



Introduction

Modern school education is a complex business. To the general public, it sometimes appears to be a maze compared to the straightforward system they 'went through'. Professor Richard Smith of Central Queensland University has suggested that schools of old, and our old idea of schools, were something like a stockyard with two gates. No wonder you 'went through', or that it felt like a tick dip in the middle. Stockyard or maze – two metaphors for our perceptions of schools, and neither a happy one.

Instead of either, Richard Smith suggests we should think of school education as being like the London Underground, with multiple entries and exits, recursive paths, alternative ways of getting to the same place, multiple entry tickets, and so on. This is a neat metaphor for the necessary and generally beneficial complexity now faced by students and families. And it's a reassuring metaphor: education is a complex environment nowadays, but it allows for many individual choices and pathways, with an underlying orderliness. It's a metaphor to aspire to, anyway. I'm going to add another element to the metaphor. The London Tube is famous for its warnings on every platform, 'Mind the Gap'. That gap is between platform and train, somewhere you do not want to put any part of your anatomy. This conference is about another gap – the gap that can open between curriculum and assessment. In the education 'tube', we must be very

mindful of that gap. Here I'll risk stretching the metaphor too far, because I think that sometimes we need to preserve the gap, and sometimes we need to close the gap.

In Queensland at the moment we are having a major debate around the interconnected relationship of curriculum, teaching, assessment and reporting. Our local research confirms what I think is the international consensus: that these four practices have to be in alignment for the best educational outcomes. Nevertheless, there are priorities and logics, and we have to be wary of allowing say assessment or reporting to drive curriculum and teaching. At least without reflecting seriously on what it is we are trying to achieve.

The axis of curriculum, teaching, assessment and reporting can be viewed from a number of positions – from the point of view of the teacher, the parent, the student, the school principal, or the local education system administrator – or from my position, as a manager and policymaker for the system as a whole. When it comes to assessment, at the fundamental level of the teacher-student relationship, it's mainly about monitoring the student's progress, modifying teaching strategies to adjust to what you've found, and clearly reporting results. From my point of view as a system manager and a policy adviser, those fundamental uses of assessment do not fade away, but they are joined by other needs. I must be concerned with all

students, not just specific individuals, with the viability of the system as a whole, and its accountability to parents, the government and the broader community. In that context, assessment becomes a potentially valuable tool for monitoring and reporting on the performance of the system overall.

So, much as I am concerned with educational fundamentals, I am responsible to the government, and I act on policy directions of the government. What is the role of government in an education system? How much should a government intervene in curriculum and assessment? How should a government 'mind the gap'?

'Hell and the High School'

In 1925 a schoolteacher in Tennessee was tried for teaching 'evolution'. The teacher's name was John Scopes, and the trial became known as the Scopes Monkey Trial. The story is well known across the world, thanks to the movie, *Inherit the Wind*, in which Spencer Tracy and Fredric March argue the case. The title of the movie comes from Proverbs 11: 29, 'He that troubleth his own house shall inherit the wind ...'

The trial was made possible by an *Act* passed by the Tennessee legislature, known as the *Butler Act* after the politician who introduced it. An *Act* of a parliament actually forbade the teaching of evolution in universities and schools. And a teacher was actually taken to court for teaching contrary to the *Act*, and fined for his offence. The Monkey Trial is perhaps the most famous and extreme example of a government intervening in the curriculum of public schools. Though, of course, looking historically across a broader range of educational facilities, teachers have at times been burned at the stake, or forced to take hemlock. Far be it from me to suggest that such practices be considered for reintroduction.

So why do I raise this extreme, perhaps hackneyed and remote, example of a government intervening in school curriculum? Surely governments today, especially those far from Tennessee (or Kansas, anyway) would not intervene in such a way? The reason I raise it is as a limiting case. Discussion of the Scopes' Monkey Trial historical case sets the boundary of the argument.

I want to say that school curriculum is a part of a government's remit. Government must concern itself with the quality and effectiveness of school curriculum, and to that extent it must monitor the effectiveness of the education system constantly, and directly intervene at times. So curriculum, teaching, assessment and reporting form an important site for government action. However, this shouldn't be seen as compromising the need for independent, non-partisan education. It shouldn't be seen as a threat to giving accountable discretion to professional educators to see that students are taught in the best way compatible with current knowledge. In the space between the Monkey Trial and a totally laissez-faire approach, I believe governments have both room and responsibility to take action about curriculum.

The second reason I mention the Monkey Trial is to illustrate the methods used by governments to influence curriculum. The Monkey Trial is a rare and antique case, in that black letter law was used to compel compliance in public schools. I'm not an expert on the constitutional arrangements in Tennessee in 1925, but I suspect the *Butler Act* was slight overkill. You'd have to think that the Government of Tennessee could have simply issued an administrative direction to its schools. And this is more commonly the approach. In Queensland in the late 1970s, a famous contretemps occurred, in which the

government of the late Sir Joh Bjelke-Petersen, urged on by a religious lobby group, took objection to a couple of new curriculum programs. The government didn't have to legislate to remove them from state schools; it just did it. A contemporary government wishing to impose an ideological curriculum on its schools would not have to resort to the blunt instrument of Butleresque legislation. And if it was of a Machiavellian bent, it might not even rely on administrative mandate. It would simply require that students would sit tests in which knowledge of creationism was included and knowledge of evolution omitted. And it would fund schools against the outcomes of the tests. It would leave time pressures and resource hunger to do the rest. This mechanism is available to governments for benign purposes as well, of course. And the mechanism is effective even in a federal system, where the federal government does not run schools, and may not usually make laws about them. In two familiar federal systems, ours and that of the United States, the federal government provides significant funds to education. So even though lacking a direct mandate over schools, it may use performance-based arrangements, including test-based accountability regimes.

Testing and policymaking

Teachers are well aware that assessment can be used for a range of purposes. In teachers college, they are informed that there are two main types of assessment – formative and summative. I am advised by folk with classroom experience that out at the chalkface, there are other handy uses for assessment – for example, punitive and sedative. I am further advised that the sedative type can be profitably used last lesson Friday, and that students should mark each other's papers to avoid impinging on the teacher's weekend. However, I treat this advice with scepticism.

But tests certainly have other uses for system administrators and policymakers – system monitoring and reform prominent among them. One American expert (Linn, 1994) summarises the appeal of tests for advancing reform thus:

Assessment has great appeal to policymakers as an agent of reform for a number of reasons.

(1) Tests and assessments are relatively inexpensive. Compared to changes that involve increases in instructional time, reduced class size, attracting more able people to teaching, hiring teacher aides, or programmatic changes involving substantial professional development for teachers, assessment is cheap.

(2) Testing and assessment can be externally mandated. It is far easier to mandate testing and assessment requirements at the state or district level than anything that involves actual change in what happens inside the classroom.

(3) Testing and assessment changes can be rapidly implemented. Importantly, new test or assessment requirements can be implemented within the term of office of elected officials.

(4) Results are visible. Test results can be reported to the press. Poor results in the beginning are desirable for policymakers who want to show they have had an effect. Based on past experience, policymakers can reasonably expect increases in scores in the first few years of a program ... with or without real improvement in the broader achievement constructs that tests as assessments are intended to measure.

Of course, those advantages are pragmatic. They presuppose that testing is a relevant and effective way of achieving the specific necessary reform.

At the general level, another expert (Stecher, Ch 4 in Hamilton et al, 2002) summarises the plusses and minuses that

tests may bring in their impact on policymakers:

Positive

- *Help policymakers to judge the effectiveness of educational policies*
- *Improve policymakers' ability to monitor school system performance*
- *Foster better allocation of state educational resources.*

Negative

- *Provide misleading information that leads policymakers to suboptimum decisions*
- *Foster a 'blame the victims' spirit among policymakers*
- *Encourage a simplistic view of education and its goals.*

So it depends on getting the balance right. Testing is a necessary part of formal education. Done well, it is an ideal way of assessing some aspects of what students can do. Good classroom assessment is vital, as research here in Queensland has shown. But high-stakes testing as a means of assessing systems and schools, students and teachers, of imposing accountability or of motivating reform, needs to be considered very carefully.

In the early 1970s, Queensland abolished the external senior examination. From that time, we have pioneered internal, progressive assessment in our schools. In time, this became criterion-referenced. And we have buttressed it with external moderation and a standardised benchmarking test. The system works well for our senior students. It is an effective solution to the problem of balancing a range of concerns.

On the one hand, less assessment or less centralised assessment may be seen to foster or serve:

- responsiveness to local needs and diversity
- pedagogical innovation

- professional discretion and educational leadership
- ideological neutrality
- primacy of curriculum.

On the other hand, more assessment or more centralised assessment may be seen to foster or serve:

- portability and buying power of credentials
- communication of expectations and achievement
- motivation to excellence
- intellectual rigour and stretch
- alignment of curriculum, teaching, assessment and reporting.

Most of these, on both sides of the balance, are desiderata of a publicly supported education system in a diverse modern democratic state. The resolution cannot be as simple as more or less assessment, or centralised or dispersed control of assessment.

Why we need consistency and standards

At the moment there are compelling reasons to examine the side of the balance in favour of common standards and assessment that supports those standards. Queensland has advocated strongly for greater national consistency in curriculum, assessment and reporting standards, and presently leads the Nationally Consistent Curriculum Outcomes project and Performance Management and Reporting Taskforce under the auspices of MCEETYA. One reason for our commitment is our recognition of demographic reality. I've heard that each year across Australia, more than 80,000 children migrate between States. That's more than the student population of Tasmania. In Queensland alone, we have 11,500 students arrive each year from interstate. Within Queensland, we find that 25% of our students change addresses each year. And thousands of our teachers and

educational leaders change jobs or locations. When we argue the need for a nationally consistent curriculum, we are not talking about everyone doing *The Harp in the South* in Semester 1 of Year 11 in Townsville and Birdsville, let alone Cairns and Perth. But we are talking about a core of essential curriculum and consistent standards that those migrating students will take with them.

But for me the issue of consistency and standards is not just about pragmatic recognition of demographic reality, or the journey across State borders. It is about a bigger journey still. If you are a student, it is about you knowing yourself, knowing the measure of yourself, and making your way in the world.

Tania Major is a young aboriginal woman who hails from Kowanyama in Cape York. She has a degree in criminology and criminal justice, and now works in the Cape York Leadership Institute under Director Noel Pearson. In 2003, as the Cape's youth representative, she made a speech to the Prime Minister when he visited Kowanyama. In part, Tania said this to John Howard;

... when I went to school in Brisbane it was as if I had missed out on my primary education. ... There is a huge gap between what we get in communities and what other kids get in cities. I got straight As at Kowanyama but when I got to Brisbane I was getting Cs and Ds. ...Prime Minister, we need to review the curriculum in these communities because it's pitched at a very low level. ... Education should be uplifting, not serve to reinforce lack of self-esteem and the heart wrenching low expectations that my mob suffer from. If we cannot get education right then we are doomed.

Standards are the benchmarks of aspiration. They are the currency of achievement – the legal tender recognised in another state or another country. Like it

or not, they are the coin of the educational realm. But don't get the idea that I'm talking about standards at just a low level of basic skills, as reflected performance on a test. Standards must be about high-level intellectual engagement. They are not just about skills, and learning is not simply reducible to testing. Assessment and reporting must reflect and enact high level intellectual engagement, too. This is an absolutely essential way that curriculum, teaching, assessment and reporting must be aligned. You cannot assess high-demand curriculum with dumb tests, nor communicate high-order learning with simplistic reports. All must set high expectations and give realistic, sound feedback.

Recently, the Australian Council of Educational Research published an evaluation of the results for Australian Indigenous students in the first OECD Programme for International Student Assessment (PISA) in 2000. The evaluation found that the positive relationship between educational aspirations and achievement was one of the highest of all the contextual variables. Across all types of literacy tested, as the achievement of an Indigenous student increased, so did their aspiration for further study (De Bortoli and Cresswell, 2004).

Now perhaps that is the causal direction – higher achievers have higher aspirations. But I believe there is also good reason to believe that it works the other way, too – that higher aspirations generate success. And part of our contribution to students' aspirations are our expectations for them and of them. In teaching, this effect is well-known. It is called the 'Pygmalion Effect' where students live up to our high expectations of them – as Galatea was modelled on the expectations of Pygmalion, or as Eliza Doolittle was given a moderately extreme make-over by Henry Higgins.

We must also beware the anti-Pygmalion Effect, where students live down to our low expectation, even where these are the articulation of good intentions. In its worst form this is 'deficit thinking'. In its best form it is just the failure to demand enough in the desire to give every child success, however illusory.

The Tania Majors we teach must know that our expectations for them, and our judgments of them, have been pitched against the best.

Some disturbing trends

Nevertheless, while there is a need for consistency and standards, this should not be turned into a caricature of itself. Education has become a complex undertaking, and it is easy to appeal to people by saluting the simplicities of bygone days. But we should not deliberately appeal to ignorant and authoritarian impulses that are out there in the population. The Australian Government has recently called for a 'Plain English Report Card'. Their example report card gives a hypothetical (male) student receiving it a criterion-referenced 'grade': A, B, C, D, E. It also gives him a norm-referenced rating, by identifying his place in a quartile of the class: Top, 2nd, 3rd, Bottom. Published comments by the Minister, Dr Nelson, indicate that these quartiles will literally be of a single class, with on average 25 to 30 students.

Coincidentally, we in Queensland are about to do something in terms of plain English and criterion-based reporting, so there is potential for some common ground with the Commonwealth. But I'm hoping the example report they have published is a draft, and that they will accept some amendments to it, because I am very reluctant to accept Top and Bottom, and I am very reluctant to accept placement in quartiles of students in a

class of 28 - or fifteen? or seven? And what sort of class - who knows? Placement in a quartile wouldn't have helped Tania Major. It would have been at least as misleading as her straight A's. We need to tell students in plain terms how they fare against solid standards. On the one hand we should be careful not to boost them to false pride with easy praise. On the other hand, we should not subject them to prejudicial language and meaningless pseudo-statistics more suited to pop psychology than solid psychometrics.

I repeat - curriculum and teaching and assessing and reporting must be aligned, in their standards, not just their content. They must all have intellectual rigour to them, not just simplistic appeal. You cannot smarten up in your teaching, and dumb down in your reporting. You can speak plain English and still be rigorous; intellectual rigour is more than just tough talk. I also worry about generational prejudices. Recently Dr Jason Sternberg, a lecturer in media and communication at Queensland University of Technology, has written of the prejudice the television current affairs programs display against young people, knowing that it will play well with their older audience. We adults today are a grumpy old generation. We shouldn't let our schools participate in that generational disappointment and disparagement. The British poet, Peter Dixon, reminds us in a satirical tone, how easily school hierarchies and standards can reinforce social inequalities and gratify personal (or generational) grudges and inadequacies.

Oh bring back higher standards - the pencil and the cane -

if we want education then we must have some pain.

Oh, bring us back all the gone days. Yes, bring back all the past ...

let's put them all in rows again - so we can see who's last.

*Let's label all the good ones (the ones like you and me)
and make them into prefects – like prefects used to be.*

*We'll have them back in uniform, we'll have them doff their caps,
and learn what manners really are ... for decent kind of chaps!
So let's label all the good ones, we'll call them 'A's and 'B's –
and we'll parcel up the useless ones and call them 'C's and 'D's.
... We'll even have an 'E' lot! ... an 'F' or 'G' maybe!!
... so they can know they're useless ... and not as good as me.*

*For we've got to have the stupid –
And we've got to have the poor
Because –
 if we don't have them ...
 well ... what are prefects for?*

The issue of how we should educate – especially children – should certainly be an arena for robust discussion. But it should not be the place for settling antagonisms, or setting up straw men. The field of young people's education above all must be kept free from point-scoring in the name of 'culture wars' or generation spats. Otherwise, that verse from Proverbs will come back with a vengeance, 'He that troubleth his own house shall inherit the wind ...'

Test-based accountability: some concerns

When I emphasise the need for consistency and standards, I am not calling for a reactionary cartoon of irrelevant and intellectually dubious 'standards'. I am also not meaning an excessive use of high-stakes accountability testing on the American model. As I've said, school assessment plays an essential role as an educational tool. And chunked up to a higher logical level, these tests – for example, the national literacy tests for Years 3, 5 and 7

– provide useful system monitors. But as one American authority (Linn, 2001a) has put it,

Assessment systems that are useful monitors lose much of their dependability and credibility for that purpose when high stakes are attached to them. The unintended negative effects of the high-stakes accountability uses often outweigh the intended positive effects.

On the American experience, there is grave risk of perverse behaviours resulting once high stakes (such as school closure or lost funding) are introduced. There is widespread evidence of American teachers and schools responding in ways that are perverse with respect to the object of the accountability regime, but perfectly explicable in the circumstances. One survey (Koretz et al, 2001) summarised teachers' responses as positive, negative and ambiguous.

Positive:

- *providing more instructional time*
- *working harder to cover more material*
- *working more effectively.*

Ambiguous:

- *reallocating instructional time*
- *aligning instruction with standards*
- *coaching by focussing on incidental aspects of the test.*

Negative:

- *cheating.*

The overall consequences of large-scale, high-stakes testing have been inflated scores (especially if the exact same test is repeated), teaching to the test, narrowing of the curriculum, neglect of higher order skills, and the quaintly named 'Lake Wobegon Effect', whereby everybody miraculously scores above the norm. The discussions of the resulting problems frequently dissolve into a 'trilemma', with the following options:

- general tests that are invalid with respect to the local curriculum as taught;
- locally applicable tests that are valid but ungeneralisable to national comparisons; and
- artificial validity achieved by matching curriculum to the test.

Some test proponents have even contested the description 'artificial' in the last option, and opted heartily for it. They have staked a claim on behalf of 'Tests worth teaching to' (Hamilton, 2005). Test-based accountability has had more subtle, less obvious effects as well. For example, it has encouraged teachers to focus on 'bubble kids' - those who happen to be near a significant cut-off that determines school performance and perhaps funding. They realise that getting those kids over the line is an easier way to improve the school's standing than making an effort to genuinely raise the performance of all students or the most at risk. But of these the most pervasive effect is the narrowing and lowering of expected curriculum outcomes. This is sometimes the effect of the test instrument (multiple choice), or just of the selectiveness of what is tested (unavoidable if you insist on giving the same test to everyone). The result, if the stakes are high enough to shape the curriculum, is a disaggregation of learning that is educationally regressive (Shepard, 1989):

The notion that learning comes about by the accretion of little bits is outmoded learning theory. Current models of learning based on cognitive psychology contend that learners gain understanding when they construct their own cognitive maps of the interconnections among concepts and facts. Thus, real learning cannot be spoon-fed, one skill at a time.

Of course it was E M Forster who wisely warned us that, 'Spoon feeding in the long-run teaches us nothing but the shape of the spoon'. Had he addressed the issue

of low-level testing, he might more crudely have added that, 'Regurgitation in the long run teaches us nothing but the shape of the bucket'.

As educators, you are all familiar with Bloom's Taxonomy of Educational Objectives, with its (perhaps arguable) hierarchy that runs from 'knowledge' up to 'evaluation'. A product of teaching to the test is that all objectives reduce to knowledge objectives. Cognitively higher-order tasks can be imitated - learned by rote and regurgitated.

Generally the alignment of assessment and curriculum is a positive condition, but this alignment brings with it considerable risks - a necessarily incomplete assessment instrument typically drives a curriculum that is reductionist in depth and convergent in range. Napoleon Bonaparte, the bugbear of flexible and relevant curriculum designers everywhere, would have approved.

I hope that we are in no risk of going down the path of high-stakes accountability testing, American style. But to the extent that we use assessment for the generally constructive purpose of system monitoring, we need to be aware of the potential impacts of tests that are too high in stakes, too limited in format, too confined in scope, and too frequent in application.

Validity is not as simple as you might think

Every student teacher knows that validity is one of the big two criteria of good assessment, along with reliability. Every student teacher also knows that to be valid, an assessment has to adequately and accurately reflect the relevant learning objectives. However it is unlikely that every student teacher is aware that the validity of an assessment relates not just to the learning objectives, but to the

purposes for which the assessment and its reports will be used. These include the purposes of the system and the public (Baker and Linn, 2002):

... validity is not a property of a test, but rather a property of the specific uses and interpretations that are made of test scores. Hence, it is not appropriate to make an unqualified statement that an assessment is valid. Rather, the assessment that has a high degree of validity for a particular use may have little or no validity if used in a quite different manner.

This is another way of expressing the risks that arise when someone uses an educational test for an accountability purpose. But it also brings up the issue of league tables. There are plenty of things to object to with league tables. It's always hard to sort out the illogic from the ill will. League tables display a fallacious inference about schools, and invite a fallacious inference about students. Even the BBC, that pillar of journalistic rectitude, publishes league tables with headlines like, 'Hackney's schools worst again'. They seem unaware that league tables are not a news item. They are not a report of mere data, but an interpretation of data, and a tendentious one at that. Why should governments or education systems object to league tables? Because they misrepresent the data we have collected, and invalidate the assessments we have made. And this is invalidity in the fully technical sense, not just a metaphorical or moral sense. In the home of the BBC, only England still publishes league tables. Scotland, Wales and Northern Ireland have either resisted them or abolished them. In Australia, governments are dealing with league tables in various ways, from legislation to negotiation. We want to be transparent, but we don't want transparent misinformation. Here in Queensland we will be entering into a dialogue with the news media, and I hope and believe we

will see responsible public reporting of school results.

'Dumbing down'

For some years now, Queensland has engaged in significant research around curriculum, teaching, assessment and reporting. One major piece of research involved 1,000 observations of selected classrooms. There was good news. Contrary to expectations conjured up by Pink Floyd, the observers didn't witness 'dark sarcasm in the classroom'. The classrooms were found to be strongly socially supportive. But the bad news was a disturbing frequency of low levels of intellectual engagement and intellectual demand. One example of this has had several outings, and you may have heard it before via Allan Luke, one of the lead researchers (Luke, 2001). He described a primary school lesson, a shared book experience on Flipper:

It was about a forty-minute lesson and it was a wonderfully orchestrated lesson. The Year 3 and 4 kids watched a video on Flipper, and they did some enlarged print materials with Flipper. All of these things in a wonderfully socially supportive environment. If I were a principal or teacher supervisor who walked past the lesson I would have said, 'Five out of five! A beautifully run lesson, no behaviour management problems, great teacher.'

At the end of lesson, at the end of forty minutes, what had the kids gained? Flipper. They knew Paul Hogan had starred in Flipper. They knew that Flipper was a dolphin that could talk and that it had been a movie, but that was about it. Now that's what I mean by 'dumbing down' of the curriculum.

So there you have it. Not everyone loves the king of the sea. Keeping it in proportion, we should remember that this was a teacher selected as innovative and effective. I wouldn't want him or her to be forever pilloried for just that one Flipper lesson. The lesson is evidence of success,

in relationship building and commitment, if it is not entirely a success itself. And it is complex in the ways that it is successful and less than successful. And it is far from the worst. We should be grateful to have the lesson on record, for it provides us with a nice example of one type of 'dumbing down'. It is difficult to say what is behind this type of lowered intellectual expectation. We could line up the usual suspects. Is it teacher training, school management, inadequate curriculum or poor communication thereof? Or is it some fuzzy pop educational theory in the ether that gives teachers the impression they have to entertain students, or that they must never compel them to deviate from their own way, that we must never challenge each boy or girl's own journey?

One temptation is to pull this kind of laxity into line using the tool of assessment for one more purpose. In government policy generally we know that 'What gets funded gets done', or 'Move the money, and the behaviour will follow.' That is of course no less true in the education sector. But here there is another attractive lever - assessment - and it has its own parallel platitude, 'What gets assessed is what gets taught'. If standards and curriculum fail, we could try to tackle the Flipper syndrome with tests. There is evidence, sadly enough, that they have a more potent effect.

There is some value in using assessment to bolster standards. We certainly would hope that benchmark tests of the literacies send a message about what we think is fundamental. We would hope that they are valid and precise enough to warn teachers and schools about inadequate practices. In Queensland we will be instituting some common assessment and testing across the state in the early and middle years of schooling, and in part we would hope that those assessments will drive greater consistency and higher standards in curriculum and teaching. But

using assessment to drive curriculum and teaching has its risks. Some obvious ones I have already mentioned. Maybe, 'What gets assessed is what gets taught', but that 'is' had better not be mere identity. We want a lot more to be taught. And it is difficult to drive higher-order cognitive engagement with standardised tests. We can test the floors, but it's hard to test ceilings, especially if you don't want to create any.

The other risk if you use assessment to drive curriculum and teaching is the message it sends. Some people will get the idea that that's the right way around. But if assessment drives curriculum, that's as a matter of behaviour, not philosophy. Curriculum should come first, or curriculum and assessment should both fit into a framework. Otherwise, the assessment tail may wag the curriculum dog. To see assessment, especially reportable assessment, as defining educational outcomes is:

- a logical mistake (since assessment cannot be comprehensive);
- a philosophical error (since it inverts educational values); and
- a practical risk (since minimum or partial outcomes tend to become maximum or total).

So our teachers and educational leaders need to know that that is not the logic. We need to clearly communicate the conceptual context that assessment fits into. As I've said more than once before, we have to ensure that curriculum, teaching, assessment and reporting are all aligned. And they need to align vertically as well as horizontally, they need to align in terms of intellectual challenge, as well as content covered.

We should recognise that dumbing down in schools can occur in the name of either amorphous personal growth or reductionist test-fixations. We need to resist the challenges to intellectual rigour

that come from either. One will take us down the road of the boy's or girl's own journey, content-free play-learning, ignoring evidence and resisting standards. The other will take us down the road of fact-grubbing, teaching to the test, simplistic regression to past nostrums, and branding league-table pariahs. There is a workable, consistent position between the extremes, and it is not an obscure or abstract or intellectually difficult position.

Aligning curriculum and assessment

We are back to getting the balance right. On the one hand, then, we have the importance of aligning curriculum, teaching, assessment and reporting. On the other hand we have the risks that follow from aligning curriculum and assessment too tightly, or from aligning curriculum to assessment, rather than vice versa. On the one hand we have the usefulness of assessment for reinforcing standards, communicating expectations, motivating improvement, monitoring the system. And on the other hand we have the risks of perverse consequences, like narrower curriculum, and of misinterpretation, like league tables. On the one hand we have the educationally essential role of formal and informal classroom assessment that is continual and integral with learning. And on the other hand we have the challenge of supporting teachers to do it well. We need to mind the gap between curriculum and assessment. And sometimes we must close it, and sometimes we must preserve it.

What we have found and are doing in Queensland

In Queensland we have accepted for some years that it's a key axiom of educational theory that curriculum, teaching, assessment and reporting need to be brought into proper alignment for us to get the educational results and outcomes we want. I mentioned the sustained

research effort that we have conducted for five years or more now. Among many other things, it has revealed that three areas of intervention are separately necessary and jointly sufficient for making a difference: social support, basic skills, and intellectual engagement. All three are required:

- Classrooms may be socially supportive but intellectually undemanding, as the research observations had disclosed.
- Basic skills in literacy, numeracy and facts/information are necessary – they provide a 'floor' or 'threshold', but they do not of themselves guarantee sustained achievement gains of at-risk learners.
- Higher-order, critical, intellectually demanding work and connection to the world are critical.

When it came to making these curriculum features happen, the research found that curriculum alone wouldn't do it. Changing curriculum doesn't work if you don't change pedagogy, for pedagogy is the way that curriculum happens in classrooms. Changing curriculum doesn't work if assessment and reporting pulls in a different direction. As I've mentioned, the system that we use in the Queensland senior school, which combines moderated internal assessment with a benchmarking test of core skills, we believe works very well and has earned a great measure of community confidence. Early and middle schooling, on the other hand, while working from a common syllabus framework, lack a consistent moderation system. It is more difficult for schools to ensure that their pedagogical and assessment practices are consistent with each other or comparable with a state-wide norm. And that came out in the research. It showed non-alignment in the educational factors of curriculum, teaching, assessment and reporting. The Queensland School Reform Longitudinal Study looked at curriculum and

assessment, and demonstrated 'a marked inconsistency between the quality and nature of pedagogy and those of teachers' assessment practices.' There was a generally low level of demand for intellectual quality in assessment tasks, especially in the more informal setting of the early and middle school. However, since then we have found that adopting a rigorous assessment system is workable. It challenges teachers, but the important thing is to provide them with appropriate support and a feedback mechanism. Under those conditions, they demonstrate the skills and commitment to meet the demands of good assessment, state-wide comparability and school accountability. That rigorous assessment system brings a cluster of benefits. It establishes standards for student work. It supports effective implementation of the curriculum. It enables comparative reporting on the quality of student work.

We are now at work bringing curriculum, teaching, assessment and reporting into systematic alignment in primary and junior schooling. In April, the Premier and the Education Minister announced the development of the new Queensland Curriculum, Assessment and Reporting (QCAR) framework. Despite appearances, it is not a framework for unmarked police vehicles. The QCAR framework will:

- define what is essential curriculum for all students in Years P-10;
- assist schools to continue to provide other curriculum appropriate to local needs;
- set standards of student achievement in the essential curriculum;
- create a bank of assessment tools for teachers that link to the essential curriculum and standards;
- establish, at three key points in the P-10 years, rigorous comparable assessment against the defined standards, which will result in consistent teacher judgments of student achievement;

- specify a common framework for reporting student achievement against standards; and
- provide for ongoing review and refinement of Queensland's P-10 syllabuses based on a rigorous evidence-based approach.

Once we've established state-wide standards, we will be in a better position to monitor the achievement and progress of individual students. We will be able to report to parents at key points in their children's schooling in a way that allows a broader view of their child's progress. And teachers, students, parents and the community will have a clear, common base for talking to each other. The framework is designed to ensure that standards and assessment have a positive effect on student learning. Rigorous comparable assessment against defined standards at key points will allow students to demonstrate a broad range of learning and deep understanding. The new assessment will also support school and system-level planning by providing more comprehensive and comparable data on school performance. All reports for state school students will include some common elements such as descriptions of what has been taught, the expected standards of student performance, and how well the child has performed against these standards and their peers. For parents, the framework will deliver easy-to-read reports that show how well their children are performing compared with others, and what is expected at their year level.

Nevertheless, true to the Queensland spirit and the Queensland reality of a vast, diverse State, the QCAR framework is just that – a framework, not a Napoleonic code. Like the reforms that have come before it – the Radford review of 1970, and the Scott review (ROSBA) in 1978 – the QCAR framework will support curriculum that has local applicability and

general portability. The framework will provide integration, but integration that can be applied flexibly at the school level, and accommodate local initiatives and community needs.

Conclusion

Perhaps this speech has been a cavalcade of ifs and buts, a litany of serial qualifications. I don't really seek to escape that charge. Policy is an art, and education is an art, and educational policy must be an art. Like all arts, It is about balance and nuance. It is the art of steering your vessel between the anarchic whirlpool of Charybdis and the sniping heads of Scylla.

At last I will define my topic. It seems to me my themes have been the role of assessment, and its relationship to curriculum, and the role of government with respect to both. What is the role of government in these things? To refrain from making this an arena for ideology and partisan manoeuvring. To maintain accountability, but not through mechanisms that corrupt educational process or defy normal reason. To encourage a clear view of the interrelationship and priority amongst curriculum, teaching, assessment and reporting. And what is the role of assessment? At its most fundamental level, assessment is about seeing our students - seeing what they are, and seeing what they might be. And communicating those things to them and their families in a way that informs but does not inhibit, inspires but does not inflate. It seems almost impossible to get the right balance. But it is everyone's task - be they educators, political leaders, parents or bureaucrats - to work to get the balance right. For out of the right balance will come improved educational outcomes for everyone, not just the privileged, in our community.

References

- Baker, Eva L. and Linn, Robert L. (2002), *Validity Issues for Accountability Systems. CSE Technical Report*, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Behar-Horenstein, Linda S., Gail S Mitchell and Teresa A. Dolan 2004, 'A multidimensional model for analyzing educational processes', *International Journal of Leadership in Education*, Volume 7, Number 2 / April-June 2004, pp. 165-180.
- Bloom, Benjamin S., J Thomas Hastings and George F. Madaus (1971), *Handbook on formative and summative evaluation of student learning*, McGraw Hill, New York.
- Bond, Linda Ann (1995), 'Rethinking Assessment and Its Role in Supporting Educational Reform' North Central Regional Educational Laboratory, accessed at <http://www.ncrel.org/sdrs/areas/issues/methods/assment/as700.htm>.
- Campbell, W. J. and Robinson, N. M. (1979), *What Australian society expects of its schools, teachers and teaching*, Department of Education of the University of Queensland.
- Commission on Instructionally Supportive Assessment (2001), *Building Tests To Support Instruction and Accountability*, A Guide for Policymakers.
- Dietel, R. J., J.L. Herman and R.A. Knuth (1991), 'What Does Research Say About Assessment?' North Central Regional Educational Laboratory, accessed at http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm
- De Bortoli, Lisa and Cresswell, John (2004), *Australia's Indigenous Students in PISA 2000: Results from an International Study*, ACER Research Monograph No 59.
- Department of Education, Queensland, *Destination 2010*.
- Department of Education, Queensland (2002a), *Literate Futures, Report of the Literacy Review for Queensland State Schools*.
- Department of Education, Queensland (2002b), *Report of the Assessment and Reporting Taskforce*
- Department of Education, Queensland (2004a), *Schools Reporting Consultation Paper*.
- Department of Education, Queensland (2004b), *Changes to Schools Reporting*.

- Department of Education, Queensland (2005), *Smarter Learning: The Queensland Curriculum, Assessment and Reporting Framework*.
- Gleeson, Denis and Husbands, Chris (2003), 'Modernizing schooling through performance management, a critical appraisal', *Journal of Education Policy*, 18 (5), pp. 499-511.
- Gunzenhauser, Michael G (2003), 'High-Stakes Testing and the Default Philosophy of Education', *Theory Into Practice*, January 1 2003 Vol. 42 Issue 1.
- Hamilton, Laura S., Brian M Stecher, and Stephen P Klein (eds.) (2002), *Making Sense of Test-Based Accountability in Education*, RAND.
- Hamilton, Laura (2005), 'Assessment as a Policy Tool', Chapter 2 in *Review of Research in Education*, 27.
- Hannaway, Jane, and Nicola Woodroffe (2005), 'Policy Instruments in Education', Chapter 1 in *Review of Research in Education*, 27.
- Hargreaves, David (2005), 'Self-discipline will test the profession', *The Australian*, 11 April 2005.
- Hopmann, Stefan Thomas (2003), 'On the evaluation of curriculum reforms', *Journal of Curriculum Studies*, Jul 2003 Vol 35 Issue 4, p. 459.
- International Reading Association (1999), *High-Stakes Assessments in Reading, A Position Statement*.
- Koretz, Daniel M. (2002), 'Limitations in the Use of Achievement Tests as Measures of Educators' Productivity', *Journal of Human Resources*, September 1 2002 Vol. 37 Issue 4.
- Koretz, Daniel M et al (1992), 'National Educational Standards and Testing, A Response to the Recommendations of the National Council on Education Standards and Testing. Congressional Testimony', Center for the Study of Testing, Evaluation, and Educational Policy Rand Corp Santa Monica CA Inst. on Education and Training.
- Koretz, Daniel M., McCaffrey, Daniel F. and Hamilton, Laura S. (2001), *Toward a Framework for Validating Gains under High-Stakes Conditions*, CSE Technical Report, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Kulieke, M., J. Bakker, C. Collins, T. Fennimore, C. Fine, J. Herman, B. F. Jones, L. Raack, and M. B. Tinzmann (1990), 'Why Should Assessment Be Based on a Vision of Learning?' North Central Regional Educational Laboratory.
- Linn, Robert L. (1994), *Assessment-Based Reform: Challenges to Educational Measurement*, Educational Testing Service, Princeton, NJ.
- Linn, Robert L. (2001a), *Assessments and Accountability: CSE Technical Report*, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Linn, Robert L. (2001b), *The Design and Evaluation of Educational Assessment and Accountability Systems*. CSE Technical Report, Center for the Study of Evaluation; National Center for Research on Evaluation, Standards, and Student Testing.
- Linn, Robert L. (2003), 'Accountability, Responsibility and Reasonable Expectations', *Educational Researcher*, v32 n7 p. 3-13, Oct.
- Luke, Allan (1999), 'Education 2010 and new times, Why equity and social justice still matter, but differently', paper presented to the Education Queensland Online Conference, 20 October 1999, accessed at <http://education.qld.gov.au/corporate/newbasics/docs/onlineal.doc>.
- Luke, Allan (2001), 'How to Make Literacy Policy Differently, Generational Change, Professionalisation, and Literate Futures', paper presented to the ALEA/AATE (Australian Literacy Educators Association / Australian Association for the Teaching of English) Conference Hobart 2001, accessed at <http://education.qld.gov.au/curriculum/learning/literate-futures/pdfs/alan-luke.pdf>.
- Macdonald, Doune (2003), 'Curriculum change and the post-modern world, is the school curriculum-reform movement an anachronism?', *Journal of Curriculum Studies*, Mar 2003 Vol 35 Issue 2, p. 139.
- Perry, William G. Jr. (1963), 'Examsmanship and the Liberal Arts, A Study in Educational Epistemology' in *The Norton Reader, an anthology of expository prose*, any edition, and accessible online at <http://www.people.fas.harvard.edu/~lipoff/miscellaneous/exams.html>.
- Rohl, Mary 2005, 'The Practices of Effective Teachers of Reading/Literacy, A Summary of the Research Literature', discussion paper for the Committee and Reference Group for the National Inquiry into the Teaching of Literacy, February.
- Rowntree, Derek (1987), *Assessing Students: How shall we know them?* Kogan Page, London.

2005 Curriculum Corporation Conference

Shepard, Lorrie A. (1989), 'Why We Need Better Assessments', *Educational Leadership*, v46 n7, April, p. 4-9.

Schindehette, Susan; Rozsa, Lori; Harmel, Kristin; Frey, Jennifer; Russell, Inez; and Bresnahan, Angela (2004), 'Stressed by the Tests', *People*, 01/11/2004, Vol. 62 Issue 18, p. 82.

Smith, Frank (1995), 'Let's declare education a disaster and get on with our lives', *Phi Delta Kappan*, April 95 Vol. 76 Issue 8.

Sternberg, Jason (2005), 'Demonising youth of today tonight', *The Age*, 13 April 2005, accessed at <http://www.theage.com.au/text/articles/2005/04/12/1113251624108.html>.

Thompson, Michael C. (2000), *Curriculum as Profound Engagement with the World*, Keynote Speech to the National Curriculum Networking Conference, Center for Gifted Education, The College of William and Mary, March 7, accessed at <http://cfge.wm.edu/documents/CurriculumProfound.html>.

Winch, Christopher, and Gingell, John (2004), *Philosophy and Educational Policy: A Critical Introduction*, RoutledgeFalmer.