



This paper argues that international, national, and State and Territory programs measure up in terms of the planning of programs and the quality of data collection procedures. The area in need of greatest attention is in relation to reporting and using the data. Key stakeholders – school administrators and teachers – are the most neglected when it comes to making information from monitoring programs accessible to them.

Introduction

Assessment systems in Australia are grounded in the belief that being explicit about learning goals and measuring students' progress toward these same goals will help to improve student learning. To this end, assessment programs are developed to measure in terms of learning goals. Information from the programs then provides an indication of how students, schools and systems are travelling and how successful programs and initiatives have been and provide the information needed to shape programs at the macro level of public policy and the micro level of teaching programs within the classroom. Eva Baker and Robert Linn (2004) refer to this as a theory of action and Eva Baker (2004), describes the process in these simple terms:

Figure out what should be taught, be prepared to teach it, help students learn it, measure their learning, and continue the cycle until desired improvement is met.

To put it more formally, this process can be described as a series of steps.

Step1	What students should know and be able to do is agreed and explicated.
Step 2	The extent to which this is being achieved is measured.
Step 3	The broader educational enterprise that includes bureaucrats, administrators and teachers throw their efforts behind ensuring students reach those goals.

In 2002 Margaret Forster, in a presentation to the Seventh Roundtable on Assessment in Canberra, conceptualised a framework for judging the quality of system-wide monitoring programs. She began by providing a checklist for judging how system assessment programs measure up. These came under three headings:

- Planning the program (clarity of purpose, resourcing and sustainability);
- Collecting the data (validity and reliability); and
- Using the data (informing policy and reform).

Using this checklist and the framework is a useful way to interrogate the efficacy of testing programs, particularly the reporting processes.

Planning the program (clarity of purpose, resourcing and sustainability)

Testing programs in which Australian States and Territories participate measure up well in terms of their clarity of purpose and their alignment with the

designated curriculum and reporting frameworks. All States and Territories have been assessing and reporting system performance for at least a decade. Tasmania and New South Wales have been doing so for much longer, since the late 1970s and late 1980s respectively. All jurisdictions have units dedicated to overseeing work done by contractors on their behalf or they have established the in-house capabilities. All jurisdictions have contributed resources to support both national and international assessment programs. Technical, logistic and financial resources are made available by all jurisdictions to ensure the sustained operation of international, national, and State and Territory programs.

Websites of all testing programs in which Australian States participate make it clear that this is a shared vision of the purpose of all those programs. They describe their fundamental purpose as being to lever change by informing action that improves student outcomes. For example, the purpose of the OECD study, Programme for International Student Assessment (PISA) is described in the Australian report as follows:

PISA was designed to help governments not only understand but also enhance the effectiveness of their educational systems.

The Victorian Curriculum and Assessment Authority's (VCAA) website describes the purpose of the AIM program as:

... providing an indication of how well the literacy and numeracy skills of students are developing ... The results provide information used to plan new programs and a useful source of feedback and guidance to students, parents and teachers.

Tasmania's Department of Education, in its Assessment Monitoring & Reporting Policy, states:

It is intended that, through this policy, student learning outcomes will be improved by assessment, monitoring and reporting practices that:

- *are integrated into teaching and learning processes;*
- *inform decision-making about teaching and learning;*
- *provide useful and timely feedback to students, parents and teachers; and*
- *enable accountability requirements to be met at student, school, Department and government levels.*

It is apparent then that testing programs in which Australian students participate intend to stimulate appropriate educational reform by providing information and insight to stakeholders so that the required interventions can be made. Being clear about the purpose of assessment programs does not, however, in itself make them successful.

Collecting the data (validity and reliability)

The testing programs in which Australian students participate measure-up well in terms of the quality of the data collected. Jurisdictions ensure that the processes of measuring performance are robust. International and national assessment programs have national committees that oversee and endorse processes and procedures. These committees require, and are given significant information to assure them of technical and measurement veracity. Since 1998 when jurisdictions around Australia began to collaborate to ensure reporting of national comparable data, the psychometric and curriculum integrity of jurisdictions' own testing programs have been subject to close scrutiny, by each other and by the Commonwealth. While psychometric rigour is necessary for a successful system-level assessment program it is not in itself sufficient.

Using the data (informing policy and reform)

This is the area of greatest weakness in all assessment programs undertaken by Australian students. The talent and creativity that is invested in the development of the assessment and the analyses is not matched when it comes to disseminating and using the information at the local level. The degree to which programs are integrated into the larger educational context has been relatively limited, as is use of information to shape programs at the macro level of public policy and the micro level of classroom programs. This is a significant weakness because if the accountability mechanisms do not positively affect the quality of public policy, school practice and classroom teaching then regardless of their other strengths, the accountability mechanisms themselves are failing.

Dissemination of significant information from the monitoring program in a way that is accessible to a range of stakeholders

While there is undoubtedly a wealth of information from the various assessment programs, it often does not get communicated in a particularly timely or accessible way. To illustrate and explore this assertion I will draw largely on Western Australian experience, however discussions with colleagues from other jurisdictions suggest that observations hold true across other States and Territories.

Joan Herman (2005) refers to both the symbolic and technical functions of accountability systems. I believe that sample testing programs such as Western Australia's MSE program, the national Primary Science Assessment Program (PSAP), the international programs, PISA and TIMSS, currently have negligible symbolic or technical function at the classroom level because the reports, while comprehensive, do not 'speak' directly to

classroom practitioners or even school administrators. Sample programs cannot provide valid individual student level information and often can provide only limited school level information. The complexity of the analyses and depth of the reporting result in quite a time lag between testing and reporting, so that the limited information that is given back to schools may refer to a cohort of students that has moved into a different phase of schooling.

This in itself does not make the information irrelevant, but in the school environment where the operational challenges always have the face of a student, parent or teacher attached to them, this information has a level of abstraction about it that allows it to be put in the 'Interesting, I'll attend to that when I get time' pile. However it is my observation that when time is made for teachers and administrators to learn about a testing program such as PISA and the implications of its findings, they are highly responsive to the information and make the connections of its relevance to their own context very quickly.

For example, the National Advisory Committee for PISA in early 2003 were advised of the importance of Australia meeting stringent sampling requirements, and it was reminded of the struggle it had been to get the sample 'over the line' in 2000. The demand on schools' time and resources was greater in 2003 and in Western Australia we could see that for many schools it would be so much easier for schools to say 'no' than 'yes' to a request to take part in the 2003 sample of a testing program that many had never heard of. A plan to tap into the symbolic power of PISA as an accountability mechanism was developed.

All schools selected in the sample were invited to send their principal and one other staff member to a half-day meeting.

At this meeting information was given about PISA, an overview of the results of the 2000 study were shared and their implications (including the significance of PISA internationally, nationally and for schools) were discussed. There was a plenary session that ended up with the group 'problem solving' the school-level impediments to successful participation in the sample. Key people in the sample schools went away knowing some essential things about PISA. In a morning they acquired sufficient information for them to place value on the results and their participation in the program and the outcome of this very modest professional development session was that all Western Australian Government schools subsequently agreed to participate in PISA 2003.

Recognising the importance of taking time to explicate aspects of the programs for key stakeholders grew out of our experience with the Western Australian Literacy and Numeracy Assessment (WALNA) program. This program had been introduced in 1998 under a storm of opposition and anger from educators. In early 1999 an evaluation was conducted to formally gauge parent and teacher reactions to the program. The results indicated relatively high levels of mistrust and dissatisfaction amongst teachers with WALNA.

Fairly incidental work done with schools in interpreting their data suggested that, even in schools where they were trying to use the data for school improvement purposes, the level of knowledge about the assessment was an impediment to efficient and effective use of the data. We also observed more than the occasional instance where teachers and their principals were over-interpreting the data and, as a corollary, their teachers were disposed to over-prepare for the assessment at the cost of curriculum balance.

From 1999 to 2003 two significant programs were introduced to support better use of the data by key stakeholders of WALNA – principals and teachers. The first to be introduced was the Data Club which targeted principals. The Data Club supports school leaders in making performance judgements based on their school's WALNA data.

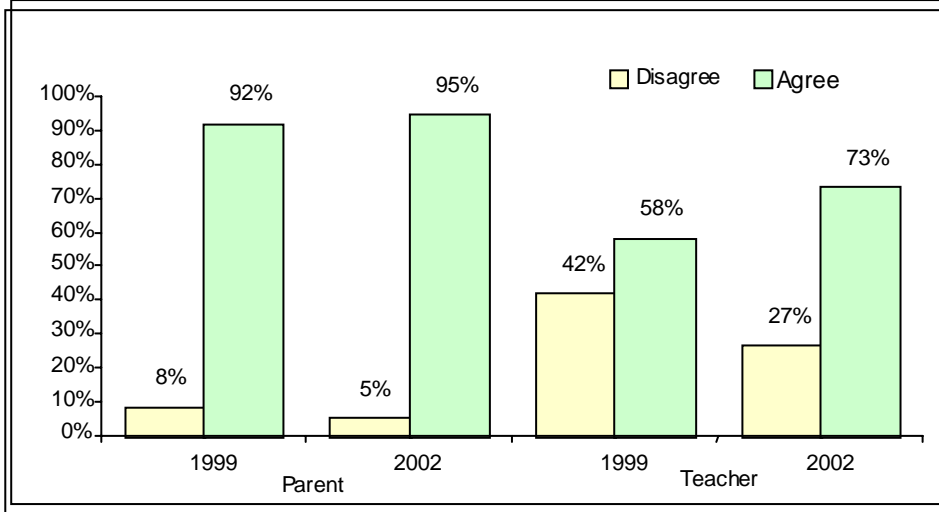
The second program which, targeting teachers, began life as the 'Teachers' Data Club', has since been rebadged as 'Assessment for Improvement'. The aim of this professional development program is to increase teachers' confidence in judgements they make from a range of assessments. Teachers' analysis workshops have been specifically designed to build teachers' ability to blend their classroom monitoring with WALNA results to judge student achievement and plan for future teaching and learning.

While both programs were about understanding data, they were built to meet the needs identified by principals and teachers. Beyond the initial data provided, the displays and graphs included in WALNA reporting have been ones requested by principals and teachers, rather than all the ones that a powerful statistical software packages can generate.

A subsequent evaluation was carried out at the end of 2002. Using the (then) original questionnaire, a representative sample of teacher and parents was canvassed as to their opinion of the key aspects of the WALNA. Parents never doubted that external assessment would support schools to provide better literacy teachings (92%) and this view firmed up over the intervening three years (95%). In 1999 close to half (42%) of the teachers surveyed disagreed that system level test information would assist schools in providing better literacy teaching. By 2002

that percentage had reduced significantly (down to 27%).

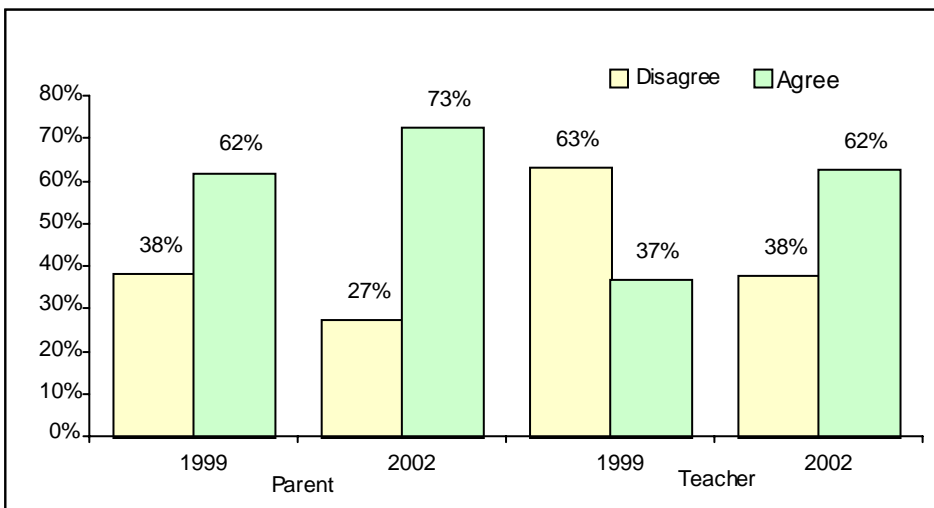
Question: Assist schools to provide better literacy teaching for all students



There was also a significant shift in how useful teachers perceived the WALNA results to be. In 1999, while parents found the additional information provided by WALNA useful, a majority of teachers (63%) did not agree that they found the information useful. By the 2002 evaluation, 62% of teachers agreed that the information was useful – a significant shift on opinion.

Question: The report gave me additional information not available in the regular school report

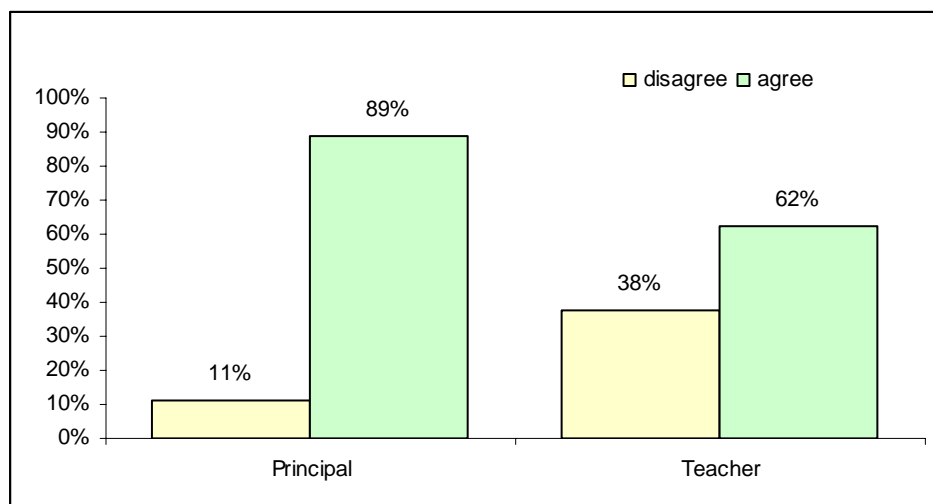
Question: The test results provided me with valuable diagnostic information about my students



The influence of Data Club is suggested in the following chart. Remember that Data Club began in earnest in 2000, so school principals had earlier access to professional development on system-level data than teachers. In the 2002 evaluation, both principals and teachers were asked about the usefulness of the results for diagnostic purposes. Principals responded significantly more positively.

Question: The data on individual students is useful for diagnostic purposes

Question: The test results provided me with valuable diagnostic information about my students



Interestingly, principals who had not had access to the Data Club differed significantly in their responses to other principals. These principals were less likely to:

- provide the data to the current or next year's teachers or the school;
- think their staff were confident in using the data;
- involve the school council in interpretation of the data or share the data with the P&C;
- take it into account when reviewing curriculum plans ;
- belong to the Data Club;
- find the data on individual students useful for diagnostic purposes or for determining rates of learning; or
- use it to track student performance.

Furthermore, these principals were more likely to disagree that the results confirmed strengths and weaknesses of the school curriculum or that the results gave a good overview of teaching in the WALNA years.

Lessons learnt about reporting system level assessment to schools

The main lesson is that if you want to engage teachers in discussion about the

meaning of system assessment, talk to them about the students that they teach and are crucially interested in helping. Use the data to help them articulate what their students can do and what they are struggling with. Let the data shape teachers' 'stories' about their students. Abstractions about strengths and weaknesses observed in the performance of subgroups and the whole population are less useful as a starting point. It follows therefore that it is easier to engage teachers with the meaning of test data when you are working with population testing data. You are talking to them about 'their kids' and their teaching program, and it is irresistible!

Sample programs have richer information, but you have to work harder (but not that much harder) for teachers to 'get it'. It is intrinsically interesting for teachers to dwell on the effect of their teaching programs and to think about why students respond the way they do. It's the staple of teaching. Teachers think a lot about how to engage students, how to counter a misconception, and how to provide the step that overcomes a block to further development.

All jurisdictions, I contend, are committed to the deep reform that positively alters learning outcomes for students. They demonstrate this by their considerable investment in international, national and local testing programs. Yet a key aspect of the cycle has been missed. Accountability systems will fail if teachers, who have been trained for decades to mistrust test data, are expected to work it out all by themselves.

This is not about producing more complicated graphs. Reporting to schools can be exhaustive and I suspect often exhausting. From the perspective of the test analysts, it can be hard to accept that more is not necessarily better. Getting the amount and pitch of the information sent to schools right can be difficult, especially when the statisticians know they have an amazing set of graphs that they can produce. However, it is really important to listen to school administrators and teachers when planning reporting packages. Once they are 'warmed up' to being 'assessment literate' they will tell you what is useful. There is no point in sending out a small forest worth of paper reports if it simply gets filed away.

Good reporting is about guiding teachers to intelligent interpretation of data that is useful to their work - teaching students effectively. An assessment system that fails to support school personnel in this enterprise is a highly imperfect mechanism.

References

Baker, E.L. (2004), *Aligning Curriculum, Standards, and Assessments: Fulfilling the Promise of School Reform*, CSE Report 645, Los Angeles University of California National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Cresswell, J., Greenwood, L., & Lokan, J. (2001), '15-up and counting, reading, writing, reasoning : how literate are Australian students?', *PISA 2000 survey of students'*

reading, mathematical and scientific literacy skills, Melbourne Acer Press.

Baker, E. L. & Linn, R. L. (2004), 'Validity issues for accountability systems', in S.Fuhrman & R. Elmore (eds.), *Redesigning accountability systems for education*, Teachers College Press, New York, pp. 47-72.

Herman, J (2005), *Making Accountability Work to Improve Student Learning*, CSE Report 649, Los Angeles University of California National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Forster, M. (2002), *Assessing the quality of systemwide monitoring programs: How do we measure up?*, Presentation to the 7th Roundtable on Assessment 'Equating', Canberra.

The Victorian Curriculum and Assessment Authority, (VCAA)
<http://www.vcaa.vic.edu.au/prep10/aim/>
Department of Education, Tasmania, Office of Educational Review, Assessment, Monitoring and Reporting Policy,
www.education.tas.gov.au/oer/AMRpolicy/default.htm