

What is the nature of evidence that makes a difference to learning?

John Hattie^a

^a*The University of Melbourne (Melbourne Graduate School of Education),
jhattie@unimelb.edu.au*

Abstract

In relazione al sistema scolastico abbiamo la disponibilità di una gran mole di dati e la tendenza all'accountability esige che se ne raccolgano sempre di più. Il seguente lavoro assegna agli insegnanti un ruolo cruciale nel ciclo dell'evidenza, presentando un modello che li può aiutare a percepire e utilizzare l'evidenza per trarne vantaggio nei processi di apprendimento. Questo modello permette agli altri diretti interessati (dirigenti scolastici, responsabili a livello ministeriale, genitori, allievi) di condividere tale evidenza. Il lavoro descrive a grandi linee gli esiti nelle scuole che lo hanno sperimentato per poi sviluppare un modello di rendicontazione dell'intero sistema basato sull'evidenza che fa la differenza nell'insegnamento come nell'apprendimento.

Parole chiave: evidenza, accountability, obiettivi di apprendimento, incremento del rendimento scolastico, efficacia.

Abstract

Schools are awash with data, and the accountability movement is requesting that they collect even more. This presentation locates the teachers as critical in the 'evidence' cycle. It demonstrates a model for assisting teachers to ascertain the nature and use evidence to make a difference to learning. This model permits other key stakeholders (principals, Ministries, parents, students) to then share this evidence. It outlines studies in schools that have been using the model and then develops a system-wide accountability model based on this evidence that makes the difference to teaching and learning.

Keywords: evidence, accountability, learning outcomes, achievement progression, efficacy.

Schools are awash with data, and I have yet to find a Department or Ministry of Education which does not have so much data that debate is more concerned with issues such as data warehouses, executive information systems, web pages, data portals, and the use of Access, Oracle, or other mega-data systems. Soon after this bounty is collected, someone begins to ask “How can we return it to the schools?” At last year’s Round Table on Assessment in Sydney, for example, there were many discussions about the volumes of data that can be readily returned, and how it could be ‘massaged’ and presented to schools in the most digestible form. It was also noted, in passing it seemed, that the schools were not that enamoured with receiving so much data – they were not sure what to do with it, and were concerned by the time and workload involved in reading and digesting it. Hence, there is the desire to find more acceptable ways to return ‘their’ data back to the schools. It seems, once again, there is an effort to solve the problem in front of us rather than the problem that should be in front of us. Asking whether and how to send data back to schools is the wrong question. A major theme of this presentation is that we must be more mindful of the ‘interpretations’ we wish to make from any data collected as it is the ‘interpretations’ that are critical, rather than data itself. Of course, the quality of the data reflects on the validity of the interpretations, but it is the latter, which should be uppermost in our minds when we (a) collect data, and (b) return interpretations to those we wish to influence.

In the meantime, while volumes of data are extruded about and from schools, teaching continues without the benefits of such data. There is still a philosophy that assumes teachers know how and what data to collect to best enhance learning, and many of these assumptions are based on folk philosophies, poor measurement, and shaky data. We still teach in a manner we did 150 years ago (see Cuban & Tyack, 1995), with a preponderance of talking (about 70–80% of the time, see Yair, 2000), deciding on activities that aim to engage rather than choosing activities that reflect on curricula intentions that aim to challenge. We are loosing the minds and hearts of the students (particularly during early adolescence, when disengagement is already a ‘cool’ attribute) and we are also losing the voters as their belief about the quality of schooling declines.

Because of such criticism (and also because it seems good practice), it is not uncommon for systems then to invent ‘accountability’ systems to drive the teachers to get more and more learning out of their charges. One form of accountability assumes that if only we could name, shame, and blame evidence, we could get those teachers operating at higher levels of efficiency. Another form of accountability assumes that if only we could collect sufficient system-wide evidence, we could convince the parents/voters not to be critics. Both miss the mark.

Most depend on the thermometer theory of traction – although the abject failures of this model are already causing untoward damage to our profession of teaching, to the role of principals, and leading to students’ lack of engagement. Perhaps the most visible form of accountability that illustrates these issues is the US ‘No Child Left Behind’. While I see some merit in some of its claims (e.g., ensuring all students, and not just the ‘average’ student, succeeds within a school) the implications of this USA-wide accountability system have become most clear in its negative effects. It has made the teachers teach what they expect is coming in the test; it ensures students are focused on this teaching; it judges the success of the school in terms of whether teachers are doing this job of teaching to the test; it rids the school day of ‘peripherals’ that are not tested (such as physical education, music, art, and self-respect); it cuts vocational and career education programs in high schools which are desperately needed by many students whose

alternative is to drop out, and it punishes those who do not do their job and teach to the test (see Hattie, Brown, & Keegan, in press; Linn, 2003; Shepard, 2000). It ensures that (a) there is a quick gain as all learn how to 'game' the test, (b) that the curriculum is altered downwards to ensure that there is reasonable success for more students, (c) it introduces procedures to remove those who may bring scores downward (e.g., 'accommodating' special education students out of the test room, retain back students from moving up a grade, using suspensions and not enrolling students who may detract from the scores of high schools, and so on). It is not the model worth moving towards.

As Robert Linn (2003) has demonstrated, it would take an innovation of atomic bomb proportions to get the average yearly gains needed to reach the stated goals by 2012 – the target year (he estimated that it will take 150 years at the rate of annual yearly progress of the past 10 years to reach the targets set for 2012). As Australia moves towards national testing, it will become more awash with data, it will de-contextualise schools, lead to more claims for 'school choice', increase the flight out of the public schools, will lead to more schools in lower socioeconomic areas stumbling and more schools in higher socioeconomic areas cruising, and, most of all, it will feed the belief that the quality of schooling in the State/Territory/Australia is declining. I see none of this enhancing the quality of teaching and learning.

But we need to remind ourselves who is asking for more tests – it is incorrect to blame the politicians. They are clearly listening to the voters – who want more accountability (which they interpret as tests and data) in the same way politicians wish to return evidence that their investment in schooling is paying off. Let me make two claims here.

First, schools have failed in their efforts to provide appropriate and defensible data to parents about their children – hence the clamour for more tests. We (Hattie & Peddie, 2003) published a study based on school reports to parents from 156 schools in New Zealand. Only 12 included information relating to the official curriculum levels; half included no information on achievement relative to any standard; half talked about students in agricultural terms (developing, needing more, emerging, growing); and half included a specific section relating to effort. On the basis of these reports 98% of students had positive comments about their achievement, were putting in effort, and were 'a pleasure to teach/joy to have in my class'. With few exceptions, the majority of students in these schools were achieving above average! No wonder parents demand more 'tests', accountability, and 'teacherproof' information from our schools.

Second, there is not a lot of evidence that the massive increases in state/federal monies have made a difference to the quality of teaching and learning. Hanushek (2005) has presented information (in current dollars allowing for inflation) of changes in public schools' resources in the United States over the past 40 years (Figure 1). The achievement curve (from NAEP) has remained constant over this period. If we, as educationalists in classrooms and schools do not provide the evidence that increased resources make a difference to student learning and outcomes, then we will soon be on the back foot, arguing why there should not be decreases in resources.

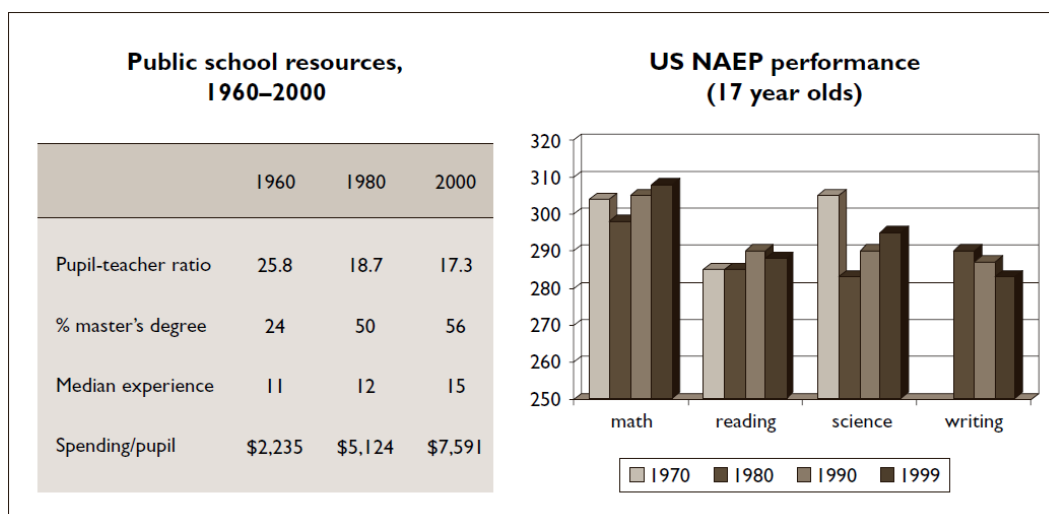


Figura 1. Indicators of changing public school resources and NAEP achievement over the past 40 years in the USA

My major theme is that we need models of school/teacher/student accountability located at the **system and school level** that maximises the probability of enhancing learning and outcomes. Indeed, we must develop an accountability system that is located from the student level upwards, directly involving and influencing the **teacher and principal level**, as such a system is more likely to have major effects on the quality of teaching and learning. Such a system, which I intend to outline, can also serve the systems' needs of providing evidence of curricula, resources, and equity issues.

What makes a difference to teaching and learning?

The reason for locating the power of data to enhance student outcomes at the teacher level comes from the many recent studies on the epicentre of casual effects on learning: the teachers. At this same ACER conference, two years ago I presented on the factors that make a difference to teaching and learning and divided them into six parts of the cake (Hattie, 2003): This is a summary of what is, not what should be – as I certainly can note the power of peers as co-learners, the role of principals to make a difference to instructional leadership, and so on. It is clear, that the major factor in this equation is the student – but most of you have to take what the neighbourhood produces and discussions of 'choice' too often means that schools get to choose the students they want (and many students in certain neighbourhoods are denied the choice they want). Maybe there is merit in 'choice' but most of us get what comes through the school gates from the local areas. Similarly absurd notions of brain waves, learning styles, multiple intelligences and other pop-educ claims are more befitting brain surgeons than the cut and thrust of the teaching and learning conundrums. The major influence on student learning is the teacher, and here is where I wish to locate the issue of 'What data would support a teacher to enhance teaching and learning?' and thus how can we devise systems to ensure that such data is obtained, and when obtained that it makes a difference? While there are other sources of data useful to a system, the key to any accountability model should orient around this question.

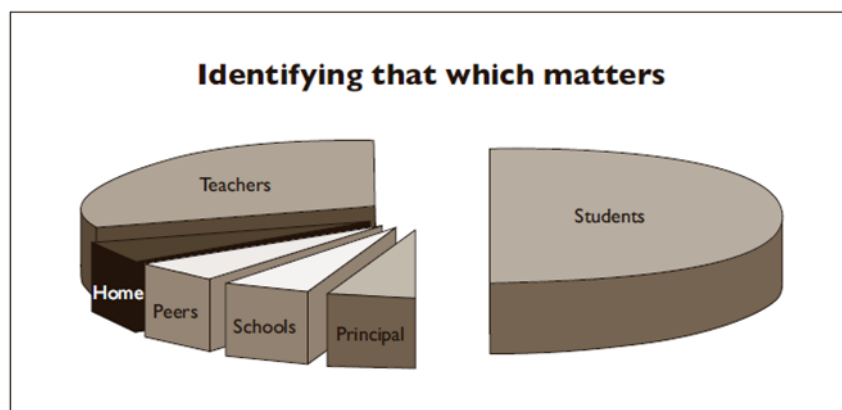


Figura 2. Percentage of achievement variance

What are 'learning outcomes'?

This question begs the question: What is it that we wish to enhance? This question has occupied the minds of curricula reformers for decades, and we seem to experience a once-a-decade bump where the old curricula is repackaged, new names invented, much is added and little is subtracted, and the classrooms continue on much as before. The latest craze, begun by the OECD is to include key competencies or 'essence' statements and this seems, at long last, to get closer to the core of what students need. Key competencies include thinking, making meaning, managing self, relation to others, and participating and contributing. Indeed, such powerful discussions must ensue around the nature of what are 'student outcomes' as this should inform what kinds of data need to be collected to thence enhance teaching and learning.

Outcomes from curricula must have a sense of achievement progression. From our New Zealand research, it is most defensible to claim that a common understanding among teachers of progression is probably the greatest chokepoint to the enhancement of learning outcomes for students. While there can be sharing of activities and stories about students and incidents, it is rare to hear discussions among teachers about the levels of understanding, the degree of challenge and expectations required and attained – such that each year teachers revisit the students in terms of their internal beliefs about what levels of performance are required – allowing students to gain or drop according to these (often untested) beliefs about the desired levels of progression (Robinson & Lai, 2005; Timperley, 2005). One of the major purposes of an accountability system is to assist in articulating a common language of progression.

The nature of 'data'

Before venturing into the recommended model, it is important to comment on the nature of 'data', as this is a most contested term. A current fad radiating out from the United States is the notion of evidence-based decision making – and this term has been hijacked to mean a very narrow form of evidence. Liberty and Miller (2003), for example, consider 'evidence-based' relates to meeting peer-review standards, and including evidence directly impacting on children's learning (not correlates, see Scriven, 1988). This cuts out so much of today's literature and I note an excellent summary of the surviving literature

by Alton-Lee (2003). But an extra condition has been added, that of the type of research designed to collect data: preferably random assignment to various groups (Mosteller & Boruch, 2002). While this may be exemplary, it is not the only design of merit. Moreover, in classrooms, teachers still need to base their evidence on data from their students and from their teaching, and rarely does random assignment occur. It is this form of teacher-available data that is of interest to my forms of accountability.

Such classroom-based data is also contested – and while it can consist of scores on tests, it can also consist of teacher judgements, student ratings, and so on – provided such evidence can be defensibly accumulated and is open to scrutiny. It is the judgements or interpretations based on these data that is of most interest. The asTTle model outlined in the presentation allows such evidence to be defensibly accumulated and contested – and this is how it should be. We must contest the evidence – as that is the basis of a common understanding of progression.

The location of ‘evidence’ starts in the classroom

The argument in this presentation is that the location of evidence that makes a difference to teaching and learning must be located at the ‘teacher’ level. Of course, the students are implicitly involved – but they are not the core. This is because it is most common to locate students in groups (i.e., classrooms) critically influenced by the teacher. Indeed, my theme is that if we form the accountability model around providing teachers with excellent diagnostic and formative evidence, we have not only an excellent model but one that influences teaching and learning. Basing a model on students can help those students who learn in a diagnostic and formative manner about such accountability evidence but this would exclude most students. Similarly, basing it on parent’s privileges (those who have the home-resources to add value to this evidence) would again exclude so many parents, particularly those who do not have command of the language of schooling and learning (Clinton & Hattie, 2005).

The first part of the model is to address teachers’ expectations and target setting, as these are key drivers in the enhancement of learning – or can be the greatest barrier to such enhancement (Rubie, Hattie, & Hamilton, in press). These expectations also are underpinned by the teacher’s conception of progression. During the conference presentation I will demonstrate a target setting process for individual students that allows immediate aggregation to the class and school level to ask whether the target setting is reasonable, enhancing, and defensible. The critical features include the following: it is in the language of teaching and learning and not assessment; it leads to discussion among fellow teachers about the nature of teaching and learning; and it provides school leaders with information to form a school-wide discussion about targets. Similarly, I will demonstrate a school profile also provided by the asTTle package that shows current performance and how it can be used to evaluate the degree of attaining these targets. Similarly, for both sets of evidence the national norms (for the country, or for ‘schools like mine’ can be interpolated).

The emphasis is on growth, and avoids many of the current problems with value-added models. The latter have been too dependent on measuring only at two time points, with all the incumbent problems (Cronbach & Furby, 1970). The current model, however, incorporates many time points and is thus conducive to an interrupted time series analysis – which has much more power to provide information on the value added by teachers and

schools (see Hattie & Rowe, 2004). At least it moves the discussion beyond the status of the students, which is what must accrue from state-/nation-wide models and to include the critical questions relating to *growth*.

The asTTle model is based around three major questions: Where are we going? How are we going? and Where to next? Thus, target setting is critical, as is evidence of the gap between current and targeted performance, and the manner in which teachers are going to reduce this gap for all students. Other graphs from the asTTle application will be shown that will demonstrate how a national system can provide evidence on these issues, in an immediate way, to teachers and students. There is evidence of individual student achievement, class achievement, the distribution of achievement across cohorts, schoolwide analyses, and linkages to appropriately challenging curricula materials. These analyses can be conducted at the individual as well as at the cohort, class, and school levels.

Evidence-based curricula development

Curriculum is also a contested domain, and too often, it is resolved by asking a group of experts to devise a new version – often tinkering at the edges, choosing new names to dominate the centre, and the teachers do much the same as they did before. Instead, it is argued, curriculum development should start with evidence based on what students know and can do.

Take mathematics as an example. It is easy to imagine a group of ‘experts’ arguing for some new twist or development in mathematics. The current vogue seems to be number strategies, and in New Zealand a group has decided there are six of these strategies, they are hierarchical, and that it is desirable that students, as early as possible, learn to strategize using the highest step in the hierarchy. My point is not to question the merit of this claim (although see Ell, 2002; van Gardaren, 2002) but to highlight that number operations are considered in most need of curriculum innovation.

We have accumulated evidence based on about 25,000 students undertaking over 1500 items from across the mathematics curricula (from the asTTle norming sample). Then we can present the growth of number (in its three forms) and can see a steep learning curve right throughout the Years 5 to 12. But in Geometric Knowledge we can see a shaky start in primary school; there is a decline and then no growth during Years 5 to 7; and then over the latter years of schooling, a less steep growth than for Number. There should be a major set of questions here about the teaching of geometry in primary schools – perhaps dropping it completely! We can drill down deep below this level of aggregation and also ask about specific objectives within Number and within Geometry, and this is the nature of evidence-based curriculum development. Such discussion, based on evidence about learning can contest deeply held beliefs about what should be undertaken in the name of curriculum form, and can lead to asking direct questions about where the curriculum needs to be reformed, and where to be left alone.

Evidence based within school development

There are many within-school debates about the nature of evidence that makes a difference to learning? Let me illustrate six.

1. The importance of asking relative questions of effectiveness

If you could sum up all the studies on what makes a difference to students' achievement, there are very few that do not report some success. Nearly everything enhances achievement, thus any teacher claiming that they can show evidence of enhanced learning is not saying much. For example, based on my syntheses of evidence on this question (Hattie, 1999; in prep), I have determined the effect-sizes of over 100 major innovations from over 300,000 studies. For example: zero is when there is no effect on achievement, a negative effect is when the innovation reduces achievement, and a positive is when the innovation enhances achievement. These innovations include structural changes (reducing class size, ability grouping), curricula innovations, teacher effects (questioning, direct instruction, reciprocal teaching), and so on. Virtually everything we do enhances achievement (note how few are below the zero effect-size). The critical question is whether we can implement those effects that enhance achievement by more than the average (.40 effect-sizes). Anything less is holding back a student, as at least half the effects can attain growth greater than .40.

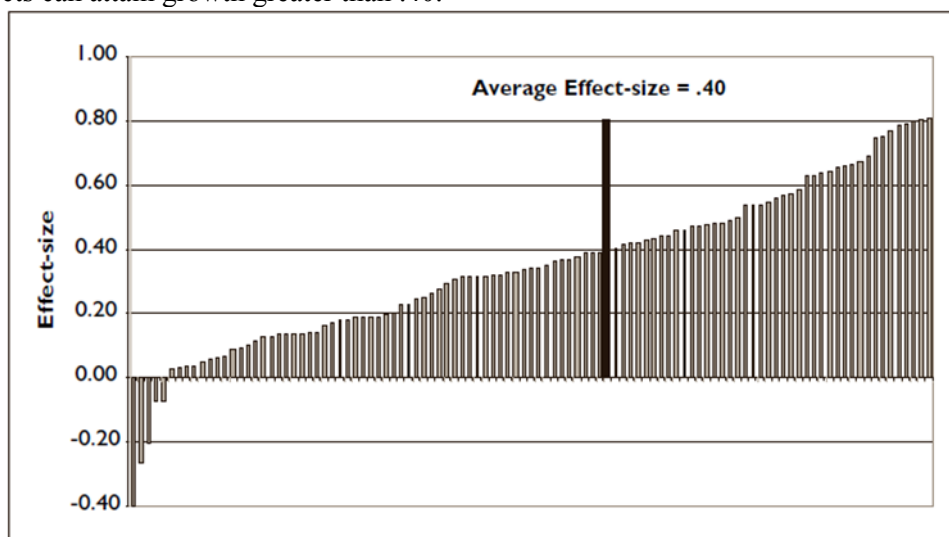


Figure 3

2. The use of effect-sizes in classrooms to underpin the discussion on effectiveness

The power of effect-sizes (the difference between two groups or between two time points divided by their pooled standard deviation) is relatively easy to implement in schools. Phillips, McNaughton, and MacDonald (2001) have used effect-sizes in their implementation of school-wide literacy programs in schools from lower socioeconomic areas, with much success. Their success is not only to provide policy makers with evidence of the success, but more importantly to assist teachers in the delivery of the literacy program. Another advantage of using effect-sizes is that they force schools to have clear goals and standards of student performance, as only then can teachers collect and review information to inform themselves about their levels of success with their students in reaching those standards (Newmann, King, & Rigdon, 1997; Timperley, 2004). Turning such evidence into tools for teachers is the key to evidence-based teaching. Timperley (2005), for example, worked with teachers in one low socioeconomic area, and began by collecting a range of literacy achievement information: This information, on individual student profiles and held in each teacher's filing cabinet, was

vast and encompassing. The more formal assessments (using standardised measures like the Reading Observation Survey, Clay, 1993) was considered by the teachers as something collected for the assistant principal's use, not theirs. Teachers considered that the most relevant planning evidence was anecdotal observational data collected on a daily basis in their classrooms. They considered such data was relevant and trustworthy in contrast to the more formally collected information. The assistant principal, however, was concerned about the low quality of these anecdotal observation data particularly because they did not give the teachers an understanding of the adequacy of their students' progress in comparison with other students in the country. When explaining the national data for their students, teachers had many reasons to exclude the information (the national kids are not like mine, I teach to the best of my ability given whom I am given, I should not 'teach to the test', the tasks are not 'authentic', others fail to understand what my kids can do, I have too many students in my class, I need more time if this is going to impact on me, and so on. Timperley (2005) Timperley highlighted the power of 'surprise' to ensure more ownership by the teachers: 'One of the ways in which data can be powerful in creating change is the possibility that they may be discrepant with previous beliefs and create surprise, thus challenging those beliefs' (Schutzwahl, 1998; Timperley & Robinson, 2001). When teachers compared their students' growth with that of students of other teachers, they were surprised. The most important aspect of this study was moving the teachers from expressing outcomes in terms of the students they received, the working conditions of teaching and learning, to a set of contingencies based on learning outcomes.

3. The importance of learning intentions and success criteria.

Using effect-sizes, or any evidence of enhancement comes back to the issues of merit and worth of the outcomes. Within the classroom, we have articulated these as learning intentions and success criteria. Our work in schools too often shows that students rarely know the learning criteria for a particular lesson, are confused as to what success would look like for this intention (often claiming that something long, spelled correctly, and neat is indicative of the success criteria), and do not see how the assessment relates to the success criteria nor the learning intentions. We have spent much time writing about making learning intentions and success criteria explicit, and have seen many classes and schools transform with these simple but powerful ideas (Clarke, Timperley, & Hattie, 2003). To illustrate:

- **Learning intention:** 'To understand the causes and effects of events that have shaped the lives of a group of people.' The context might be the diseases that affected Maori after the arrival of the British colonists.
- **Success criteria:** By week 3 of this unit, students will be able describe the trends in Maori population between 1820 and 1920. By the end of the unit, the students will be able to explain the effect of British colonisation on Maori health at the beginning of the twentieth century and how it influenced Maori population trends and make predictions about the health effects on indigenous peoples by colonising countries.
- **How success criteria will be assessed:** Students will be able to write a paragraph that relates three pieces of information: the arrival of British diseases, the population trends of Maori, the contribution of previously unknown disease to the decline in population.

Evidence is now easy – it relates the teacher's intention (from the curriculum) to the task and activities, clearly specifies the criteria the teacher would use to judge student

learning, and indicates how data could be collected specific to these criteria. And even more powerful if the learning intention, success criteria, and assessment are shared with the students (as they commence the task). At a minimum, it stipulates the notion of what the learning outcomes are, can lead to debates about sufficiency, challenge, appropriateness, time, resources, and can indicate to other teachers and students (and parents) the level and depth of the learning.

4. Assessment data is optimised when teachers conceive such data as about them (and not about the students)

One of the powerful ideas in evidence based models of teaching and learning is that teachers need to move away from considering achievement data as saying something about the student, and start considering achievement data as saying something about their teaching. If students do not know something, or cannot process the information, this should be cues for teacher action, particularly teaching in a different way (the first time did not work!). Merely ascribing to the student the information that they can or can not do something is not as powerful as ascribing to the teacher what they have or have not taught well. A similar powerful idea is that teachers have differing conceptions of assessment (Brown, 2004), and understanding these differing conceptions may be critical before encouraging teachers to collect more evidence. Brown (2004) has discovered four major conceptions: assessment improves teaching and learning, assessment makes schools and teachers accountable, assessment makes students accountable, and assessment is irrelevant. If teachers consider assessment is irrelevant, then this needs to be attended to before inviting such teachers to consider evidence-based models of teaching and learning. They will depend overly on anecdotal evidence, believing that completion of assigned tasks (regardless of difficulty and challenge) and similar such engagement-related activities are more critical than any 'surprises' and evidence based on dependable testing procedures. It may be necessary for teachers to listen to students more closely and thus use other sources of classroom evidence. Bishop, Berryman, Tiakiwai, and Richardson (2003) interviewed Māori students about how to best improve their educational achievement. The students claimed that the major changes needed to be how teachers related and interacted with Māori students in their classrooms. Too often, these interactions were based on deficit theorising by teachers about these students, and too often these relationships were based on denying that the students had a rich cultural heritage that they brought to the classroom. This led to low expectations of Māori students and collecting evidence to confirm these beliefs, thus creating a downward spiralling, self fulfilling prophecy of Māori student achievement and failure. Based on these student experiences, the Team developed a professional development intervention, that when implemented with a group of 11 teachers in four schools, was associated with improved learning, behaviour and attendance outcomes for Māori students. Similarly, Irving (2005) has found that students are very adept at identifying excellence in teaching and the major question may be 'Why primary and secondary teachers do not use more student evaluation of teaching?' Irving used the standards of the National Board for Professional Teaching Standards to create a student evaluation instrument (for high school mathematics). Using a sample of NBC and non-NBC teachers, he found that students could dependably discriminate between these two groups of teachers. The data are there but is the courage to use it there?

5. Movement towards student empowerment of teaching and learning

If you believe in student self-assessment, self-monitoring, self-teaching, self-learning,

and self-responsibility – then it is critical that the student has dependable evidence on which to base their decision-making. Instead, we so often promote the power of self-regulation but fail to realise that it is premised on evidence of learning performance.

6. Enhancing teacher performance to improve student learning is conditional upon evidence

Timperley (2005) recently noted that ‘the notorious lack of success of teacher PD is too well known to keep hiding or assuming that we should continue as if this evidence is not aplenty (DuFour & Eaker, 1999; Lewis, 1997; Louis & Leithwood, 1998; Timperley, 2005; Wald & Castleberry, 2000). “A major reason for this lack of success is that too much professional development for teachers does not have enhancement of student learning as the contingency of success. Too often, PD is more related to working conditions (of teachers and students), and correlates of student learning. Indeed, in her recent synthesis of literature Timperley was able to locate only 17 articles that related the effects of PD on student learning!

She continued:

Generic delivery models of much external professional development have often proved ineffective in creating the depth of shared professional knowledge needed if staff are to address complex teaching and learning issues in their schools, particularly in those schools facing challenging circumstances (DuFour & Eaker, 1999; Lewis, 1997; Louis & Leithwood, 1998; Wald & Castleberry, 2000). Part of the depth required is an understanding of the contextual conditions in which the new learning must be applied (King & Newmann, 2000). Every school contains a diverse mix of teachers and students with varying competencies and attitudes and a unique set of social, cultural and political conditions, all of which have a powerful influence on teaching and learning (Bryk, Sebring, Kerbow, Rollow, & Easton, 1998; Lytle & Cochran-Smith, 1994). These complex conditions often present obstacles for teachers attempting to apply new ‘generic’ learning from conventional professional development programs to their own classroom practice (Clement & Vandenberghe, 2000; DuFour, 1999; Hord, 1997; Lashway, 1998; Leo & D’Ette, 2000; Leonard & Leonard, 1999; Louis & Leithwood, 1998; McLaughlin, 1993; Rosenholtz, 1989; Smylie, 1995).

From such an analysis, Timperley recommends developing a culture of using data to support learning and how this ‘needs a mind shift that will rock the foundations of what we do and how we do it’. She proposed five elements of professional learning communities:

1. The development of shared values and expectations about children, learning, teaching and teachers’ roles and the relationship of these to the environment (Bryk et al. 1999; Louis et al. (1996);
2. The collective focus on student learning that then becomes part of the normative control of the professional community (Bryk et al., 1999);
3. Collaboration, whereby professional communities foster the sharing of expertise and faculty members call on each other to discuss the development of skills and create shared understandings of effective practice;
4. Deprivatised practice, and much time and opportunity to talk to each other about teaching;
5. Reflective dialogue implies self-awareness about one’s work as a teacher through engaging in in-depth conversations about teaching and learning (Louis, Marks et al., 1996).

These all require a serious commitment to evidence, debates about the contested nature and value of evidence, and actions based on evidence. This is a major culture shift for many schools, where privatised teaching occurs, discussion is more about curriculum and students and less about teaching, and evidence of growth in learning is rarely shared across the school.

Concluding comments

The major argument of this presentation is to move the discussion away from data towards interpretations, from student outcomes to teaching successes and improvements, and from accountability models located about schools to located first in the classroom to support such evidence-based teaching and learning. The asTTle model, which has been developed in New Zealand, will be used in the Keynote presentation to demonstrate such a model. By locating evidence in the classroom we can improve the quality of information and interpretations sent to students, parents, Ministries, Ministers, and thence the community. We can influence the major agent that influences student and learning – the teacher, can highlight the debate about what is worth teaching, and, most importantly, can begin to establish a teacher-shared language about the achievement progression. The model is based on target setting, on ensuring the implementation of the curricula, and by comparisons to appropriate national and local standards of performance. The major sources of evidence relate to diagnosis and formative assessment models and are centred on three major questions: Where are we going? How are we going? and Where to next? All analyses can be conducted at the individual as well as at the cohort, class, and school levels. The evidence can also be used to contest deeply held beliefs about what should be undertaken in the name of curriculum reform, and can lead to asking direct questions about where the curriculum needs to be reformed, and where it should be left alone. Within schools, this evidence-based accountability model can be used to ask relative questions about the effectiveness of teaching, can be recast in terms of learning intentions and success criteria, and evidence provided about the quality of teaching rather than the quality of the students that a school receives. It is important to consider teachers' conceptions of assessment, and to use evidence as the basis for professional development programs. Perhaps students' evaluations of teaching could be also used as part of this evidence base. The move to collecting more data needs to be stopped and the move to making more defensible interpretations about teaching and learning upgraded to priority levels. Evidence that informs teachers about their teaching is the most critical evidence that can be provided and too many current models ignore such evidence. It is possible to devise a national accountability model based on evidence critical to teachers, and such a model can also serve to evaluate the state of learning in the nation, to provide evidence for curriculum reform, to create debate about what is worth learning in our schools, and to develop a common language about the progression of this learning as students advance through their schooling.

References

- Alton-Lee, A. (2003). *Best evidence synthesis*. Ministry of Education, Wellington, New Zealand. <http://www.educationcounts.govt.nz/publications/series/2515/5959>.
- Bishop, R., Berryman, M., Tiakiwai, S., & Richardson, C. (2003). *Te Kōtahitanga*:

- The Experiences of Year 9 and 10 Māori Students in Mainstream Classrooms*. Poutama Pounamu Research and Development Centre/Māori Education Research Institute (MERI), School of Education, University of Waikato.
- Brown, G. L.T. (2004). *Teachers' conceptions of assessment*. Unpublished doctoral dissertation. New Zealand: University of Auckland.
- Bryk, A., S. (1999). Policy lessons from Chicago's experience with decentralization. In D. Ravitch (Ed.), *Brookings Papers on Education Policy*. Washington, DC: Brookings Institution Press, pp. 67-128.
- Bryk, A. S., Sebring, P. B., Kerbow, D., Rollow, S., & Easton, J. Q. (1998). *Charting Chicago school reform: Democratic localism as a lever for change*. Boulder, CO: Westview Press (ED421594).
- Clarke, S., Timperley, H., & Hattie, J.A. C. (2003). *Assessing formative assessment*. New Zealand: Hodder Moa Beckett.
- Clay, M. M. (1993). *An observation survey of early literacy achievement*. Auckland, New Zealand: Heinemann Education.
- Clement, M., & Vandenberghe, R. (2000). Teachers' professional development: A solitary or collegial (ad)venture? *Teaching and Teacher Education*, 16, pp. 81-101.
- Clinton, J. M., & Hattie, J.A. C. (2005). *When families learn the language of schools: Evaluation of the Flaxmere Project*. Ministry of Education, Wellington.
- Cronbach, L.J., & Furby, L. (1970). How should we measure 'change' – Or should we? *Psychological Bulletin*, 74, pp. 68-80.
- Cuban L. and Tyack, D.B. (1995). *Tinkering toward utopia: a century of public school reform*. Cambridge Mass: Harvard University Press.
- DuFour, R. (1999). Challenging role: Playing the part of principal stretches one's talent. *Journal of Staff Development*, 20 (4), pp. 1-4.
- DuFour, R., & Eaker, R. (1999). *Professional learning communities at work: Best practices for enhancing student achievement*. Bloomington, IN: National Educational Service (ED426 472).
- Ell, F. (2002). *Strategies and thinking about number in children aged 9-11 years* (Tech. Rep. No. 17). Auckland, NZ: University of Auckland, Project asTTle.
- Hanushek, E. (2005, May). *Alternative Policy Options: Class Size Reduction and Other Possibilities*. Keynote presentation at the International Conference on Learning Effectiveness and Class size. Hong Kong.
- Hattie, J.A.C., & Peddie, R. (2003). *School reports: 'Praising with faint damns'*. Set: Research information for teachers, 3, pp. 4-9.
- Hattie, J.A.C. (1999). *Influences on student learning*. Inaugural Professorial Address, University of Auckland.
- Hattie, J.A.C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Oxford, UK: Routledge.

- Hattie, J.A.C. (2003, October). *Teachers make a difference: What is the research evidence?* Keynote presentation at the Building Teacher Quality: The ACER Annual Conference, Melbourne, Australia.
- Hattie, J.A. C., Brown, G.T., & Keegan, P. (2003, January). *The Advantages and Disadvantages of State wide and/or National Testing*. Hawaii International Conference on Education, Honolulu.
- Hattie, J.A.C., & Rowe, K. (2004). *Review of current indicators constructed using NEMP data: Consultancy report to the New Zealand Ministry of Education*. Ministry of Education, Wellington, New Zealand.
- Hord, S. M. (1997). *Professional learning communities: Communities of continuous inquiry and improvement*. Austin, TX: Southwest Educational Development Laboratory.
- Irving, E. (2005). *The development and validation of a student evaluation instrument to identify highly accomplished mathematics teachers*. Unpublished Doctoral dissertation. University of Auckland.
- King, M. B., & Newmann, F. M. (2000). *Will teacher learning advance school goals?* Phi Delta Kappan, 81, pp. 576-580.
- Lashway, L. (1998). Creating a learning organization. *ERIC Digest*, Number 121: ED420897.
- Leo,T., & D'Ette, C. (2000). Launching professional learning communities: Beginning actions. *Issues About Change*, 8 (1).
- Leonard, L. J., & Leonard, P. E. (1999). Reculturing for collaboration and leadership. *Journal of Educational Research*, 92, pp. 237-242.
- Lewis, A. (1997). A new consensus emerges on the characteristics of good professional development. *The Harvard Education Letter*, 13 (3), May/June.
- Liberty, K., & Miller, J. (2003). *Research as a resource for evidence-based practice*. Dunmore Press, Palmerston North, NZ.
- Linn, R.L. (2003). Accountability: responsibilities and reasonable expectations. *Educational Researcher*, 32 (7), pp. 3-13.
- Louis, K. S., & Leithwood, K. (1998). From organizational learning to professional learning communities. In K. Leithwood & K. S. Louis (Eds.), *Organizational learning in schools*, pp. 275-285. Lisse: Swets & Zeitlinger.
- Louis, K. S., Kruss, S. D., & Marks, H. M. (1996). School wide professional community. In F.M. Newmann & Associates (Eds.), *Authentic achievement: Restructuring schools for intellectual quality*, pp. 179-207. San Francisco: Jossey-Bass Publishers.
- Lytle, S. J., & Cochran-Smith, M. (1994). Inquiry, knowledge, and practice. In Sandra Hollingsworth & Hugh Sockett (Eds.), *Teacher research and educational reform: 93rd NSSE yearbook*, Part I, pp. 22-51. Chicago: National Society for the Study of Education, University of Chicago Press.
- McLaughlin, M. (1993). What matters most in teachers' workplace context? In J.W. Little & M. McLaughlin (Eds.), *Teachers work*, pp. 79-103. New York: Teachers College Press.

- Phillips, G., McNaughton, S., & MacDonald, S. (2001). *Picking up the pace: Effective literacy interventions for accelerated progress over the transition into decile 1 schools*. Wellington: Ministry of Education.
- Mosteller, F., & Boruch, R. (2002). *Evidence matters: randomized trials in education research*. Brookings Institute, Washington DC.
- Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. *Harvard Educational Review*, 67, pp. 41-74.
- Robinson, V. M. J., & Lai, M. K. (2005). *Practitioner research for educators: A guide to improving classrooms and schools*. Thousand Oaks, CA: Corwin Press.
- Rosenholtz, S. (1989). Workplace conditions that affect teacher quality and commitment: Implications for teacher induction programs. *The Elementary School Journal*, 89, pp. 421-439.
- Rubie, C. M., Hattie, J.A. C., & Hamilton, R. J. (2006). Expecting the best for New Zealand students: teacher expectations and academic outcomes. *British Journal of Educational Psychology*.
- Schutzwohl, A. (1998). Surprise and schema strength. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, pp. 1182-1199.
- Scriven, M. (1988). The new crisis in teacher evaluation: the improper use of 'research-based' indicators', *Professional Personnel Evaluation News*, 1 (1), pp. 4-8.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29 (7), pp. 4-14.
- Smylie, M.A. (1995). Teacher learning in the workplace: Implications for school reform. In T. R. Guskey & M. Huberman (Eds.), *Professional development in Education: New paradigms and practices*, pp. 92-113. New York: Teachers College, Columbia University.
- Timperley, H. S. (2004). School-based influences on teachers' learning and expectations. In F. Columbus (Ed.), *Teaching and Teacher Issues: Nova Science*.
- Timperley, H. S. (2005). Instructional Leadership Challenges: The Case of Using Student Achievement Information for Instructional Improvement, In Review. Timperley, H. S., & Robinson, V. M. J. (1998). Collegiality in schools: Its nature and implications for problem solving. *Educational Administration Quarterly*, 34 (Supplemental December), pp. 608-629.
- Timperley, H. S., & Robinson, V. M. J. (2001). Achieving school improvement through challenging and changing teachers' schema. *Journal of Education Change*, 2, pp. 281-300.
- Timperley, H. S., & Robinson, V.M.J. (2003). Partnership as intervention strategy in self-managing schools. *School Effectiveness and School Improvement*, 14 (3), pp. 249-274.
- Timperley, H., Robinson, V. M. J., & Bullard, T. (1999). *Strengthening education in Mangere and Otara: First evaluation report*. Auckland: University of Auckland.

- Timperley, H. S. & Wiseman, J. (2003). *The sustainability of professional development in literacy: Part 2. School based factors associated with high student achievement* Wellington, New Zealand: Ministry of Education.
- van Gardaren, D. (2002). *Paper and pencil strategy assessment items for the New Zealand Number Framework: Pilot study*. (Tech. Rep. No. 18). Auckland, NZ: University of Auckland, Project asTTle.
- Wald, P. J., & Castleberry, M. S. (Eds.). (2000). *Educators as learners: Creating a professional learning community in your school*. Alexandria, VA: Association for Supervision and Curriculum Development. (ED439099).
- Yair, G. (2000). Educational battlefields in America: The tug-of-war over students' engagement with instruction. *Sociology of Education*, 73, pp. 247-269.