Political positioning

## 1. Aspiring to quality - and equity – in education

- The quality of education is high up in political debate. It is a political hot button:
  - internationally, eg. at the G8 since 2000
  - nationally, a major election issue in every OECD country
  - sub-nationally, an issue in the States of the US, the Provinces of Canada, the Lander of Germany, the Cantons of Switzerland...
- Equity is also profiled often – but not always;

## 2. Evaluation, accountability and indicators

### Two premises

- Every teacher – and parent – knows that evaluation is part of the process of education.
- Accountability is normal and necessary in public services

Issues arise when evaluation of systems, schools and/or teachers is combined conceptually with accountability. The political debate over performance testing arises out of this conceptual combination, not to say conceptual confusion.

The debate in many countries is often based on loose assumptions. An example is the assumption that census testing will provide more valuable information about system performance than indicators. There is no evidence to support this assumption.

## 3. Testing and Ranking

-

Across the OECD politicians are under pressure to come up with answers to the universal aspiration to improve quality in education. Testing and ranking is all too often their response.

That response is stimulated by the widespread notion that efficiency in public services can be enhanced by applying the practices of business enterprises. Among those practices is the widespread use of what is known in the jargon as “metrics”, ie measuring outputs in relation to inputs, “cost-effectiveness” if you like.

Measurement is in fashion. But what does measurement mean in education?

We should also mention **PISA** - the Program of International Student Assessment. PISA is a programme for comparing national performance, using as an indicator the results of tests from a random sample of 15 year-olds in each country on literacy, numeracy, scientific awareness, and capacity to solve problems. ACER is a key player in developing and implementing PISA. 67 countries participated in 2006, with the results to be released this December. PISA results attract front page attention in many countries. Beyond the headlines about whether Japan has fallen behind Korea, etc., we think PISA has helped to provide some useful data showing that good education policies, as in Finland, can achieve both quality and equity. PISA data has also been useful in addressing issues of migrant education in Denmark and Germany, for example. It must be underlined that PISA is based on sampling and on indicators – not on census testing.

#### **4. Examples of testing and ranking elsewhere**

##### **England**

Others have gone down the road of census testing and ranking. The best known of course is England. The National Curriculum in England is divided up into four key stages. At the end of Key Stages 1-3 (for 7, 11 and 14 year-olds), teachers are expected to conduct end of key stage National Curriculum assessment. A level of professional judgment is allowed to teachers for 7 year-olds. This flexibility does not apply to 11 year-old testing (end of Key Stage 2) and 14 year-old testing (end of Key Stage 3). The 11 year-old and 14 year-old tests take place in May, on fixed days and are externally marked. Schools are evaluated on the percentage of children achieving the appropriate level 4 in the National Curriculum tests of English, Maths and Science. Then, at age 16, schools are evaluated in performance tables on the percentage of young people who receive 5 A-Cs in GCSEs, including English and Maths. The same public exam benchmarks do not exist for age 18.

Note this: neither Wales, nor Scotland, nor Northern Ireland have end of Key Stage testing. So there are no school performance tables, a.k.a. League Tables in Scotland, Wales or Northern Ireland.

In England, changes are in any case foreseen for public examinations for both 16 and 18 year olds. A new national Diploma has been introduced which covers both vocational and academic skills. This places a big question mark over the future of school performance tables for 14-19 year olds. The Diploma is being trialed from September 2008.

### When were school performance tables invented in England?

The Conservative Government introduced school performance tables in 1994 after it had introduced in 1987 the National Curriculum and its assessment and testing arrangements. Subsequently, local authority performance tables were introduced.

While National Curriculum tests and school performance tables were Conservative Party inventions, the idea of national targets based on school test results is very much a new Labour invention. National targets based on test results were introduced in 1997. The National Curriculum targets are used to evaluate the performance of local authorities. They are also used as benchmarks to evaluate the performance of individual schools. A new invention from New Labour in the last two years has been to introduce the idea of contextual value added when it comes to school performance tables. The contextual factors of race, ethnicity and deprivation are factored into the results, but they are ignored by the press who use the raw test results for the school performance tables.

While League tables have had much media attention, it is worth recalling that England has a separate national school evaluation system: a national inspection agency called OFSTED. The role of OFSTED is to evaluate the quality of teaching and learning in each school. Often OFSTED's findings on the performance of each school vary markedly from the published test results for each school. OFSTED can be critical of a high performing school which it considers to be coasting (not doing its best by its highly able children) while that school receives a high ranking in the school performance tables. Conversely, OFSTED can praise a school with high numbers of children from socially deprived backgrounds and yet the school appears very low in the school performance tables. These are contradictions which the Government has not resolved.

The most succinct rationale for the use of school league tables comes from Sir Michael Barber who was, until recently, the Head of the Prime Minister's Delivery Unit for Public Service Reform. In a recent book, he gives the following rationale for league tables.

*"Not everyone in the public service likes league tables, but I love them. I have spent much of the last decade advocating them, usually in front of skeptical or even hostile audiences of head teachers. They make the evidence about performance public, they focus minds on the priorities they encompass, and they make sure, in whatever system they are applied, that something is done about the individual units at the bottom of the league table – whether they are schools, hospitals, police forces or local authorities. In fact there is no more powerful drive for equity."*

The purpose for school league tables in England is to exert direct control over individual schools and local authorities which the government believes are underperforming. School performance tables, targets and tests results are very much the weapons of a Government intent on using its own powers to “lever up standards within individual schools”.

The key problem with school performance tables and national targets, as seen by NUT:

- teachers are constrained to teach to the tests which contribute to the league tables of each school’s performance;
- the rest of the curriculum suffers because of teaching to the test;
- schools are unfairly identified as failing even though they may have had good OFSTED inspection results;
- national targets create enormous strains in the system and even Government finds that it has shot itself in the foot since it has not been able to achieve the ambitious national targets it set itself 10 years ago (leading to the resignation of a Secretary of State, Estelle Morris when the targets were not achieved in 2002);
- school performance tables means that enormous resources are pumped into supporting children who may achieve a benchmark level 4 or 5 at the expense of those who need extra support but who can’t make level 4 or 5. This particular issue is relevant for 11 year olds at Key Stage 2 where enormous amounts of resourcing are focused on moving youngsters over the level 4 borderline;
- school performance tables confuse public forms of accountability for schools and lead to conflicting messages between evaluations of the quality of teaching and learning by OFSTED and the raw data yielded by narrow tests.

In short, the impact of National Curriculum tests, performance tables and national targets has narrowed considerably the National Curriculum, undermined the achievements of the Government’s own National Literacy and Numeracy Strategies, and penalised particularly low achieving youngsters where support has been withdrawn from them in favour of youngsters on the borderlines of National Curriculum levels in Mathematics and English.

**United States**

In the United States we have also seen moves toward nation-wide performance testing, but with an approach significantly different to that of England. As in Australia, issues of national targets and performance testing inter-act with issues of Federal/State responsibility and resourcing. In 2001, the Bush administration and Congress set a federal requirement that all students be tested in math and reading-annually in grades 3 through 8, and at least once in grades 9 through 12. NEA pointed

out at the time that the new requirement risked disrupting the more-comprehensive student and school assessment programs that had been developed by states over the previous 15 years.

Let's look at what the US Federal Law says:

- States must develop and administer tests in math and reading to students in grades 3-8 and once to all students in grades 9-12, beginning with the 2005-06 school year;
- States must also administer science assessments to all students once in grades 3-5, 6-9 and 9-12, beginning with the 2007-08 school year.

The law goes on to say that states must:

- *design or purchase* tests that are aligned with state content and performance standards. If standards span more than one grade, teachers must be informed as to what portion of those multi-grade standards are to be taught at each grade.
- *design or purchase* tests that are the same for all children (with appropriate accommodations as needed), but are valid and accessible for all students, including students with limited English proficiency and students with disabilities.
- *design or purchase* tests that are consistent with nationally recognized professional and technical standards and use multiple measures that include higher-order thinking skills. Tests must also objectively measure academic achievement, knowledge, and skills without evaluating or assessing family beliefs and attitudes.

Finally, the law requires that:

- states must produce and provide individual reports of student performance to parents, teachers, and principals in a comprehensible and uniform format.

Many teachers and other educators are concerned about heavy reliance on standardized testing. Pencil and paper tests may not go far enough in measuring the skills of students—skills that will certainly contribute to their ability to compete and succeed in tomorrow's workplace. Federal requirements for annual testing in math and reading have already affected state assessment programs that measure student progress in other subjects such as science, history, and geography.

NEA, AFT and their affiliates are working to:

- analyze the impact of standardized testing requirements on learning.
- urge Congress and the Education Department to shape the law to ensure that tests are not used as the sole measure of a school's success.
- remind policy makers and members of the public who are judging their students and schools to rely on a fuller picture that includes other "outputs," such as graduation rates, safety record, numbers, and percentages of students taking challenging courses.

Now, here is a crucial point about the US experience. As in the case of England, we have seen confusion arising out of the imposition of national laws on testing and targets while retaining pre-existing and established evaluation and assessment systems.

The US has long had NAEP – National Assessment of Educational Performance. In the 1970's under Gough Whitlam, there was a move to introduce NAEP in Australia. I represented ATF on the committee set up at the time. The committee recommended against NAEP for Australia, and it was not introduced.

Returning to the United States today, The NAEP definition of Proficient is not synonymous with "proficiency" in a subject. Yet, under the No Child Left Behind Act (NCLB), states are required to report the percentages of students achieving proficiency in reading and mathematics for grades 3 through 8. For each subject and grade combination, the percentages vary widely across states. For grades 4 and 8, these percentages can be compared to the estimated percentages of students achieving proficiency with respect to the standard established by the National Assessment of Educational Progress (NAEP). Again, large discrepancies are observed. This variation could derive from differences in both content standards and student academic achievement from state to state, as well as from differences in the stringency of the standards adopted by the states. There is no way to directly compare state proficiency standards because states are free to select the tests they employ and to establish their own performance standards.

There is a major debate among specialists in the United States around these inconsistencies.

## **Poland**

Poland is the most populous country in Europe. Currently it has an ultra-conservative government, although elections have just been called following a crisis within the governing coalition. Nation-wide testing was introduced in 2002, with census testing of all pupils at given ages. Individual school results are published. EI's affiliate ZNP reports that the main stated aim of creating a national testing system was to enhance the quality of education, as well as to gain better information about student learning outcomes. However, straight after the first test results were published, they turned into leagues tables, and were reported in the media as as such. ZNP points out that applications of students for secondary schools are influenced, and special classes are created for the 'worse' students. Research is underway on value-added assessment.

## **Norway**

The Norwegian government decided to bring in national tests in 2004. In 2005 national tests were compulsory for all students in years 4, 7, 10 and the first year in upper secondary. The results were to be published, but were not supposed to be presented as league tables. The tests created debate and resistance among teachers, parents and pupils, and many students in year 10 and upper secondary boycotted them.

Some of the reasons for the opposition against the tests may have been the way it was introduced to the schools. The schools got the information late and it was unorganised. The tests meant extra workload for teachers and school-leaders which was not compensated.

The purpose of the tests was not clear. The results were supposed to meet many interests; to help the teachers to adapt their teaching to the needs of the students, but also to meet the needs of government to monitor the student performance in the country.

The debate after the first implementation of national tests was very much about publishing of school results. Although the government said they didn't want league tables, there was no problem for the newspaper to compile and publish them as such.

UEN did not oppose the national tests, but was skeptical of the system for publishing the results, calling for the results to be part of school-based evaluation-work.

The reactions against the national tests lead to an evaluation of the tests. Many experts from universities and elsewhere advised against using the tests. In 2006 the tests were stopped by the new Labour-led government. This autumn the tests, revised, will be carried out again, but only in a few subjects and only in year 5 and 8. The purpose of the new tests is to measure if the students can achieve the competence objectives for basic skills in numeric and literacy in the new curriculum. As under their predecessors, the new government says that they will not provide league tables, but we'll have to wait and see how the media will use the information about the results which will be published in a national web portal.

## **5. Lessons to be drawn**

There are several common features of experiments with national performance testing in other countries:

- In general, performance testing with publishing of school results has been introduced in stages; there are strong grounds for believing that a stage by stage process is underway in Australia.
- There has been confusion between established systems for evaluation and assessment and new regimes of performance testing;
- The process for arriving at performance testing and ranking of schools into league tables has been driven by political agendas, with an eye more to media impact than to the educational interests of children and young people.
- Performance testing of all students at given age levels is highly disruptive, distorts the teaching of a broad curriculum, and adversely influences resource allocation and equitable school admission policies.
- There is little or no empirical evidence that performance testing of all pupils actually does anything to improve educational quality, while there are strong indications that such experiments have an adverse impact in terms of equity.

## **6. Quality, equity and evaluation - getting it right**

Performance testing comes out of the 'back-to-basics' movement. Yet education, more than ever, must be broadly based. One of the criticisms of PISA is that it is very limited – does not cover geography or history, for example. The same criticism applies to nation-wide performance tests. Simplistic reporting of testing and ranking is an unsatisfactory substitute for serious community-wide debate

about education, including what parents expect of their schools. Achieving the goals of quality and equity in education for all Australian children and young people will take hard work, widespread consultation, partnership between educators and their communities, agreements on effective instruments for monitoring and accountability, and resources. Short cuts will be self-defeating and counter-productive.