Large cohort testing – How can we use assessment data to effect school and system improvement?

Dave Wasson[1]

Department of Education and Training, New South Wales

Dave Wasson began his teaching career in 1975, working as a classroom teacher in a number of secondary schools before becoming a regional consultant. He later held executive positions in schools, including the role of deputy principal at Whalan High School in the Western Sydney Region. In 1994 he was seconded to the Quality Assurance Directorate. He subsequently held various Chief Education Officer (CEO) positions related to quality assurance, corporate performance, school effectiveness indicators, and school reporting, before becoming the Principal of Erina High School. He was then appointed School Improvement Officer, Parramatta District, and later became CEO, School Reviews.

Since 2004 he has held the position of Director of the Educational Measurement and School Accountability Directorate (EMSAD), with responsibility for supporting school improvement initiatives across NSW and ensuring school accountability in the areas of: school reviews; annual school reporting; implementation of statewide and, more recently, national testing programs; and supervising student selection and placement in selective high schools and opportunity classes.

___

1 The opinions expressed in this paper are those of the author and not necessarily those of the NSW Department of Education and Training. I wish to acknowledge the significant contribution to this paper made by many former and current members of various iterations of the Educational Measurement and School Accountability Directorate over many years.

Abstract

This paper discusses the introduction and use of data from large cohort testing into teaching and learning in New South Wales schools. It highlights the conditions that existed towards the end of the 1990s when a number of influences and initiatives coalesced to enable large cohort testing to impact positively on student outcomes. It then considers how some of these lessons might be employed to enhance the impact of the new era of national testing heralded by the introduction in 2008 of the National Assessment Program – Literacy and Numeracy (NAPLAN).

In order to contain the scope of the discussion, this paper begins with an examination of the NSW experience with the use of the data from the Basic Skills Test in literacy and numeracy in Years 3 and 5 from 1996 to 2007.

To assess the impact on teaching and learning this paper also looks at a range of school effectiveness indicators used in NSW to drive school and system improvement, including the notions of measuring growth, and value added and relative effectiveness. In addition, it traces the development of the Like School Group structure employed in NSW to more meaningfully compare the performance of schools. It also evaluates the utility of various tools in supporting the analysis and interpretation of these indicators at both a school and system level.

Finally, the paper highlights the merits of a transition from current pencil-and-paper testing to an online environment to enable the assessment of a greater range of syllabus outcomes and to provide more timely feedback to teachers, students and parents.

Introduction

Large cohort testing in literacy and numeracy in Australia is a relatively new activity. The jurisdiction with the longest history is New South Wales which began full cohort testing of students in Year 6 with the introduction of the Basic Skills Test (BST) in 1989. Its introduction was vehemently opposed by the NSW Teachers Federation and by a number of members of the Primary Principals' Association (PPA).[2]

More recently the outcomes of large cohort testing and associated resources in NSW have largely been welcomed by teachers and principals across both primary and secondary schools. But there are still a number of pivotal questions: How did this culture of acceptance of the outcomes of large cohort testing develop? And, can large cohort testing improve school and system performance? If so, how?

Cizek (2005) argues that high stakes (accountability) tests are incapable of providing high-quality information for instructional purposes and doubts if relative group performances have anything meaningful to contribute at the school level. The NSW experience supports the contrary view: that testing and assessment programs can effectively serve two purposes at once, if the design of the tests is appropriate and there are mechanisms in place to convey the critical diagnostic and performance-related messages to the right people in a flexible and timely manner.

The NSW Department of Education and Training has addressed these issues by:

- Providing a relevant curriculum framework in the form of a high-

___

2 Chadwick, V., *NSW Legislative Council Hansard*, 28 April 1992.

quality syllabus upon which the tests are based

- Ensuring that the statewide testing programs reflected what teachers were teaching

- Providing to teachers sophisticated, relevant and accessible diagnostic information relating to the performance of their students

- Ensuring that teachers can access relevant resources and support to address areas of identified need.

The sophisticated analysis of student performance and the capacity to access high-quality resources electronically are features that teachers and principals can access through the highly valued and supported *School Measurement, Assessment and Reporting Toolkit (SMART)* software.

This paper will provide a historical overview of the development of large cohort testing in NSW, highlighting some critical developments. It will then discuss current developments, including the support provided to schools for the current National Assessment Program – Literacy and Numeracy (NAPLAN) tests. Finally, the paper will pose some future challenges in relation to large cohort testing to ensure its utility and effectiveness in promoting school and system improvement.

Historical overview of large cohort testing in NSW

The Greiner Liberal Coalition Government introduced a Basic Skills Test for all Year 6 students in NSW in 1989, providing outcome information in literacy and numeracy. In 1990 the decision was taken to expand the test to include Year 3 students. At this stage the tests were not developed on a common scale and the notion of measuring growth between testing points was not considered.

In 1994 the decision was made to move the test from Year 6 (at the end of primary schooling in NSW) to Year 5. This was an acknowledgement of the concerns from primary principals that the information from Year 6 testing came too late for teachers to meaningfully address any identified issues from the data. As a subsequent Minister for Education observed: 'The previous Government changed the Basic Skills Test from Year 6 to Year 5 after finally realising what nonsense it was to hold basic skills tests in Year 6 when it was not possible to diagnose the results'.[3][4]

In 1996 and until the end of the BST in 2007, the Year 3 and 5 tests were developed on a common scale for literacy, and a separate common scale for numeracy. The reason for this was to provide an accurate and reliable comparison of the performance of students across the two year levels. The reports could now reflect an individual student's development from Year 3 to Year 5. The reporting language was still the same but now it also had the same meaning in Year 3 and Year 5.

The method by which this was done was to link the tests by having common questions in both. Extensive trialling identified suitable questions to act as link items.

The BST was originally developed by the Australian Council for Educational Research (ACER) using the Rasch measurement scale. Analysis by ACER showed that the scale underpinning the BST satisfied the requirements of the Rasch model (local independence, unidimensionality, specific objectivity).

Each year extensive trialling of items took place and only items that fitted the Rasch model were considered for the final test. A combination of common person equating and, since 1996, common item equating was used to place new tests on the historical literacy and numeracy scales. In the equating process, items in the equating test that showed significant misfit were not used.

As a result of these processes, stable and reliable estimates of student and cohort achievement on common literacy and numeracy scales were obtained. It is thus valid to compare individual student scores over time and also examine cohort trends to see whether improvements have occurred.[5]

The use of a common scale for both Years 3 and 5 allowed for the first time the depiction of growth between testing points. In a large and diverse jurisdiction such as NSW, this was a critical development in ensuring greater acceptance of the utility and accuracy of the data provided to principals from the administration of large cohort testing. They had argued, rightly, that comparisons based on the raw performance of student cohorts in schools was flawed and indefensible as schools serve communities with diverse demographics.

An internal review of the BST undertaken in 1995 (Mamouney, 1995) made a number of recommendations, including:

- Provide BST results on computer disk with appropriate software to enable schools to analyse the data on site for school-specific purposes

- Improve analysis of the BST data to look for patterns of performance which could inform the use of data for the benefit of individual students, schools and system

3 Aquilina, J., *Legislative Assembly Hansard,* 9 April 1997.

4 Lind, P., Interview by Dave Wasson, 2 July 2009. Peter Lind is a Senior Data Analyst with the Educational Measurement and School Accountability Directorate, NSW Department of Education and Training.

5 Lind, P., Interview by Dave Wasson, 2 July 2009.

- Provide better ways of supporting school use of BST data through training programs.

In 1996 there was pressure from the NSW Primary Principals' Association to provide the information from the BST electronically and, in 1997, the first iteration of what was to become known as the *School Measurement, Assessment and Reporting Toolkit (SMART)* was released.

In 1996 it was also apparent that the percentage of students in the lowest band in the BST for Year 3 (Band 1), and the lowest two bands in Year 5 (Bands 1 and 2), was unacceptably high. There was a need for a new approach to the teaching of both literacy and numeracy in NSW schools. In 1997 the State Literacy Strategy was launched. This was accompanied by a new syllabus (K–6 English Syllabus, 1998), an unprecedented level of professional learning for teachers and a large bank of practical teaching resources, as well as enhanced central and regional consultancy support.

According to the Director at that time of the Curriculum K–12 Directorate, the State Literacy Strategy:

> … drew fragmented philosophical strands together and focused on *explicit and systematic* teaching and buried the prevalence of learning by osmosis. The Strategy provided a secure foundation for literacy learning and revolutionised the way teachers and educators in NSW talked about learning. It provided

the confidence that NSW was moving in the right direction regarding literacy teaching and largely neutralised the debate between the whole language and phonics camps.[6]

The new K–6 English Syllabus was released in 1998 and the State Literacy Strategy evolved into the State Literacy Plan in 1999. The Plan provided for an increased concentration of resources in terms of personnel, support materials and professional learning for teachers. It was accompanied by the comprehensive assessment of student literacy skills via the Basic Skills Tests and the provision of sophisticated electronic analysis of individual, group and school performance via SMART.

The BST for primary schools was subsequently complemented by a new literacy assessment for secondary students in 1998, the English Language and Literacy Assessment (ELLA), followed by the Secondary Numeracy Assessment Program (SNAP) in 2001.

An extensive evaluation of the State Literacy Plan was undertaken in 2003 by the Educational Measurement and School Accountability Directorate (EMSAD). The evaluation (NSW Department of Education and Training, 2004) confirmed that the Plan was highly successful and that teaching practice had indeed changed. The evaluation also indicated the resources

___

6 Wasson, L.J., Interview by Dave Wasson, 19 September 2007.

developed were focused and valued and that teachers were now better equipped to identify areas of student need.

The following table illustrates the trends from 1996 to 2007 for students placed in the bottom and top bands in BST literacy. While the outcomes from a large cohort testing program such as the BST are subject to volatility from year to year, there is a noticeable improvement trend, with a reduction in the percentage of students in Band 1 from about 17 per cent in 1996 to about 11 per cent.

It is important to note that the underlying scale for the development of the BST in NSW did not change over this period. This indicates a level of genuine improvement of student outcomes from 1998, when the percentage of students in Year 3, Band 1 for example was reduced from 15.4 per cent in 1998 to 10.7 per cent in 1999.

So, in 1998 there was a convergence of initiatives thatconspired to positively impact on the learning outcomes of students in NSW: student outcomes data from large cohort testing; the implementation of a high-quality syllabus and a statewide training and development program; and the provision of sophisticated diagnostic information on student performance for teacher use via SMART.

Table 1: Literacy percentages in Bands

Literacy Percentages in Bands												
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Y3 Band 1	17.0	16.0	15.4	10.7	11.1	11.8	10.7	12.2	10.8	11.5	10.6	11.1
Y3 Band 5	16.7	17.3	13.2	13.9	15.1	19.8	18.1	17.7	16.6	20.4	19.4	19.5
Y5 Band 1&2	9.0	8.2	8.5	5.9	7.5	6.2	5.4	6.0	6.9	7.1	6.9	6.8
Y5 Band 6	19.2	23.7	20.2	19.6	19.5	23.0	24.9	25.6	27.8	23.8	25.0	26.7

Table 2: Numeracy percentages in Bands

Literacy Percentages in Bands												
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Y3 Band 1	10.8	10.8	13.8	10.3	14.7	10.6	9.3	8.1	10.1	9.2	9.1	8.5
Y3 Band 5	23.6	17.7	21.0	16.0	15.3	15.4	17.9	17.1	15.1	21.8	21.8	19.3
Y5 Band 1&2	7.5	5.9	5.7	7.4	8.0	6.4	6.3	6.0	6.4	6.6	5.4	6.5
Y5 Band 6	20.9	20.9	23.3	23.2	19.6	23.2	25.1	23.1	24.9	23.9	29.6	32.6

It is also apparent that this percentage has stabilised and that further reduction, including reduction of students at this level in NAPLAN, will require a new approach.

Over the same period for numeracy the improvement is not as pronounced, perhaps reflecting a greater emphasis on literacy in NSW at both policy and operational levels.

An evaluation of assessment and reporting processes and outcomes in NSW was undertaken in 2003 by Eltis and Crump. At this time, Eltis and Crump detected a major shift in attitude to the outcomes of large cohort testing. They observed that there was 'overt support for testing programs' and that there was a marked increase in the 'quality of information available to schools as a result of statewide testing programs.'

They further noted that 'statewide tests have come to be valued by teachers and parents for their perceived diagnostic assistance for each student … (Eltis & Crump, 2003). In addition, Eltis and Crump commented on the quality and support for an earlier iteration of the SMART software which '… allows schools to analyse their results by viewing achievement levels, student results and questions, and question details. Results and graphs can be printed (and the software) provides hyperlinks to resource materials.'

Key developments

The developments described above were pivotal in gaining support for the outcomes of large cohort testing in NSW. In addition to these, over the last decade a number of initiatives relating to the provision of more sophisticated school performance information have been implemented that have provided additional levels of analysis to teachers, principals and their supervisors. While some of this additional information was welcomed in schools, the data presented school performance in new and challenging ways that meant even some high-performing schools in purely raw terms were not performing as expected when their school intake characteristics were taken into account.

Growth

A most important type of additional information presented was the measurement of growth. The depiction of growth between testing points, where the underlying measurement scale was common, was possible with the implementation of a common scale across Years 3 to 5 from 1996. The notion of growth between testing points levelled the playing field, when the two variables that have the greatest impact on the quality of student outcomes in NSW are taken into account: socioeconomic status and geographic location. This initiative was relatively quickly understood by principals and largely embraced.

It was depicted in SMART in a way that allowed the growth of individual students to be identified, and for that information to be aggregated for a custom group of students or for the entire cohort. (See Figure 1, opposite.)

Value added and relative effectiveness indicators for secondary schools

A second significant type of additional information presented was the measurement of value added. Work on value added and relative effectiveness indicators was undertaken from 1995 (NSW Department of School Education, 1997a) and the models stabilised in 1998 (Smith, 2005). The notion of value added, as distinct from growth, is to use performance on one measurement scale at a particular point in time to predict subsequent performance on a different measurement scale. For example, using student performance in the Basic Skills Test to predict and measure subsequent performance in the NSW School Certificate.

These additional levels of analysis for school and system use, such as growth, value added and relative effectiveness enabled school, regional and central personnel to grapple in sophisticated ways with school effectiveness issues. Principals could see that a system performance analysis was based on more than just raw scores.

Table 3: Correlations between BST Year 5 predictor scores and School Certificate student course scores for 2008

Course	Correlation
English	0.75
Mathematics	0.78
Science	0.73
History	0.65
Geography	0.67
Computer Skills	0.72

An additional value added indicator for secondary schools is the use of the Year 10 School Certificate aggregate measure as a predictor of subsequent Higher School Certificate performance (correlation = 0.790). An example of the depiction of value added in SMART is shown in Figure 2.

This was a difficult notion for some principals and teachers to accept and understand, and required a lot of professional learning before it was accepted as legitimate and became valued in schools.

Curriculum links (Teaching strategies)

Arguably the most important development in securing support of large cohort testing in NSW and the subsequent use of the information to drive school and system improvement was the linking of test items with high-quality teaching strategies.

In 1999 the decision was taken to better support teachers with high-quality support for statewide tests, and in the same year hard copy teaching strategies linked to the skills underpinning a number of the test items were developed for the first time. Within the SMART software there was a page reference provided to direct
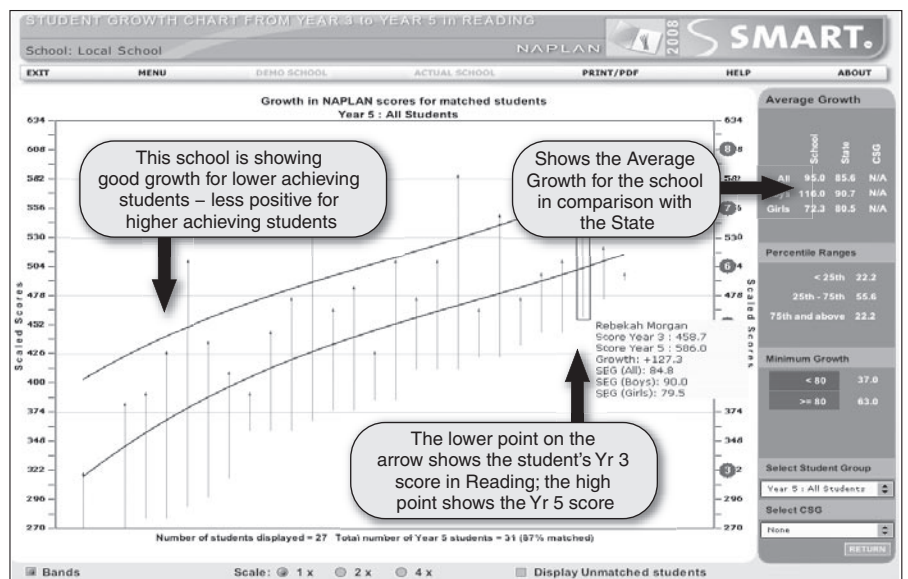

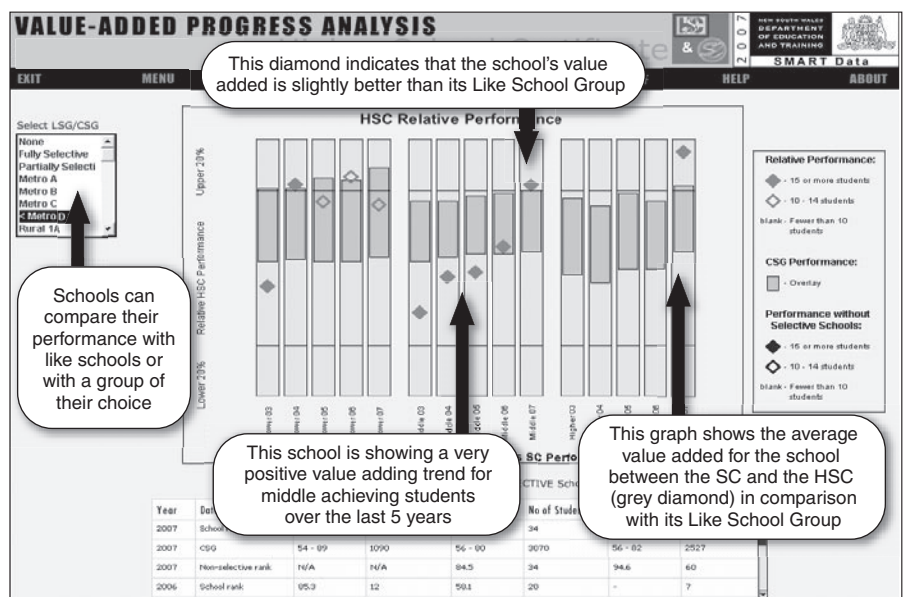**Figure 1:** Reading Growth – BST Yr 3 2006 to NAPLAN Yr 5 2008

Figure 2: Value added between Year 10 School Certificate and Year 12 Higher School Certificate

teachers to the relevant hard copy page in the Curriculum Link document.[7] From 2005 the Curriculum Links were made available electronically within SMART. This process began with the BST and subsequently included ELLA and SNAP.

In 2008, the quality and scope of the teaching strategies was significantly increased to coincide with the implementation of the first NAPLAN test. The strategies were delivered as HTML documents via the Web – as had been the case in 2007 – but every test item in NAPLAN in literacy and numeracy, and for all Years 3, 5, 7 and 9, was linked to the NSW curriculum and the skills underpinning the items were addressed with highly effective and classroom ready teaching strategies. For 2008, in excess of 800 electronic pages of teaching strategies were developed to better support teachers, many with hyperlinks to relevant sites on the Web.[8]

In addition, the strategies were developed within the NSW Quality Teaching Framework[9] and in many cases included a range of strategies for the one skill area for students at different ages and at different levels of ability: strategies for students who require *modelled teaching, guided teaching or independent teaching strategies.*
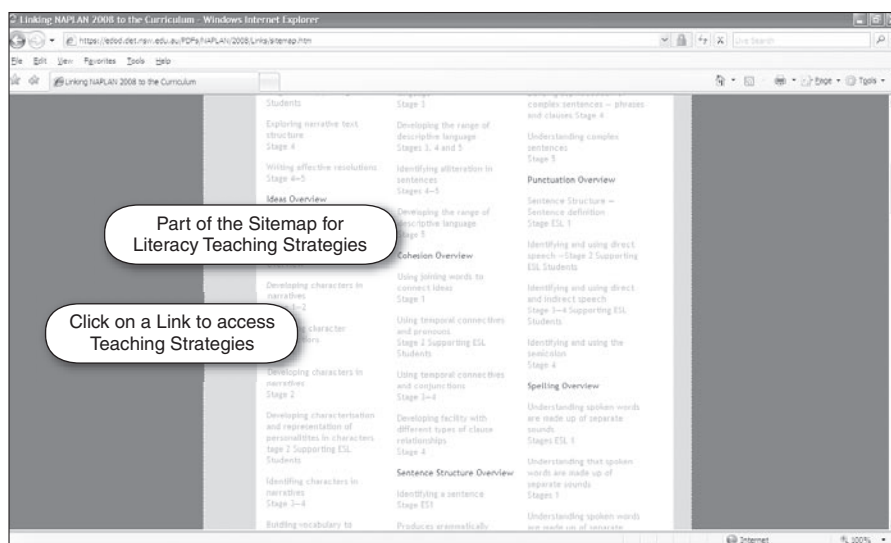

Figure 3: Teaching strategy links in SMART for NAPLAN 2008

The guiding principles for the development of the NAPLAN teaching strategies were:

1. The NSW Quality Teaching Framework (QTF)

2. The Modelled, Guided and Independent teaching cycle

3. The National Statements of Learning for English (SOL)

4. Strategies, and activities to support those strategies

5. Critical aspects of literacy development K—10 continuum (NSW Department of Education and Training, 2008).

This focus on student diagnostics and supporting teachers has been particularly successful in gaining support for large cohort testing across the NSW educational community. For 2009, the NAPLAN teaching strategies will be further developed to address skill areas that were tested for the first time in 2009, or where existing strategies require enhancement or redevelopment.[10]

School and regional performance graphs

In 2005, EMSAD undertook further development work on school and regional performance indicators based on assessment data from large cohort tests. These data were presented on XY scatter plots, using variables that research undertaken by Dr Geoff Barnes (from EMSAD) indicated had the greatest influence on student learning outcomes. These variables were IRSED (Indicators of Relative Socio-Economic Disadvantage), ARIA (Accessibility/Remoteness Indicators for Areas), student attendance and teacher attendance. They were used for the 2006 and subsequent tests. It is important to note that the research undertaken by Barnes indicates that there is no correlation in NSW between teacher attendance and the quality of student outcomes.

The two performance measures analysed in relation to these variables was raw performance, for example, average Years 3 and 5 mean scores for 2008 NAPLAN; and value added measures for junior and senior secondary schools. As of 2010, growth will be included between Years 3 and

---

7 Cordaiy, R., Interview by Dave Wasson, 30 June 2009. Robert Cordaiy is the Manager, School and System Measurement and Analysis, Educational Measurement and School Accountability Directorate, NSW Department of Education and Training.

8 O'Donnell, K., Interview by Dave Wasson, 3 November 2008. Kate O'Donnell is R/ Assistant Director for the Educational Measurement and School Accountability Directorate, NSW Department of Education and Training. Ms O'Donnell is the NSW DET representative on the NAPLAN Project Reference Group.

9 Further information about the Quality Teaching Framework can be found at: http://www.curriculumsupport.education.nsw. gov.au/qualityteach/index.htm

10 O'Donnell, K., Interview by Dave Wasson, 3 July 2009.

5, Years 5 and 7, and between Years 7 and 9. The kind of performance information depicted in Figure 4 below has been used extensively to identify and share best practice, and to identify schools at a regional level for closer monitoring and specific support through the 'Focus Support School' model which is having a demonstrable impact on a number of schools.

At the same time, a Like School Group (LSG) methodology was developed to meaningfully compare schools. This was welcomed by principals, especially when their school was remote; in a low socioeconomic status area; had a high proportion of Indigenous students; or more especially if all three factors were present. These principals maintained it was indefensible to compare their performance with that of the state average, for example. Comparisons with a LSG to a certain extent levelled the playing field and were largely supported (more than 60 per cent of NSW government schools voluntarily report their outcomes against their relevant LSG in mandatory annual school reports). A LSG structure was developed that is reflected in Figure 5 below.

While this model represented a significant step forward in terms of interpreting school performance within the context of the two community factors that explain the greatest amount of variation of performance in NSW (SES and remoteness), the relatively arbitrary cut-points for the various groupings created disquiet amongst some principals. For example, there were 239 primary schools in the Metro C group. This meant that while there may have been some justification for comparison with the mean performance of schools in Metro C, no one could argue that a school at the cut-point with Metro D was similar to a school at the cut-point with Metro B. A more defensible and more equitable model was required.
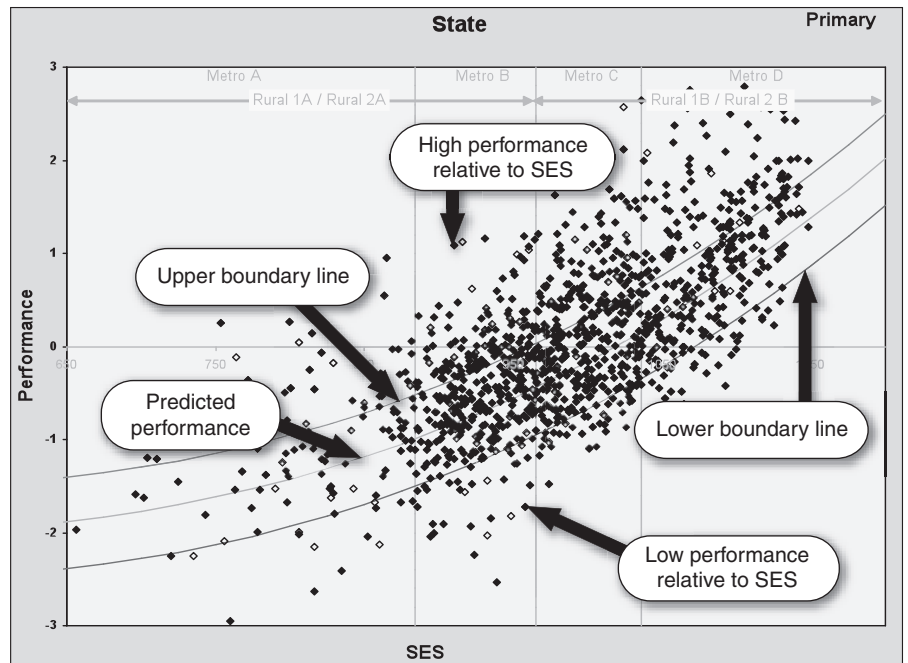

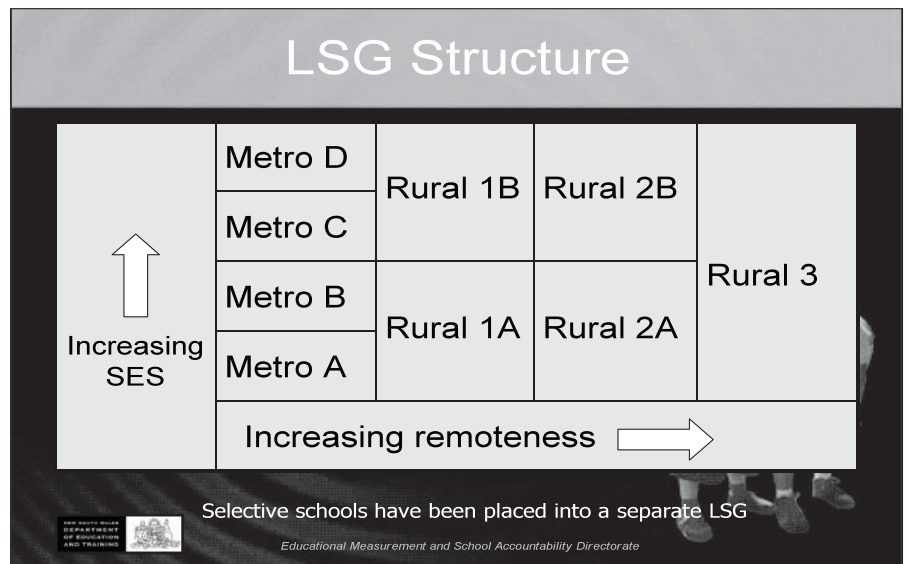**Figure 4:** School performance relative to SES

Figure 5: NSW Like School Group structure – 2005– 2008

Assessment and Student Learning: Collecting, interpreting and using data to inform teaching

School Community Education Advantage (SCEA)

The pathway to develop a new form of LSG model came from the work undertaken by ACER (Masters et al., 2008) and commissioned by the Department of Education, Employment and Workplace Relations (DEEWR). Masters et al. advocate a 'statistical neighbour' approach, such as that which is used in Ontario, that allows schools to compare performance with schools that are most like them on various measures.

To undertake this analysis, the three main community influences on school aggregated outcomes were used: socioeconomic status (as measured by the ABS Index of Education and Occupation); remoteness (as measured by ARIA); and percentage of Aboriginal enrolments.

The table below shows the correlations between these measures and the school performance measures.

Table 4:
Correlations between community variables and school performance

	Primary	Junior secondary
SES	.772	.653
%Aboriginal	.555	.428
ARIA (rural schools only)	.293	.274

Note:
1. SES correlations are based on the ABS IEO (Index of Education and Occupation) SEIFA measure.
2. Correlations based on analyses of NSW DET data.

Schools are ranked according to their values on the SCEA scale. The graph below plots the SCEA values for all NSW government schools against overall performance measures, and demonstrates the process
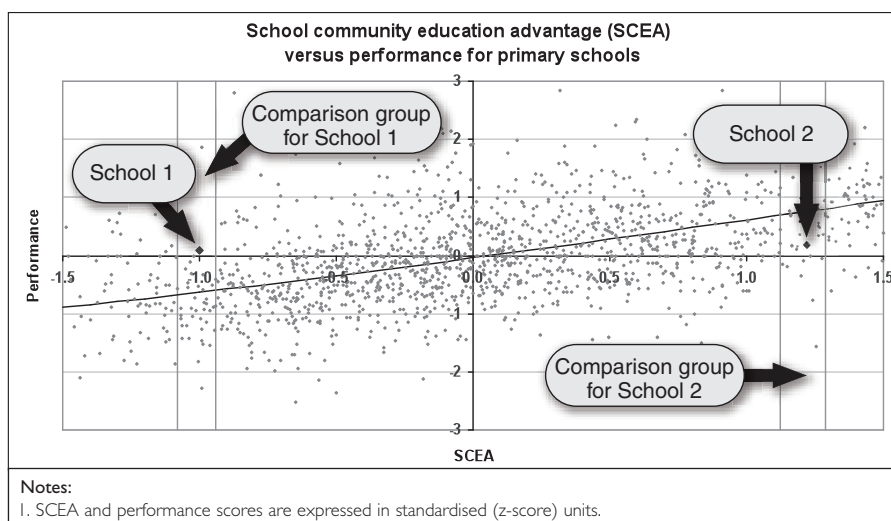

Figure 6: School performance relative to SCEA

for generating like school group comparison data. Each point on the graph represents a school. The position of the school on the horizontal axis is determined by its SCEA value. The comparison group for a given school comprises the 20 schools to the left and the 20 schools to the right of that school. For example, the vertical lines either side of School 1 and School 2 encompass the schools that would form their respective comparison groups. Note that the performances of the comparison group schools can vary considerably because of in-school factors. The average outcomes for the comparison group schools become the like school comparison data for that school (Barnes, 2009).

The significant advantage of this model over the previous NSW LSG model is that at each point along the SCEA scale the comparison group of schools changes. In this way, apart from the two extremes at either end of the SCEA scale, with about 1600 primary schools in NSW, there is potentially 1520 different, or 'floating' LSGs. Discussions with executive members of both the Primary Principals' Association and the Secondary Principals' Council in NSW indicate strong support for this revised form of LSG comparison model.

Future challenges

EMSAD is working towards implementing online testing for large cohorts which potentially has numerous advantages over current pencil-and-paper approaches. These include, primarily, the capacity to assess a greater range and depth of syllabus outcomes and the provision of more timely diagnostic feedback to teachers, parents and students. With the current four-month lag between testing and reporting in NAPLAN, for example, the relevance and utility of the diagnostic information provided is sometimes questioned.

The Essential Secondary Science Assessment (ESSA) will extend earlier online developmental work undertaken with the previous Computer Skills Assessment for Year 6 (CSA6), and transition to a fully online science test for Year 8 students in 2011. There is already an online element to ESSA – the Online Practical Component (OPC). This is an innovative approach to the assessment of science as it creates the elements of a science laboratory online so that sophisticated scientific experiments can be replicated. See Figure 7 for an example of one aspect from an online experiment.

Figure 7: Replicating a science laboratory online in the Year 8 science test – ESSA

Large Cohort Testing Quadrant

Efficiencey in testing & marking →

Range & depth of assessment →

○ C   ○ E

○ B

○ D

○ A

Possible locations of the following:

A – Standard pencil & paper

B – CSA or other simple computer-skills tests

C – Assessment Item Databank (AID) concept, or other e-learning systems

D – Current HSC

E – ESSA Onling Practical Component

Figure 8: Dimensions of testing – Efficiency versus range of syllabus outcomes

Figure 8 presents various forms of testing in NSW on a matrix, in terms of efficiency and immediacy of feedback on the vertical axis, and capacity to measure a range of syllabus outcomes on the horizontal axis.

The limitations of standard pencil-and-paper large cohort tests, represented by 'A' in the matrix are arguably that they are inefficient, they do not provide diagnostic information back to teachers and the system in a timely manner, they are limited in their capacity to assess a range of syllabus outcomes, they are expensive and they are environmentally unfriendly.

The technological capacity currently exists to transition from pencil-and-paper tests to an online environment, where it is possible for the instant scoring of student responses, online assessment of written responses and the possible assessment of a greater range of syllabus outcomes. The challenge remains to implement the change.
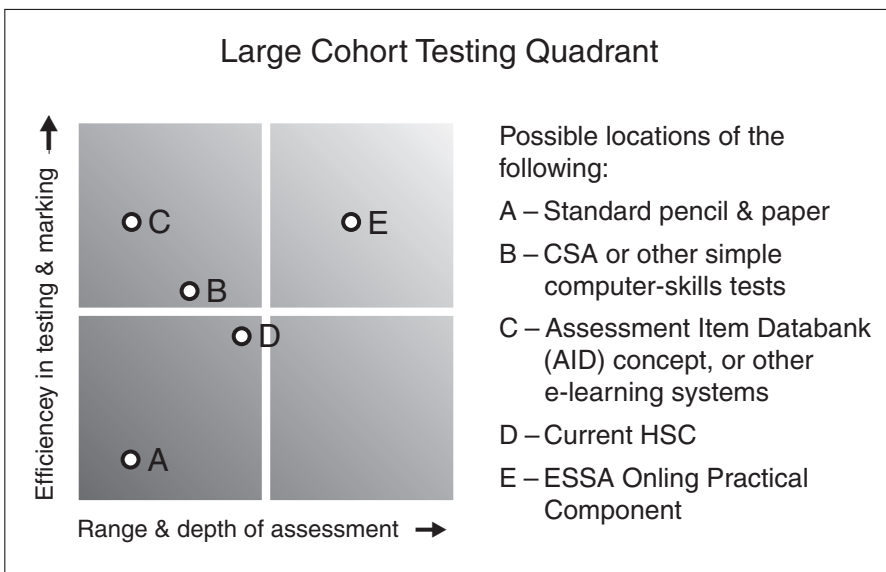
Conclusion: Lessons from the NSW experience

Large cohort testing can have a positive impact on school and system outcomes, particularly and most importantly in the area of improved student outcomes when:

- Driven by a rigorous, relevant and pedagogically sound curriculum framework

- Supported by extensive and relevant professional opportunities for teachers

- Assisted by sophisticated diagnostic tools for the analysis of individual, group, school and system performance

- Accompanied by central and local consultancy support and high-quality support materials.

References

Barnes, G., (2009). *Proposed model for constructing like school data.* Sydney: Educational Measurement and School Accountability Directorate, NSW Department of Education and Training.

Chadwick, V., NSW Legislative Council Hansard, 28 April 1992.

Cizek, G. (2005). 'High-stakes testing: Contexts, characteristics, critiques, and consequences'. In R.P. Phelps (Ed.), *Defending standardized testing.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Eltis, K., and Crump, S. (2003). Time to teach, time to learn: *Report on the evaluation of outcomes assessment and reporting in NSW government schools.* Sydney: NSW Department of Education and Training.

Mamouney, R. (1995) 1995 *Review of the Basic Skills Testing Program.* Sydney: NSW Department of School Education.

Masters, G., Lokan, J., Doig, B., Khoo, S.-T., Lindsey, J., Robinson, L. and Zammit, S., (1990). *Profiles of Learning: The Basic Skills Testing Program in New South Wales 1989.* Melbourne: ACER.

Masters, G. N., Rowley, G., Ainley, J., and Khoo, S.-T., (2008). *Reporting and comparing school performances: paper prepared for the MCEETYA Expert Working Group to provide advice on national schools data collection and reporting for school evaluation, accountability and resource allocation, Australia,* [electronic resource]. MCEETYA, 2008. Retrieved [month] [day], [year], from http://www.appa.asn.au/images/news/mceetyareporting20090420.pdf

NSW Department of Education and Training (2004). *State Literacy Strategy: Evaluation 1997–2003.* Sydney: NSW Department of Education and Training.

NSW Department of Education and Training (2008). *Welcome to linking NAPLAN 2008 to the curriculum,* [electronic resource]. Retrieved [month] [day], [year], from

https://www.det.nsw.edu.au/directorates/schoimpro/EMD/naplan/pubs/Naplan08CL/index.htm

NSW Department of School Education (1997a). *Understanding value added.* Sydney: NSW Department of School Education, Corporate Performance Directorate.

Smith, M, (2005). *Getting SMART with Data in Schools: Lessons from NSW,* Research Conference paper, Melbourne: Australian Council for Educational Research.

Interviews

Cordaiy, R., Interview by Dave Wasson, 30 June 2009.

Lind, P., Interview by Dave Wasson, 2 July 2009.

O'Donnell, K., Interview by Dave Wasson, 3 November 2008.

Wasson, L.J., Interview by Dave Wasson, 19 September 2007.