

**A RESPONSE TO THE APPA POSITION PAPER
ON
THE PUBLICATION OF NATIONALLY COMPARABLE SCHOOL
PERFORMANCE DATA**

Ken Boston
August 2009

PP = Presentation slide

I thank the Australian Primary Principals' Association for the invitation to respond to their position paper on the publication of national comparable school performance data, and to do so from the perspective of my work in England.

PP

A great deal of progress has been made in recent years in Australia towards goals in which I have always strongly believed.

You have a National Curriculum Board developing a genuinely national curriculum provision, for the first time in the history of this country.

You have implemented a program of national testing, designed primarily as a diagnostic tool to provide information about individual children, classrooms and schools for teachers, principals and parents.

You have a new set of National Goals for education, following on the earlier Hobart and Adelaide Declarations.

Ministers have agreed on a framework for the publication of nationally consistent information about school achievements within the context in which the schools function, with the objective of supporting continued growth in the quality of education.

And the new Australian Curriculum, Assessment and Reporting Authority is being established. It will - amongst other things - publish relevant, nationally comparable information on schools.

Much has been achieved, but even more remains to be done, and in that emerging work there is both opportunity and risk.

Much seems to have been agreed at a level of broad principle, but understandably not at a level of specifics and detail.

PP

The APPA paper supports the principle of transparency, on the understanding that appropriate safeguards and protocols are put in place to ensure that the release of information about students and schools has a beneficial impact on primary education and that the potential negative effects have been nullified.

It also accepts, with some important qualifications and caveats, that NAPLAN results should be one of the sources of information for comparing the performance of schools, provided their use for that purpose leads to an enhancement of the quality of schooling.

Statements made by Commonwealth and state ministers seem to be in agreement with this.

There also seems to be agreement in principle that the provision of public information - the delivery of transparency - should not be by means of league tables which simply rank schools according to test scores. But until the detail of how transparency will be achieved without the use of league tables, there is understandable uncertainty about that point.

Comparisons between schools must be fair and accurate.

There are various qualifications and provisos: on assessment preparation; on a review in 2010; on technical reports on validity and reliability being made available; on an independent review of results; and on ongoing consultation on draft proposals

Above all, the bottom line for APPA is that the transparency agenda must have a positive impact on the primary curriculum.

PP

It seems to me that the most helpful contribution I can make today is, in the light of experience in England:

to suggest some of the caveats, protocols and safeguards needed to ensure that the national transparency agenda has a positive impact on education

to look at the issue of making fair and accurate comparisons between like schools

and to consider better alternatives than league tables for the provision of transparent public information.

PP

I should begin by nailing my colors to the mast. I am:

a supporter of national testing in England, and in Australia;

opposed to the test results in England being used for purposes for which they are not fit;

concerned by the debilitating impact that the high stakes uses of the test results in England have on the school curriculum;

concerned about the archaic method of delivery of the tests

and totally opposed to league tables.

Full-cohort annual national testing has a critical role to play in providing real-time diagnostic data to inform principals and head teachers and the system as a whole about what works and what doesn't work, and to allow ministers make strategic and resource decisions based on hard and contemporary evidence.

NAPLAN has been designed primarily for the purpose of providing diagnostic information, not population statistics for policy makers. While the English national testing program began with a similarly noble purpose, it is now used for a host of other purposes - which I will come to later - and is very largely summative rather than diagnostic in purpose.

Indeed, now that the key stage 3 tests at age 14 have gone, the only remaining tests in England are key stage 2 at the end of primary schooling. NAPLAN, at two year intervals at ages 3, 5, 7 and 9 can realistically be of diagnostic benefit. But the key stage 2 tests are of no diagnostic use for the primary teachers, and the results are passed on to the various secondary schools which usually disregard them and then test the children again before putting them into 'sets' or ability groups.

In England, the government's use of the key stage tests has seriously damaged the breadth and quality of primary education. The tests have changed from an essentially diagnostic test for the purpose of school and system improvement, to a high stakes summative test on which depend - amongst other things - the pay and future employment of the head teacher and staff.

PP

As a result the school curriculum is narrower and poorer than it was when the tests were introduced in 1997. In many schools, the time spent on areas of the curriculum which are not externally assessed has contracted sharply.

Most schools prepare pupils extensively before they undertake the tests. A survey conducted by QCA in 2007 showed that 68 per cent of primary schools employed additional staff to prepare students for the key stage 2 tests, 78 per cent set additional homework, more than 80 per cent had revision classes and used practice tests they had purchased commercially.

The amount of time spent on test preparation has increased over the past 10 years: in the second half of the spring term 70 per cent of schools spent more than three hours per week on test preparation. In some extreme cases, months have been spent in the final year of primary schooling on nothing else than test preparation, to the neglect of the other areas of the curriculum and hence to the great detriment of the quality of the children's education.

Teachers have been concerned about the impact of this on the quality of education for a generation of children. But interestingly, it is employers' concerns that have made it a prominent issue.

Employers find that, despite their formal qualifications, so many 16-18 year olds are inarticulate, and unable to communicate simply and well; they cannot work collaboratively and constructively in teams; they lack initiative and enterprise; and surprisingly, given that most of their lives have been spent in school, they lack a thirst for continued learning and personal growth. They are deficient in the soft skills which form an essential component of the human capital of each individual, regardless of their

academic achievements.

The response of government has been the adoption of essentially remedial strategies: the introduction of programs for secondary school students in areas variously described as personal learning and thinking skills, life skills or employability skills. But should England really need remedial strategies in the soft skills for 16 year olds, after they have spent two-thirds of their lives in continuous education?

It is now being understood increasingly by the public - as teachers always knew - that employability skills are the product of a full, balanced and well rounded primary and junior secondary education. Teamwork can be taught and learned only over a decade of activities such as playing sport weekly, or singing in a choir. Initiative, enterprise, self-management and a thirst for learning are created by building from the early primary years, and throughout the following decade, respect for individual creativity and real achievement, through recognition and reward. Competitiveness, and striving for individual success, is at the heart of teaching and learning in initiative, planning, personal organization, problem solving and enterprise. None of these educational outcomes is achieved or enhanced by league tables: they are the result of enlightened school leadership and effective classroom teaching across the full range of the national curriculum.

There is not much wrong with the primary school curriculum in England. The real problem is that teachers and schools aren't able to get on with the business of teaching it. The teaching program focuses on what is to be tested, and on practicing for the tests, because the future of the school is dependent upon the result.

PP

This is an extract from the current performance table of primary schools within a part of

Buckinghamshire. extracted from the web.

The government in England is careful to say that it does not publish league tables - what the government publishes are performance tables, also called attainment tables. This is disingenuous: the performance tables give results alphabetically, the newspaper league tables rank schools according to score. Performance tables become league tables at the click of a mouse.

The performance tables give five sets of data, of which this slide is the key set, which everyone looks at.

The others are background; contextual value added (which I will come to when making another point later); year on year comparisons; and absentee rates.

The background for each school is no more than the total number of students - nothing about school aims and objectives, curriculum, school organization, student intake characteristics, physical and human resources, areas of special focus. There is virtually no contextual information to enable you to understand possible reasons for differences between the performance of schools.

In England there are eight assessment levels for the curriculum, from level 1 to level 8, for ages 5 to 16. The key target at age 11 at the end of primary school is for students to achieve level 4 or above in English, maths and science.

The columns show, from left to right, the name of the school, the number of students in the final year of schooling, and then the English, maths and science results at Level 4 and above, and level 5.

So Little Houghton Church of England Primary School is high achieving: every child is above level 4 in the three subjects.

Great Linford Primary is significantly underachieving, and is well below the local authority average and the England average.

But the tables give us no information to help us understand why that might be the case, or what is being done about it.

The real problem with league tables is that they are far from transparent - they are opaque. They tell you only the enrolment, the results, the yearly trend in results, and persistent absentee rates. In the absence of contextual information the league table becomes a proxy for all other information, which is inferred.

This school has good results: it must be well-led, the teachers must be great, it must be well-funded, discipline must be good, the pupils will be the sort of kids I'd like my child to have as friends. But this other school has bad results; its leadership and teachers must be poor; it is obviously run-down; my child would not be happy there.

It is inevitable of course that the media and others will construct league tables from publicly available data. They can be stopped from doing so only by the provision of fuller and better alternatives. The partial and limited nature of the league tables in England enhances their authority and plays to the fears of many parents who might well make better decisions about meeting the needs of their own child in the light of complete and genuinely transparent information.

League tables cannot be banned, but their significance can be substantially reduced; they should be given absolutely no status by ministers or education systems. It is disappointing in England that no minister has ever come out with the public strident criticism the league tables deserve, and that the government has allowed the landscape of schools and schooling to be shaped so crudely.

PP

Well, what are some of the of the caveats, protocols and safeguards needed to ensure that the national transparency agenda has a positive impact on education? I suggest three.

PP

The first is to ensure that the purpose or purposes for which the NAPLAN results are used are specific, agreed and clear, and that the results are used for those purposes alone.

National testing was originally introduced in England as a diagnostic tool, to identify how many children were achieving, on the eight point scale, level 2 at key stage 1 (age 7), level 4 at key stage 2 (age 11), and level 5 at key stage 3 (age 14). The purpose was essentially diagnostic, to inform planning and resource allocation at school, local authority and national level.

Many other purposes have since been bolted on, and the tests have become a largely summative national assessment, for a variety of purposes for which the tests were not designed and are simply not fit, including being a significant determinant of residential property prices throughout England.

PP

The results of the national tests are used for many purposes, including the following:

to determine whether national performance in English, maths and science has improved since last year, or deteriorated

to judge whether a school is a good school or a bad school

to judge the social or personal value of an individual child's achievements

to determine whether an individual student is making sufficient progress in relation to attainment targets

to identify learning needs and guide further teaching

to diagnose learning difficulties

to determine whether a child meets eligibility criteria for special education provision

to place children in sets or ability groups

to identify the general educational needs of students who transfer to new schools

to select students for entry to a school, or to distinguish between them

to identify the most desirable school for a child to attend

to decide whether institutional performance – relating to individual teachers, classes or schools – is rising or falling in relation to expectations or targets; and, potentially, to allocate rewards or sanctions

to identify institutional needs and allocate resources

to identify institutional failure and hence the need for intervention

to evaluate the success of educational programs or initiatives, nationally or locally

to guide decisions on the comparability of examination standards for later assessments on

the basis of cohort performance in earlier ones

to 'quality adjust' education output indicators for the purposes of national accounting.

Now, it can be argued many of these uses are diagnostic in part, and that all of them are laudable, but the key stage tests were not designed for all these purposes and they are certainly not fit for many of them.

For example, taking the first of these purposes, although test results might indicate a true trend over three to five years, they really don't give an accurate measure of annual change in educational performance.

After the introduction of key stage tests in 1997, performance in English, maths and science improved steadily for five or six years, but as in many other western countries the rate of improvement has since declined, to almost a plateau. At key stage 2 last year, 78 per cent achieved level 4 in maths and 81 per cent in English. The figures rise or fall by about a percentage point or less in each year. The latest figures are just out: maths and science are unchanged, English has fallen by one per cent.

Great attention is paid by Government and the media to those annual changes. An increase of one per cent is hailed by government as evidence of its policies working, and condemned by the media as evidence of dumbing down. A fall of one per cent is seen as cause for concern by government, or evidence that the QCA has failed to maintain the assessment standard and has increased the level of demand, or by the media as a real decline in performance.

PP

There are 650,000 students in the final year of primary school. Each year each cohort sits an entirely new paper. The levels are set by the examiners, taking into account the

perceived level of demand of that year's paper. At the critical threshold level, level 4, there are typically up to 4 per cent of the cohort (26,000 students) at the mark just below the threshold, and a similar number on the mark just above the threshold. Where the boundary is drawn is finally a quite subjective judgement by experienced senior examiners, all of them very experienced teachers, taking into account the perceived difficulty of the test in relation to previous years. It is not an exact science. The minor annual fluctuations tell us nothing, and yet they are the subject of national headlines and earnest interviews on radio and television, and lengthy meetings of boards of governors in schools.

Similarly, although the key stage tests results might give some partial and imperfect information, they are not an adequate instrument for diagnosing learning difficulties, or deciding on school placement, or determining teachers' pay.

The lesson for Australia is then to be absolutely clear about the purposes of testing, and don't attach additional purposes to NAPLAN for which it was not designed and for which it is not fit.

PP

Secondly, report school outcomes directly to the public, rather than as vale-added or 'contextualized' attainment measures.

School performance in England is measured and reported by adjusting the scaled results of students, that is the published results, to reflect the school's student intake characteristics and other elements of school context.

PP

This is extracted from the league tables for the secondary schools in Buckinghamshire LEA. The columns from the left show:

Name of the school.

The enrolment.

The percentage getting five GCSEs with a score from A* to C (a good pass), including maths and English

The percentage with five A* to C GCSEs in all their subjects (commonly take eight; some 12-14)

The percentage with five A* to C GCSEs in a modern foreign language

The value added score. This is explained on the website in the following terms: “This is arrived at by predicting what pupils should achieve when they arrive at school at age 11. If - on average - pupils improve on their predicted performance and do better than their fellow pupils in similar circumstances, the school will be awarded a score of more than 1000. If they do worse, it will get less than 1000.”

The average point score per pupil. An A grade is worth 270 points, a B grade 240, a C 210, a D 180 and an E 150, and the maximum score of 45 in the International Baccalaureate is 1380 points.

So, by this account Aylesbury High School, a selective state-financed grammar school had a win and added value.

The Buckingham School, which is a secondary modern which exist to take youngsters not admitted to grammar schools, by this particular piece of science, subtracted value. Aylesbury’s achievement on the one hand; Buckingham’s failure on the other.

PP

Now, there are real problems with this:

it fails the transparency test

it has a false air of precision, which is simply not justified

it suggests that a certain margin above or below expectation has the same meaning, despite the fact that the schools might be greatly different

it makes low achievement seem acceptable in schools that suffer from social disadvantage

it doesn’t tell parents whether, given the characteristics of their child, any particular

school can be expected to produce a better result than another school.

PP

Now the MCEETYA communique of 17 April 2009 said that “Ministers agreed that these reforms were not about simplistic league tables which rank schools according to raw test scores.”

We should accept the reassurance by ministers that this reform agenda is not about league tables, and all such tables are of course simplistic; it is the word ‘raw’ which makes me pause. I’m sure the ministers mean scaled standardised scores, as published, not truly raw scores, but even so there is an issue.

Undoubtedly value-added and contextualized modeling has an important role in analysing system performance and in planning at a system level. But based on experience in England, I think we need to be very cautious about using any value-added or contextualized value added data in publicly reporting school performance, rather than simply reporting the results the students actually obtained.

Value added data cannot be readily explained publicly. Parents don’t understand it, and it raises more questions than it answers. So the lesson is don’t massage the data; simply report the results, and the change in results over time, without adjustment.

PP

The third suggestion is to set targets, but use them cautiously, and never as an instrument for implementing policy.

In England, as in the United States, the attainment of minimum performance targets is now the basic dynamic setting the education agenda. It passes the transparency test, and it is easily understood. But no target is an appropriate target for every student. In England, it is clear that the current targets provide no incentive for schools to extend those students who are already well above the performance standard, and the league tables provide no incentive for schools to meet the needs of those who are far below the target and will never achieve it.

With regard to the last part of the caveat, I give an example.

Until about five years ago, students in England were required to take a modern foreign to age 16, at GCSE. The Government then abolished that requirement. The numbers taking

languages subsequently plummeted. To turn that around, it was decided that, from 2008, performance in a modern foreign language would be counted in the GCSE league tables, as in the example I have just shown you. It was left to the schools to determine how to do this and what resources to deploy, but failure to adjust would mean they would drop their ranking in the league tables.

Setting an output target without the prior input of resources, and providing sufficient lead time, is not a satisfactory way to achieve curriculum reform. For instance, I would like to see the physical fitness of children in England - and indeed in Australia - having sharper focus in the primary curriculum than it does at present, being monitored and reported nationally for schools and school systems, and aspirational but achievable attainment targets being set. But this would require a prior increased investment of people, class time, facilities and other resources. It can't be done simply by setting a target and telling the schools to get on with it.

So that is the third lesson.

PP

I turn now to the issue of making fair and accurate comparisons between like schools.

The APPA paper argues against the public reporting reporting of schools results in a 'like schools format.'

It also makes the point that like schools or schools in a common statistical neighbourhood will be similar according to some measures not others.

Quite understandably, APPA doesn't want there to be a league table of statistical neighbourhoods, in each of which there is a league table of schools.

My own thoughts on this are that there are good reasons why a like school format should not be used for public reporting on individual schools.

It is true, too, that any individual school can be grouped into a number of like groups. Some schools in a group defined as small rural schools will also be Aboriginal schools, but others will not; some suburban primary schools will also be schools with children from many language backgrounds, or large primary schools, or primary schools on newly developed housing estates, or in gentrified inner suburbs.

'Like schools' can be cut many ways, with the uniqueness of each school being reflected

in the fact that it can be included in a number of like groups.

It seems to me that the best way to learn about what is working in schools, and hence support school improvement, is to analyse the performance of groups of schools which share one or more common characteristics, while recognizing that not all schools in a particular group will have common membership of other groups.

If, when looking at a particular group of schools, we find that there are significant differences between them in the extent and rate of learning by children, we need to ask why, in order to identify what is working and what is not. It might reflect differences in curriculum, school organization, teaching methods, assessment practices, professional development, or how the school is being organized and led. Or it might be that a particular school stands out from the other schools in the group, because of the much stronger effect which the characteristics of another group to which it also belongs have upon it.

The purpose of that sort of analysis is, as I see it, not to blame or shame, but to identify good practice, to see what works and what doesn't, and to apply that learning to the improvement of the system as a whole. A rising tide raises all ships.

PP

Where does that take us?

I think we're saying that:

We're committed to a fair deal for every child.

We want every school to be a good school.

We support transparency in the publication of nationally comparable performance data.

We don't want league tables which rank schools in order of performance.

We see the point of grouping schools with some particular characteristics in common for

the purpose of identifying and reporting best practice.

But we don't want individual school performance reported on the basis of comparison with schools in any one particular group, because not all schools in any group have all characteristics in common.

PP

So, that brings me to the final question: what are the better alternatives than league tables for the provision of transparent public information?

PP

School Report Cards, as in New York City and soon to be introduced in England, are in tabular form, they give a grade or ranking for the school, and give some limited contextual information about the nature of the school. Currently they are a form of league table, although the contextual information presumably could be expanded.

PP

School profiles, now common in Australia, and readily accessible on the web. Generally, they are prepared within a template framework which requires certain performance data to be reported, but gives scope for an extended narrative prepared by the school on its aims, objectives, philosophy, curriculum emphases, staffing, facilities and so on. Provided they do include some agreed mandatory performance indicators, these are much more informative than the school report cards.

But they have two weaknesses. One is that there is great variation in the quality and range of the information provided by the schools, which could be overcome by agreeing a common format.

More importantly, from my limited scanning of some of the profiles on the web, there appears to be still no external verification of the information provided by the school. If school profiles are to be seen as more than a prospectus, and are to be accepted as valid and reliable information on which parents make decisions, their credibility in the eyes of the public depends upon them being audited or signed off as a report by a third party, the obvious one being the Australian Curriculum and Assessment Reporting Authority.

PP

External inspection: in Australia I had doubts about inspection, but after seven years working alongside Ofsted, I believe it is the most powerful and constructive force in school improvement in England today.

Its approach is from an educational child-centred perspective, not a political media-centred perspective, as with the league tables.

It begins each inspection with the school's own self-evaluation, which is along the lines of a school profile, although prepared in response to a tightly prescribed format.

It inspects on the basis of risk: many schools are inspected only at five year intervals, others on the basis of need.

It writes a report on each school, setting out the results the children achieved, the extent to which the school's objectives are being met, the strengths and any weaknesses of the school, and what needs to be done to address the weaknesses, where they exist. Where weaknesses do exist, Ofsted specifies a timeline - which is limited but adequate - and sets a date for a further inspection.

The reports are both summative and diagnostic; clearly written; highly credible; above

politics; they respect the uniqueness of each school; they are focussed entirely on the needs of children.

I have found strong support for Ofsted from most of the teaching profession, and did not expect to. There are sometimes grumbles about short notice being given of an inspection, and of the workload involved in preparing for it, but teachers generally seem to find great benefit from an intelligent, informed and independent appraisal of the school. Many head teachers bemoan the fact that they will have to wait several years for an Ofsted visit, when there are so many good things going on in the school that they would like Ofsted to authenticate publicly.

There are 22,500 maintained schools funded by government; 17,500 of them are primary schools. Currently 515 schools are receiving special support, through special measures, and they are all schools which have been chronically and seriously failing for some years, and in which the kids have been getting a very poor deal.

Since 1997, 1400 schools have been in special measures. Only two hundred of them have been closed, and 57 of these have been reopened as a new school on the same site. The remaining 1250 schools are now all very effective schools, to the great benefit of some tens of thousands of children who would otherwise never have received a decent education. It works.

Now, although Ofsted currently is a far cry from the old style inspections of HMI (Her Majesty's Inspectors) there is a long tradition of external inspection in Britain, which can't readily be transferred to Australia. I'm not proposing that an Ofsted inspection model should be set up in this country, because it is quite foreign to our education culture, and would fail to be accepted. But if you want to look to the United Kingdom for examples of best practice in school and system improvement, you will find the best of them in the approaches which Ofsted currently employs.

An obvious question is why league tables continue to have such an influence on the public perception of school performance in England, when the public has access to such thorough and independent Ofsted reports on each individual school. The provision of alternative better, full and authoritative information has not diminished the importance of league tables, as I earlier argued it should.

The answer is also obvious. Responsibility lies fairly and squarely with the current government: ministers have facilitated the easy construction of league tables; they have used the language of crude ranking in their own rhetoric; occasionally they have publicly regretted the fact that league tables exist, but then presented themselves as unable to do much about it; they have never come out with the sort of strategy needed to attack and destroy the league table culture, but in my view have encouraged it.

Ministers are ministering to all youngsters and to all schools. Education is a public good, not a positional good. Education nationally is not about ministers managing a competitive market of winners and losers, but about a fair go for every youngster in every school, public or private, across the country.

PP

So, in conclusion, what is the best and non-toxic way to achieve genuine transparency, and to ensure that it has a really positive impact on the quality of primary education for all children?

My suggestion is:

To support the analysis and publication of reports on the performance of like schools, being groups of schools which share one or more common characteristics for the purposes of that analysis. For example, achievement in mathematics in:

small rural primary schools, inner suburban primary schools, inner suburban primary schools in Melbourne, independent primary schools receiving government funding above a particular level, and primary schools in Tasmania.

To support publication of national data comparing the achievement in one group of schools with the achievement in other groups of schools, but without identifying individual schools, and hence using the reports as the means for identifying the need for improvement and support of the group as a whole (as distinct from the particular needs of any individual school)

To support the further development of the school profile as the means of providing information to parents and the community on the achievements of the school, on its aims and objectives, on its curriculum and programs, on its staffing and facilities, and its financial resources, subject to:

- the adoption of a national template for preparation of the profile; and
- the establishment of a process for third-party verification of the content of the profile.

The latter is absolutely critical. If it is a choice between believing what it says on the scoreboard (that is the league tables), or believing the material produced by the school (which has very good reason to be entirely self-serving), the public will go with the league tables.

My advice is that the price of avoiding league tables, and the price of avoiding being ranked within league tables of statistical neighbourhoods, is greater external scrutiny at the level of the individual school.

And in the interests of the young people who are currently in your care, and the future of Australian education, that price is well worth paying.