# MAKING PROGRESS— MEASURING PROGRESS

*Conference Proceedings*

NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH
TE RŪNANGA O AOTEAROA MŌ TE RANGAHAU | TE MĀTAURANGA

MARCH 2008

# Making progress—measuring progress

**Conference Proceedings**

**March 2008**

Compiled by Joanna Morton

NZCER PRESS

Wellington 2008

**NZCER PRESS**
New Zealand Council for Educational Research
PO Box 3237
Wellington
New Zealand



ISBN 978–1–877398–39–1



Distributed by NZCER Distribution Services
PO Box 3237
Wellington
New Zealand
www.nzcer.org.nz

# Contents

# Introduction

At NZCER we plan our one-day conferences to give participants the chance to examine a current educational issue of importance in a way that provides new insights and engages and extends thinking. The 2008 NZCER conference sought to examine two fundamental ideas underlying assessment: first, the idea of assessment being viewed as a form of "measurement of learning", and second, the idea that learners might "progress" in some way. These ideas raise a multitude of questions such as: Who determines what progress might be? Do you really need to measure progress in all areas of students' learning to be able to be judged an effective teacher, school, or system? Apart from the issue of cost, or indeed value for money, could you do this with any degree of validity and reliability anyway? These questions led to further important considerations. How could we as educators build our collective understanding about the usefulness and limitations of these ideas of measurement and progress? How could we expand this outwards and also build the confidence of everyone who is interested in education in what we know are sound ideas about measurement and ideas of progress? Perhaps more challenging, are our current constructs of progress useful only in the world of today and yesterday? Do we need new ways of thinking if we are to have an education system and assessment practices that really address the current inequities in outcomes?

This is truly a complex area. As educators we need to think hard about it and to continue to develop our understanding. We need a sense of what is possible and educationally defendable and what is outside the bounds of our current understanding. We ourselves need to keep learning. The papers in this publication provide a very useful source of material to support this learning. There is not just one view, and no one answer, no recipe or formula, and no silver bullet. There are, however, choices to be made, and the papers provide insights and information that will enable us to make decisions that are educationally sound. They will also inform our contribution to the ongoing debate about assessment for the classroom, the school, the system, and in terms of what we as a nation want and expect from our national education system.

**Robyn Baker**

*Director*
New Zealand Council for Educational Research

# Zooming in and zooming out: Challenges and choices in discussions about making progress

**Margaret Carr**

*University of Waikato*
*Wilf Malcolm Institute of Educational Research*

## Introduction

I have always thought that the expression *making progress* is an interesting one. We use it in an everyday sense of "moving forward towards a clearly articulated goal"; it usually means improving, getting better at, achieving more. The emphasis is on the word *progress* and the notion of "progressing". We can also use the expression *making progress* to refer to the construction of a pathway, where the emphasis is on the making. So, making progress implies something about an endpoint (the domain of interest and the direction in which we might be heading) and something about the pathway (the track that we are invited, or compelled, to take).

Sometimes the goal will be clearly articulated, and sometimes it won't; sometimes the pathway will be clearly set out in milestones, sometimes it won't. I want to argue that we can still talk about progress in the uncertain and not-clearly-set-out scenarios, but we have to talk about it differently from when we consider the certain and the clearly-set-out.

Here is an example of a task in which the goal and the pathway towards it have been constructed as (at least moderately) certain and clearly set out, and the way we talk about it in an everyday conversation. Sally Peters and I have recently completed a final report for a TLRI (Teaching and Learning Research Initiative) project (Carr, Peters, et al., in press). Our conversations at the latter stages probably used the everyday meaning of progress, where we know the goal and can recognise the steps along the way. These conversations might have included the following:

> Margaret: I think we are making progress.
>
> Sally: Yes. But there's a bit more work to do in Chapter Three.
>
> Margaret: And there are still some gaps in Chapter Four.
>
> Sally: I think we'll get it completed in time.

The TLRI requirements for a final report were quite clearly set out (the goal), and we would know when we had reached the goal (the chapters would be complete). It's not a perfect analogy, of course, because we didn't actually write the chapters in sequence; the pieces were later put together as a logical and connected pathway for the reader. And there is always a sense of a "work in progress" for a TLRI final report, because it provides the foundation for later ideas, and those later ideas may well suggest some reconfiguring of the earlier pathway.

There is something of a feeling of "fools rush in where angels fear to tread" here, since this may be one of the biggest questions in education. Writing about transfer of learning—an aspect of progress—Ference Marton points out that "As this is basically the main question of all education, we can probably agree that it should not be neglected for long" (2006, p. 499). It certainly has a big literature, and the ambiguous nature of the notion of progress and progression is reflected in the names that have been used: names that reflect different meanings, metaphors, and theories about learning. Here is a selection (the authors I have attached are the ones familiar to me, but they are not necessarily those who set out the label first).

## Progress over time

telos ("a direction of movement or change of learning (not the same as goal-directed activity)") (Lave, 1996)

lines of flight (and learning as "rhizomatic") (Deleuze & Guattari, 1980/1987)

learning trajectory (Wenger, 1998)

developmental trajectory (Bronfenbrenner, 1979)

PLODS (Possible Lines of Direction) (Whalley, 1994)

transformation of participation (Lave, 1996; Rogoff, 2003)

episodes as linked chains of interaction (Scott, Mortimer, & Aguiar, 2006, cited in Mercer & Littleton, 2007, p. 116)

co-constructed pathways (O'Connor, in press)

## Progress over time and space

propagation (Beach, 2003)

chains of signification (Lacan, 1977; Walkerdine, 1997)

tracers (Lemke, 2001)

connection-building (Gee & Green, 1998)

PFL (Preparation for Future Learning) (Bransford & Schwartz, 1999)

transfer in pieces (Wagner, 2006)

dimensions of strength (Carr, 2006)

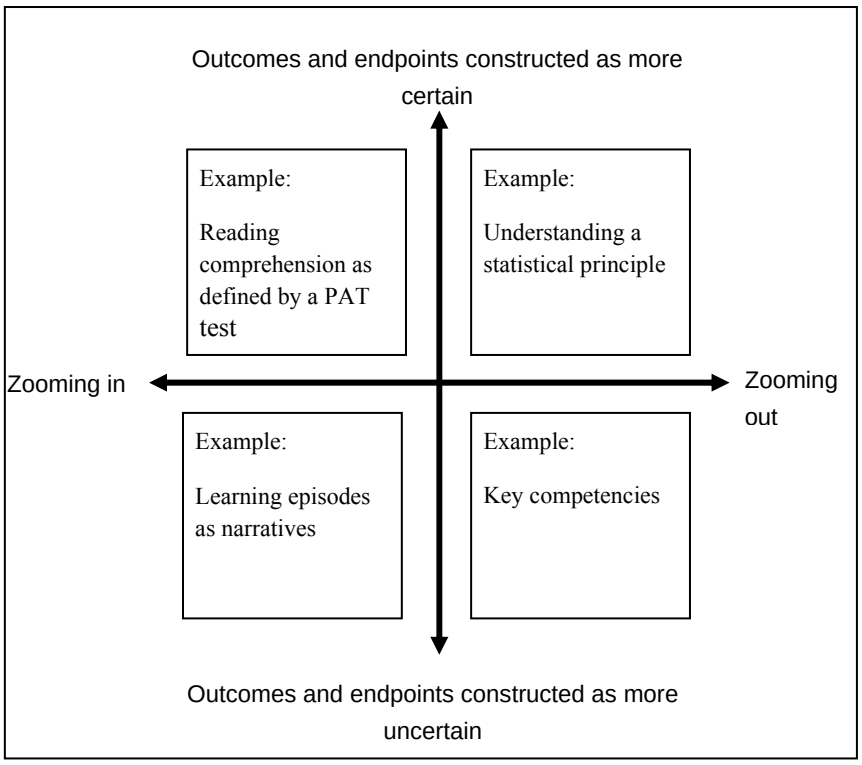
In each list it is the last two that I will mostly refer to here. I will be picking up on *episodes as linked chains of interaction* and *co-constructed pathways* and referring to progress over time and place as *transfer in pieces* and *dimensions of strength*. You will begin to see that, like many of the authors above, I am interested in metaphors associated with pieces rather than lines, but pieces as they form and re-form into sequences when new circumstances and new pieces arrive. I end up

with a notion of making progress as "progress in pieces". (As I reread this, I wonder if I am suggesting that we tear up the notion of progress; but I want to keep it, reframed.)

## A plan

I want to start by setting out some ideas in a diagram that connects two processes, in order to try to describe a complicated picture in a way that might assist us to understand the challenges and choices that face us when we consider progress. The first process is about the focus. I have called it "zooming in" and "zooming out" on outcomes, a reference to processes that I can do with my digital camera and my computer. The idea is used in education in slightly different ways in Roth (2001) and Halverston (as "telescoping": 2005); Marton (2006, p. 507) writes about "grain size" in discussions about transfer. Zooming in gives us outcomes and events that are small in the wider scheme of big-picture curriculum. As units of analysis they have fairly clear boundaries, even if for some of them (episodes of learning, for instance) their significant features are uncertain. The second process is thus about the level of uncertainty, from endpoints or pathways that have been constructed as certain or "moderately certain", to endpoints and pathways that we acknowledge to be very uncertain. I have used these as two axes in Figure 1, which sets out four quadrants.

Figure 1 **Outcomes and pathways: four quadrants**



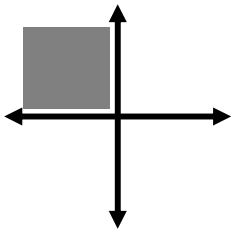

I plan to consider one or two examples in each quadrant, rather whizzing past the top two (the more certain) quadrants, where others are much more expert than I am. I will be taking more time on the bottom two quadrants, the uncertain ones. I have more to say in those contexts where the units of analysis are uncertain and fuzzy—for example, learning dispositions and key competencies—because I have read and researched more on those. And I have done more reading

and research on those because, in my view, and in the view of many others (for example, Claxton, 2002; Craft, Gardner, & Claxton, 2008; Ritchhart, 2002), they are the major contributors to lifelong learning and to wise, resourceful, creative, and considerate citizens. They take some of the more certain outcomes and turn them into action: they include motivation and disposition, and they are closely tethered to, or take account of, situation and context. I am certainly not an expert here either, but many teachers I have worked with are now becoming very familiar with these sorts of outcomes and are expertly exploring ways to recognise and document their progress (this paper includes examples from some of their classrooms).

Jean Lave has set out rather clearly some aspects of a social or sociocultural theoretical approach to learning that I have found to be a useful starting point; I quote from a 1996 paper:

> Why pursue a social rather than a more familiar psychological theory of learning? To the extent that being human is a relational matter, generated in social living, historically, in social formations whose participants engage with each other as a condition and precondition for their existence, theories that conceive of learning as a special universal mental process impoverish and misrecognize it. (p. 149)
>
> Common theories of learning begin and end with individuals (although these days they often nod at 'the social' or 'the environment' in between). Such theories are deeply concerned with individual differences, with notions of better and worse, more or less learning, and with comparison of these things across groups-of-individuals. Psychological theories of learning prescribe ideals and pathways to excellence and identify the kinds of individuals (by no means all) who should arrive …. A reconsideration of learning as a social, collective, rather than individual, psychological phenomenon offers the only way beyond the current state of affairs that I can envision at the present time. (p. 149)
>
> The region of social theory that seems richest in clues for how to conceive of learning in social terms, in my view is that of historical, dialectical, social practice theory. Such a theoretical perspective takes learning to be an aspect of participation in socially situated practices. … Further, I found that apprenticeship studies offered an especially clear window on issues about learning. But even supposing that this claim is correct, how could apprenticeship studies be relevant to learning in school settings? The argument developed by Etienne Wenger and myself (Lave & Wenger, 1991) is that learning is an aspect of changing participation in changing 'communities of practice' everywhere. (p. 150)

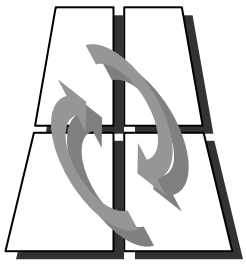
This notion of learning being a relational matter was a viewpoint that teacher-researchers in our TLRI project returned to again and again, and the work of Russell Bishop and colleagues on the Kotahitanga project (Bishop, Berryman, Cavanagh, Teddy, & Clapham, 2006) is based on it; see also Peters (in press). The following discussion wrestles with its implications for the notion of progress. This is not at all an academic exercise: it matters a great deal. One's identity as a competent learner and knower is to some extent built around the notion of making progress, getting better at something, becoming more expert, and knowing more. Our opportunities to build these identities will be impoverished and narrowly configured if they take as reference points only those occasions when the goal is certain and the pathway is clearly set out in steps. Nor will social justice be served by such simplification.



## Quadrant one: Zooming in on outcomes, and more certain endpoints and pathways

In New Zealand, and probably in many other countries, we have a very early model of development in the Plunket Book or the WellChild/Tamariki Ora Health Book. Our parents, and we too if we have been closely involved in watching the growth of babies we know, have a powerful working theory of progress that has often been accompanied by emotional content. "How is she doing?", we ask, carefully looking at the graph of weight gain or loss. This notion of progress has its home in quadrant one. I illustrated it in the presentation with the graph of Mereana's progress. She was 1.4 kg when born; the graph declined rather dramatically in the first few days, then "turned a corner" as she regained what she had lost and slowly edged her way up towards the norm. As a family we knew what progress meant; the units of outcome (weight) were, we knew, very significant, and we watched the graph with some anxiety.

I have also used as an illustration of quadrant one an example of a virtual student's level of progress, in comparison to a norm, from the recently developed PAT test of reading comprehension (Darr, McDowall, Ferral, Twist, & Watson, 2008). The paper by Charles Darr and Sue McDowall, in this volume, explains well just what this tells us about progress, and what it doesn't, and what decisions they had to make in order to construct a pathway of progress. This example is different from the graphs of weight and length for babies; the authors make the comment that "when height is measured, for example, everybody is expected to agree on which of the two objects is the taller, and that the results do not depend on what rulers are used" (Darr et al., 2008, p. 24). They point out that in the PAT test the "ruler" is constructed. We expect this construction—the criteria and the levels of difficulty and the norming process—to be explained and transparent so that we can judge whether this test is going to be relevant and helpful for our purpose.



I want to complicate the model a little further, and introduce some arrows from one quadrant to another as I proceed. In my view, apart from their value within their own "home" quadrant, the findings in the quadrant where the PAT tests sit will also contribute to our understanding of some of the more uncertain and zoomed-out pathways and domains of outcome.

The 1990s work on thinking dispositions by David Perkins and others at Harvard's Project Zero developed a three-part definition of the dispositional outcomes that belong in that uncertain and zoomed-out quadrant. He and colleagues describe these three aspects as: ability, inclination, and sensitivity to occasion (Perkins, Jay, & Tishman, 1993), and teachers and researchers have found this three-part definition to be helpful. Perkins has sometimes called them the three As: ability, attitude, and alertness. Research at Project Zero (see, for example, Ritchhart, 2002) found that sensitivity to occasion—alertness to context—was particularly important. Researchers on progress from a "situated" or sociocultural perspective have also taken an interest in this aspect of learning: Greeno et al. (1998), Bransford and Schwartz (1999), and Marton (2006), for example. Greeno et al. refer to "attunement" (p. 8). Marton writes about "the important process of discerning differences" (p. 504). The PAT test tells us a lot about progress in the *ability* dimension of that triad. Cathy Wylie and Edith Hodgen's longitudinal Competent Children research indicates that,

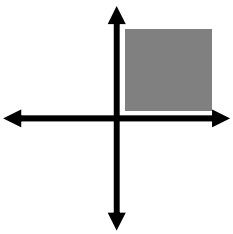
in that study, reading competence at age five did not predict competence at age eight very well. The authors comment:

> It would seem that our early literacy measure was not a good predictor of near-adult literacy and/or that something happened between age 5 and age 8 for some students, and failed to happen for others. Either way, the relative achievement in literacy at age 8 was more closely associated with achievement in literacy at age 16 than the age-5 achievement was. (Wylie & Hodgen, 2007, pp. 54–55)

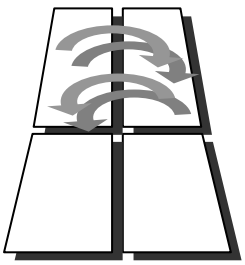
Consistent with the authors' second hypothesis ("something happened") is the theory that *being a reader* includes motivation and context (the inclination and the occasion) that attach themselves to reading ability in the early years of school in a range of important ways that are difficult to predict. My reading of the work of Alan Luke and Peter Freebody (1999; Freebody & Luke, 2003) and James Gee (2003) would support this hypothesis. See also Comber (2000). Luke and Freebody point out that literacy is not just a matter of cognitive ability, it is also about "access and apprenticeship into institutions and resources, discourses and texts" (1999, p. 5). Freebody and Luke suggest that:

> effective literacy in complex print and multimediated societies requires a broad and flexible repertoire of practices. This repertoire we have characterised as a set of 'roles', later 'practices', that participants in literacy events are able to use or 'resource'. (2003, p. 56)

These are (not in any particular order): breaking the codes (recognising and manipulating the units and symbols); participating in the meanings; using texts functionally; and critically analysing and transforming texts. Breaking the codes belongs in quadrant one, but the other practices that contribute to being a reader zoom out more, and what they look like is less precisely defined and certain (they are "broad and flexible"). In this case, insights from quadrant one are important, but they are not enough.



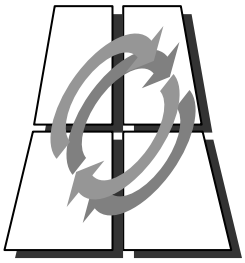
## Quadrant two: Zooming out on big-picture outcomes where the endpoint is moderately certain



As I thought about progress I was intrigued to read a paper by Joseph Wagner called "Transfer in Pieces" (2006). This is a case-study analysis of an undergraduate student's attempt to solve a series of problems or tasks that finally contributed to an understanding of an elementary statistical principle (the law of large numbers): a big-picture outcome that is, in the domain and principles of mathematics, certain. It illustrated how she came to see problems initially thought to be different as alike. In this way she slowly came to identify different problems as instances of a single principle (p. 3). Wagner comments: "The approach proposed here argues that abstractness—situated or not—is not the key to understanding transfer, but rather that knowledge supporting transfer must somehow *account for* (rather than *overlook*) [italics original] contextual differences

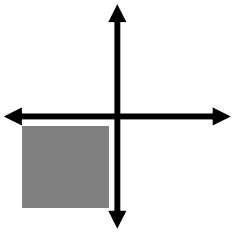
across otherwise 'similar' activities" (p. 6). He writes about this process as constructing a big-picture "contextual neighbourhood".

This notion of accounting for rather than overlooking contextual differences is also a feature of the work of Bransford and Schwartz (1999). Bransford and Schwartz write about a Preparation for Future Learning (PFL) perspective on transfer and learning outcomes. In PFL the focus of assessment shifts to assessments of people's abilities to learn in knowledge-rich environments (p. 68). "The ideal assessment from a PFL perspective is to directly explore people's abilities to learn new information and relate their learning to previous experiences" (p. 69). They comment that "the learning experiences 'set the stage' for further noticing" (p. 74). Experiences with contrasting cases set the stage for learning new information, but the opportunity for follow-up learning was needed. "The multiply embedded social settings within which people's lives unfold have a powerful effect on the degree to which they are supported in letting go of older ideas and practices and attempting new ones" (p. 81). I see a connection here with Perkins' notion of "sensitivity to occasion". There is also a connection with the idea of "episodes of linked chains of interactions" that contribute to understanding concepts in science:



> [Our] analysis of the discourse of science lessons involves an iterative process of moving backwards and forwards through time, trying to make sense of the episodes as linked chains of interactions.
>
> (Scott et al., 2006, cited in Mercer & Littleton, 2007, p. 116)

This model of progress might also be illustrated by the opportunities for learners to discuss their prior knowledge and to link ideas as they develop an understanding of a big-picture concept in science (the water cycle) in the NZCER Kick Starts resource *Key Competencies: The Water Cycle, a Science Journey* (Bull, Hipkins, Joyce, & McIntyre, 2007).



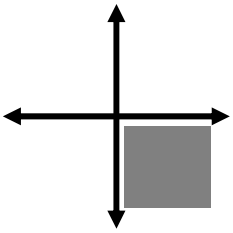
## Quadrant three: Zooming in on small pieces and uncertain pathways

In this quadrant I have placed portfolios or collections of work, including learning stories (Carr, 2001; a way of documenting learning dispositions in early childhood, with particular reference to the national early childhood curriculum, *Te Whāriki*, Ministry of Education, 1996). I think Carlina Rinaldi's nice phrase, "a dictionary of experiences" (2006, p. 76) as a support for dialogue about learning and teaching, finds a home here:

> It seems to me that we were able to structure—though the reader will be a better judge of this—what Kenneth J. Gergen calls a 'transformational dialogue': a dialogue which is able to transform our relationship and, therefore, in a certain way, our professional and group identities. Instead of adopting a 'top-down' approach, with a prior definition of identical rules, ethics and practices for everyone involved, we managed to move into a sphere of action in which children and adults alike struggled successfully, it seems, with the problem of learning in a multiple and conflictual context. We thus compiled a sort of 'dictionary of experiences' which helped us to reflect, infer, hypothesise and understand. (p. 76)

David Perkins uses the term "learning in the wild" for these bottom two quadrants, with their high levels of uncertainty. Classrooms and early childhood settings are "wild" and complex contexts that Perkins and colleagues (Perkins, Tishman, Ritchhart, Donis, & Andrade, 2000, p. 270) say "present a wilderness of vaguely marked and ill-defined occasions for thoughtful engagement". An abilities-centric view of progress can prevail where the task is prescribed and criteria for success are clearly marked (as in the first quadrant in this paper), but being "in the wild" is different from this. It includes interpretation and re-cognition. In my presentation I described an example from Mangere Bridge Kindergarten's Centre of Innovation research, where a child took his kindergarten portfolio to school, to the new entrant classroom and then into Year 1. Other children did too, and the portfolios of learning stories became literacy artefacts, and opportunities for interaction between peers and between adults and children. They provided a site for "belonging" and they represented the children's developing identities as competent and confident learners.

Many of these small units of analysis in quadrant three are stories, and we know that stories can be a valuable unit of analysis for the understanding of learning, especially when the definition of learning includes dispositions (Bruner, 2002; Sfard & Prusak, 2005).



## Quadrant four: Zooming out on big-picture outcomes and uncertain pathways

Now we arrive at the final quadrant: zooming out on big ideas to do with lifelong learning dispositions and key competencies—outcomes that include intention and inclination, that turn abilities into action, and that are closely attached to the context of relationships and support and "content". The 2007 *New Zealand Curriculum* gives this description of the key competencies: "More complex than skills, the competencies draw also on knowledge, attitudes, and values in ways that lead to action…. As they develop the competencies, successful learners are motivated to use them, recognising when and how to do so and why" (Ministry of Education, 2007, p. 12).

In the 2007 *New Zealand Curriculum* document, progress is set out from early childhood to tertiary on page 42, aligning the domains of learning disposition and working theories in *Te Whāriki* with the key competencies at school (and with a nod to the tertiary sector as well). In the presentation I noted the fading on the left-hand edge of the bands of progress, reminding us that many definitions of progress do not just begin at birth: "progress" connects back to all those who have gone before, and to the stories, discoveries, and constructions of progress that history has bequeathed us—some of which we are considering in this conference. This is consistent with a kaupapa Māori viewpoint, eloquently spoken by Tilly Reedy (1995/2003, pp. 55–56) in the early days of the development of *Te Whāriki*:

Māori tradition identified the Māori child as a valued member of the Māori world—before conception, before birth, before time. The child was the personification of the worlds of yesterday:

He purapura i ruia mai Rangiatea

E kore e ngaro

Precious seeds dispersed from Rangiatea

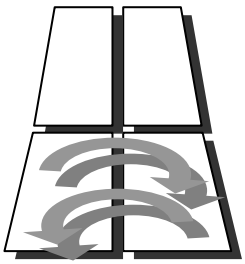
[the famed homeland of the Māori gods]

Will never be lost

Indoctrination of mana and pride, of knowledge of their aristocratic lineage, of histories, of descriptions of chiefly ornaments and cloaks, took place through the many lullabies that were composed on the birth of the child and sung to them constantly.

## Dimensions of strength

There is now considerable policy and practice interest in continuity and transfer of learning dispositions and key competencies over time and place. One attempt to find ways to make progress with these big-picture outcomes has developed from an interrogation of some of the smaller pieces from the previous quadrant in order to make the kind of connections across quadrants that I recognised as similar to Wagner's transfer in pieces. The small pieces came from data in the *Kei Tua o te Pae* work (Carr, Lee, & Jones, 2004, 2007, & in press), and the TLRI project *Key Learning Competencies Across Place and Time: Kimihia te Ara Tōtika, hei Oranga mō to Aō* (Carr, Peters, et al., in press). Four "dimensions of strength" were developed (Carr, 2006). They connect with the sociocultural literature on "affordance networks" and activity-theoretical studies (for example, Barab & Roth, 2006; Engeström, Engeström, & Suntio, 2002), and with the four curriculum principles in *Te Whāriki*. In this paper I introduce them and briefly provide some examples; a more elaborate discussion is in the 2006 paper (Carr, 2006, currently on the Ministry of Education website).



When they were first developed, the dimensions were called: *mindfulness* (as learners begin to "make these practices part of their own identity and expertise"); *breadth* (more wide-ranging contexts, as connections are made beyond any one setting or community); *frequency* (over time, in facilitating environments); and *complexity* (across mediating resources—including people). These dimensions overlap and intertwine. We have since begun to call them the ABCD of strength: agency, breadth, continuity, and distribution. *Agency* in this context now includes an increasingly critical and innovative perspective on a key competency. Greeno (2006) writes about "authoritative, accountable positioning as an appropriate condition for learning". Calling on a 2002 study by Lobato and Siebert, he uses the term "actor-oriented" (p. 544) perspectives on progress and transfer, a term that matches with Wagner's view that it is important to investigate the students' own perceptions of similar and different tasks and learning experiences. *Breadth* is about the widening connections to social practices and communities outside the classroom. *Continuity* in a conceptual domain is the recognition by the learner of continuities that link the

past, the present, and the possible future. It demands a facilitating environment that provides space and routines for conversations about learning, for practising, and for trial and error. It includes recognising what is the same and what is different in episodes of learning—alertness, discernment, and attunement. Continuity is jointly constructed by teachers and learners, and families have a role in this as well. The term *distribution* refers to the literature on distributed cognition and distributed learning, which says that all learning is distributed across or "stretched over" resources such as people and artefacts (see the chapters in Salomon, 1993; and the quote from Jean Lave earlier in this paper): hence the idea in sociocultural theory that most learning is situated in one place, tethered to a particular terrain of resources and relationships, a key idea that this paper is puzzling about. For key competencies and learning dispositions the learning is also stretched over "learning area" content (which may or may not include social practices and communities outside the classroom).[1]

The dimensions of strength acknowledge the distributed nature of learning and suggest that a key requirement of progress is therefore to hone one's "navigational" capacity, one's disposition (including ability) to recognise and construct connections, similarities, and differences across different terrains of resource and relationship. Elsewhere (Carr, 2008) I have summarised this as having agency and resourcefulness, quoting Newfield, Andrew, Stein, and Maungedzo (2003, p. 79): "In thinking about assessment in the multimodal classroom, we have placed agency and resourcefulness at the centre of what is to be assessed." The dimensions of strength are a work in progress; the aim is to have something to say about progress that accounts for, rather than overlooks, the connections to motivation (agency and continuity) and context (breadth and distribution). The following connections between small pieces (episodes of learning and learning stories) and the bigger key competencies or learning dispositions might contribute to these ongoing discussions about progress.

## Examples of dimensions of strength

*Agency.* An example of agency (from Carr et al., in press) is Ethan, who is interested in art and has been practising looking closely in art activities. One such activity is a still life, where he paints a vase of flowers that has been set up for this purpose at the art table. He decides to use cotton buds as one of his tools for part of the painting. He is producing a "text" that represents

---

[1] The 2007 *New Zealand Curriculum* says of key competencies:

> Opportunities to develop the competencies occur in social contexts. People adopt and adapt practices that they see used and valued by those closest to them, and they make these practices part of their own identity and expertise.
>
> The competencies continue to develop over time, shaped by interactions with people, places, ideas, and things. Students need to be challenged to develop them in contexts that are increasingly wide-ranging and complex. (Ministry of Education, 2007, p. 12)

Four possible dimensions of strength are implied here:

(i) People make the practice of using key competencies *part of their own identity and expertise.* They "own" them and are motivated to use them in different circumstances, interests, or topics.

(ii) They are about interactions in contexts that are *increasingly wide-ranging.*

(iii) Competencies *develop over time*: they are not acquired or possessed at some point in an education; they develop through practice.

(iv) They are about interactions in contexts that are *increasingly complex*, shaped by interactions with people, places, ideas, and things.

some key features of the flowers (the "spikey" nature of the petals). Three weeks later, and not at the art table, he notices the coat hooks as he hangs up his backpack, and "re-cognises" them, saying to the teacher that they look like elephants. She responds by inviting him to draw them as elephants and to turn that into a story; she moves the coat hooks closer to where he is drawing so that he can continue to look closely.

*Breadth.* An example of increasing breadth is provided by the portfolios of 18 children who began school in one teacher's (Yvonne Smith's) new entrant class and moved in the following school year into the same teacher's Year 1 classroom. These portfolios set out educational progress with a particular emphasis on reading, writing, and mathematics by including samples of the children's work, often with a comment (for instance about the reading level) from the teacher. Progress in key competencies was described in learning stories, in particular in what we have called the *personalised group stories* and the *individual stories*. The portfolios record the widening connections with communities of reference: asking at home for information; going on visits; and the invitations to families (usually taken up) to make comments in the portfolios. Lily's family sent a story from home, dictated by Lily and with photo attached, of the day that she went "Biking without trainer wheels". Yvonne frequently comments on what might be called "possible selves", an aspect that keeps in mind the wider community: Abby is a librarian; Amy is a song writer; Luke's stories include the heading "geologists"; Liam, Mark, and Niko "see themselves as inventors"; Nadia is an author; "He [Nathaniel] could be a mediator".

*Continuity.* In terms of continuity, there is an aspect of social justice that appears in these portfolios, made available by the key competencies and the narrative documentation of them. Lily and Amy, for instance, are no great shakes at writing, and their portfolios include work samples and teacher commentary that indicate this. Yvonne's programme includes improving their reading and writing. However, their learning stories also reveal other aspects of their learning, and the portfolios document the increasing strength of these dispositional aspects as well. The stories record the increase in leadership activities for Amy, and those activities of "high focus". Lily, at free choice time, took the lead in a number bingo game, encouraged others, and praised their effort. On another occasion she was the organiser and the caller for alphabet bingo, with another child. She and Amy shared their news, and the teacher commented: "Lily was demonstrating that she is able to adapt to the task to cater for others' needs. She … knew to ask questions to encourage Amy's participation. Lily's empathy for others is a great strength." Lily led the class for hip-hop sessions: "Lily has a great memory for these dances and is tireless in the performance of them. She teaches by example …. Lily will be a great resource person to maybe lead the creation of other dances."

*Distribution.* Mark's and Max's school portfolios track their increasingly complex interactions with a diverse peer support group (an aspect of increasing distribution across supporting people and resources). When Max builds train tracks with Mark early in his first year of school Yvonne comments, "this is the first time I have seen Max play with another child. He usually plays by himself." (She adds: "This demonstrates to me just how important 'choosing time' is to these young children beginning primary school.") In another free-choice time, Mark, Liam, and Niko play with the Mobilo: they "negotiated and contributed ideas, shared, explained how the construction worked …."

## Co-constructed pathways of learning

Two Christchurch teachers, Nikki O'Connor and Susie Greenslade, developed the notion of a "co-constructed pathway of learning" (CPL). Nikki comments as follows:

> A CPL is like a journey—an exploration journey. The journey may have no predetermined pathway and in fact no destination in sight. Rather it is an exploration of the learning landscape. It is unique to the learner. The learner carves his or her own pathway into the learning landscape. The journey may be fuelled by a variety of things. Passion, curiosity, development and ideas could be some of the driving factors.
>
> Along the way there will be people guiding them, telling stories of terrains they have not yet been to, helping them overcome obstacles and equipping them with the tools they need to navigate their way through the landscape. People will share the journey with them. In this sense the journey is co-constructed. The learning pathway is embedded within the environment. It is entwined with people, places and things. (Carr et al., in press)

The case study of one student's (Diana's) co-constructed pathway about participating and contributing will be written as a working paper for the TLRI project (O'Connor, in press). The pathway began when the teachers brainstormed ideas about the key competencies with the children. For "participation" one of the popular ideas they came up with was "trying something new". Diana took up this challenge and began a pathway of exploration where she challenged herself to trying something new, something she had not done before. There is no doubt that, in this case study, *participating and contributing* (the later 2007 title of this key competency) is being strengthened by the opportunities to learn in Diana's classroom. The four dimensions of strength might map the progress here too. *Agency*: She was one of the children consulted about the meaning of the key competency, and, having made connections with her own words ("trying something new") she had the option of choosing interesting and challenging workshops. A group dance task provided the opportunity for the children to develop a dance piece together, and for Diana to take the lead at one point. *Breadth*: There was a "repertoire of places" established in which she could try something new, and the key competency went home with her mother (who was welcome at the school, and also wrote a learning story for Diana's portfolio). Diana's mother, too, decided to try something she had not done before. *Continuity*: This pathway reflects an "actor-oriented" view of key competency progress in the sense that Diana was, together with the teachers, her mother, and her peers, co-constructing a pathway that made sense to her. The social "stage" and the local "scripts" assisted this process. She was invited to practise being courageous in situations like the karate workshop where she was a "little scared", and the sequence of episodes was documented by Nikki and Susie, including a commentary on the learning. Diana, too, wrote a learning story. That documentation was available for revisiting and for conversations about progress. *Distribution*: Because of the structured opportunities at this school (for example, workshops with students from other classrooms), she commented that she was the only girl from her classroom to attend this workshop but thought it was okay because she was trying something new.

## Concluding comments

Given the four quadrants that I have set out here, I have become increasingly interested in the value of connecting one quadrant to another, especially from left (zooming in) to right (zooming

out) and back again, to develop contextual neighbourhoods and to make progress. It is a view of making progress as "progress in pieces". And I am reminded of the following story, told to me once by Guy Claxton:

> A person is searching on a dark night under the street light.
>
> 'Can I help?' says a passerby.
>
> 'I'm looking for my car keys; I dropped them over there by my car.' (Gestures to the dark [*uncertain quadrant*].)
>
> 'Why search here [*in the clear and certain quadrant*] then?'
>
> Reply: 'Because the light is better.'

In other words, we should not attempt to make progress for outcomes that belong in one quadrant by using only the "making" tools from another home quadrant. We should not try to only use the Plunket Book model for complex outcomes, simply because the light is better there.

## Acknowledgements

I acknowledge with thanks the invitation from Robyn Baker at NZCER to present at this conference, and therefore to think and read more about "making progress". I also want to thank the facilitators and the teacher-researchers who participated in the TLRI project *Key Learning Competencies Across Place and Time: Kimihia te Ara Tōtika, hei Oranga mō to Aō*, and the co-director Sally Peters (the authors in Carr et al., in press), together with the families and the children for their contributions and words. Also the co-researchers and writers of *Kei Tua o te Pae*, Wendy Lee and Carolyn Jones (who, with Margaret Drummond, prepared the diagrams for this paper), and the teachers, families, and children at Mangere Bridge Kindergarten and Mangere Bridge School. A number of other people also contributed ideas and readings to this paper, especially Bronwen Cowie and Jane McChesney. I also acknowledge and thank NZCER and the Ministry of Education for research support and research funding.

## References

Barab, S. A., & Roth, W-M. (2006). Curriculum-based ecosystems: Supporting knowing from an ecological perspective. *Educational Researcher*, *35*(5), 3–13.

Beach, K. (2003). Consequential transitions: A developmental view of knowledge propagation through social organizations. In T. Tuomi-Gröhn & Y. Engeström (Eds.), *Between school and work: New perspectives on transfer and boundary-crossing* (pp. 39–62). New York: Pergamon.

Bishop, R., Berryman, M., Cavanagh, T., Teddy, L., & Clapham, S. (2006). *Te Kotahitanga Phase 3. Whanaungatanga: Establishing a culturally responsive pedagogy of relations in mainstream secondary classrooms.* Wellington: Ministry of Education.

Bransford, J. D., & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61–100.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Bruner, J. (2002). *Making stories: Law, literature and life*. Cambridge, MA: Harvard University Press.

Bull, A., Hipkins, R., Joyce, C., & McIntyre, B. (2007). *Key competencies: The water cycle, a science journey.* Kick Starts series. Wellington: NZCER Press.

Carr, M. (2001). *Assessment in early childhood settings: Learning stories*. London: Paul Chapman.

Carr, M. (2006). *Dimensions of strength for key competencies: Working paper 5: Key learning competencies across place and time: Kimihia te ara tōtika, hei oranga mō to aō.* Hamilton: University of Waikato, Wilf Malcolm Institute of Educational Research.

Carr, M. (2008). Can assessment unlock and open the doors to resourcefulness and agency? In S. Swaffield (Ed.), *Unlocking assessment* (pp. 36–54). London: Routledge.

Carr, M., Lee, W., & Jones, C. (2004). *Kei tua o te pae. Assessment for learning: Early childhood exemplars. A resource prepared for the Ministry of Education.* Books 1–9. Wellington: Learning Media.

Carr, M., Lee, W., & Jones, C. (2007). *Kei tua o te pae. Assessment for learning: Early childhood exemplars. A resource prepared for the Ministry of Education.* Books 10–15. Wellington: Learning Media.

Carr, M., Lee, W., & Jones, C. (in press). *Kei tua o te pae. Assessment for learning: Early childhood exemplars. A resource prepared for the Ministry of Education.* Books 16–20. Wellington: Learning Media.

Carr, M., Peters, S., Davis, K., Bartlett, C., Bashford, N., Berry, P., et al. (in press). *Key learning competencies across place and time: Kimihia te ara tōtika, hei oranga mō to aō.* Final report to the Teaching and Learning Research Initiative.

Claxton, G. (2002). *Building learning power: Helping young people become better learners*. Bristol: TLO.

Comber, B. (2000). What really counts in early literacy lessons. *Language Arts*, *78*(1), 39–49.

Craft, A., Gardner, H., & Claxton, G. (Eds.). (2008). *Creativity, wisdom and trusteeship: Exploring the role of education.* Thousand Oaks, CA: Corwin.

Darr, C., McDowall, S., Ferral, H., Twist, J., & Watson, V. (2008). *Progressive Achievement Test: Reading. 2nd edition. Teacher Manual*. Wellington: New Zealand Council for Educational Research.

Deleuze, G., & Guattari, F. (1980/1987). *A thousand plateaus: Capitalism and schizophrenia*. London: Continuum.

Engeström, Y., Engeström, R., & Suntio A. (2002). Can a school community learn to master its own future? An activity-theoretical study of expansive learning among middle school teachers. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century: Sociocultural perspectives on the future of education* (pp. 211–224). Oxford: Blackwell.

Freebody, P., & Luke, A. (2003). Literacy as engaging with new forms of life: The "four roles" model. In G. Bull & M. Anstey (Eds.), *The literacy lexicon* (2nd ed., pp. 52–65). Frenchs Forest, NSW: Prentice Hall.

Gee, J. P. (2003). Literacy and social minds. In G. Bull & M. Anstey (Eds.), *The literacy lexicon* (2nd ed., pp. 4–14). Frenchs Forest, NSW: Prentice Hall.

Gee, J. P., & Green, J. L. (1998). Discourse analysis, learning, and social practice; a methodological study. In P. D. Pearson & A. Iran-Nejad (Eds.), *Review of research in education 23* (pp. 119–169). Washington, DC: AERA.

Greeno, J. G. (2006). Authoritative, accountable positioning and connected, general knowing: Progressive themes in understanding transfer. *The Journal of the Learning Sciences*, *15*(4), 537–547.

Greeno, J. G., & the Middle School Mathematics Through Applications Project Group. (1998). The situativity of knowing, learning and research. *American Psychologist*, January, 5–26.

Halverston, R. (2005, November). *A distributed leadership perspective on how leaders use artifacts to create professional community in schools.* Paper presented at the annual conference of the University Council of Educational Administration, Nashville, TN.

Lacan, J. (1977). *Ecrits: A selection*. London: Tavistock.

Lave, J. (1996). Teaching, as learning, in practice. *Mind, Culture, and Activity*, *3*(3), 149–164.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lemke, J. L. (2001). The long and the short: Comments on multiple timescale studies of human activity. *The Journal of the Learning Sciences*, *10*(1&2), 17–26.

Luke, A., & Freebody, P. (1999). A map of possible practices: Further notes on the four resources model. *Practically Primary*, *4*(2 June), 5–8.

Marton, F. (2006). Sameness and difference in transfer. *The Journal of the Learning Sciences*, *15*(4), 499–535.

Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. London: Routledge.

Ministry of Education. (1996). *Te whāriki. He whāriki mātauranga mō ngā mokopuna o Aotearoa: Early childhood curriculum*. Wellington: Learning Media.

Ministry of Education. (2007). *The New Zealand curriculum: The English-medium teaching and learning in years 1–13*. Wellington: Learning Media.

Newfield, D., Andrew, A., Stein, P., & Maungedzo, R. (2003). "No number can describe how good it was": Assessment issues in the multimodal classroom. *Assessment in Education*, *10*(1), 61–81.

O'Connor, N. (in press). *Diana's co-constructed pathway of learning: Working paper: Key learning competencies across place and time: Kimihia te ara tōtika, hei oranga mō to aō.* Hamilton: University of Waikato, Wilf Malcolm Institute of Educational Research.

Perkins, D., Tishman, S., Ritchhart, R., Donis, K., & Andrade, A. (2000). Intelligence in the wild: A dispositional view of intellectual traits. *Educational Psychology Review*, *12*(3), 269–293.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly*, *39*(1), 1–21.

Peters, S. (in press). Responsive, reciprocal relationships: The heart of the *Te Whāriki* curriculum. In J. Moyles (Ed.), *Learning together in the early years: Relational pedagogy* (pp. 24–35). London: Routledge.

Reedy, T. (1995/2003). Toku rangatiratanga na te mana-mātauranga: "Knowledge and power set me free …". In J. Nuttall (Ed.), *Weaving Te Whāriki: Aotearoa New Zealand's early childhood curriculum document in theory and practice* (pp. 51–78). Wellington: New Zealand Council for Educational Research. (Based on a keynote presentation to the 1995 Early Childhood convention.)

Rinaldi, C. (2006). *In dialogue with Reggio Emilia: Listening, researching and learning.* Oxford: Routledge.

Ritchhart, R. (2002). *Intellectual character: What it is, why it matters, and how to get it.* San Francisco: Jossey-Bass.

Rogoff, B. (2003). *The cultural nature of human development*. Oxford and New York: Oxford University Press.

Roth, W-M. (2001). Situating cognition. *The Journal of the Learning Sciences*, *10*(1&2), 27–61.

Salomon, G. (Ed.). (1993). *Distributed cognitions: Psychological and educational considerations*. Cambridge, UK: Cambridge University Press.

Scott, P., Mortimer, E., & Aguiar, O. (2006). The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education*, *90*, 605–631.

Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher*, *34*(4), 14–22.

Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction*, *24*(1), 1–71.

Walkerdine, V. (1997). Redefining the subject in situated cognition theory. In D. Kirschner & J. A. Whitson (Eds.), *Situated cognition: Social, semiotic, and psychological perspectives* (pp. 57–70). Mahwah, NJ: Lawrence Erlbaum.

Wenger, E. (1998). Communities of Practice: Learning, meaning and identity. Cambridge, UK: Cambridge University Press.

Whalley, M. (1994). *Learning to be strong: Setting up a neighbourhood service for under-twos and their families*. Sevenoaks, Kent: Hodder & Stoughton.

Wylie, C., & Hodgen, E. (2007). *Competent Learners @ 16: Competency levels and development over time. Report to the Ministry of Education.* Wellington: Ministry of Education Research Division.

# Yes, but how are our kids doing?

**Jeffrey Smith**

*University of Otago*

This question, which is the motivation and touchstone for this paper, was asked at a Ministry of Education meeting at the end of a presentation on the annual results of the National Educational Monitoring Project (NEMP) of New Zealand. The question reflected a desire for a simple answer to the question, one not couched in caveats and context. Although a simple answer might have been desired, it may be the case that one could not be honestly given, that there is no simple answer that sits in isolation, free from comparisons. Ed Koch, peripatetic former mayor of New York City, used to ask people he would meet on the street, "How'm I doing?" Now, usually, Mayor Koch just wanted to be told that he was doing "great". He didn't want to be compared to a phalanx of other mayors of large cities, or be compared to previous mayors of New York City. He wanted an answer about as long as the question. It occurred to me that while we at NEMP are readily able to say, "here is *what* our kids are doing", we have much more difficulty in simply saying, "here is *how* our kids are doing". The "how" implies value judgements, comparisons, and contrasts to other groups and/or other times, and context.

But we get similar questions in all sorts of settings in our lives: How are you doing? How is it going? How did that work out for you? How's the job? How are your folks? Often these questions are mere pleasantries, but when the question is about children, and asked seriously, the answer is more complex. As a parent, I am a veteran of a large number of parent/teacher conferences. I never liked parent/teacher conferences and always approached them with great trepidation. I really wanted to hear only good news, unadulterated adulation for the kids whom I thought were just terrific. But from the perspective of the teacher in these meetings, my child was just one of many, with strengths and weaknesses that needed to be communicated with the best interests of her growth in mind. Thus, the question of how my child was doing had to be placed in some context, some comparison. This context could be the other students in the class, my child's progress since the last meeting, her performance on standardised tests, or some generalised notion of how well the teacher thought she should be doing. Whatever the context, given the propensity of some children to behave differently in school than at home, after hearing my child's performance and behaviour described in terms of enthusiastic compliance to the teacher's wishes, I was often tempted to say, "We're the Smiths, are you sure we are talking about the same child?"

But ultimately, even though the question "How is my child doing?" cannot be answered simply, it should be answered directly. And what can be asked of one child can be asked of all children, thus the question, "Yes, but how are our kids doing?"

## Context and judgement

I have two grandnephews who both just passed their first birthdays. When I talk to their grandfathers, my brothers, I ask how Dashiell and Carson are doing. And they both typically say, "great" (Mayor Koch would be pleased). But for Dashiell, this means that he is beginning to use language, and understanding it better each day; that he is around the 80th percentile in height and weight; and that he can stand if he is holding on to something. He crawls all over the place, and walking won't be too far in the future. For Carson, born at 26 weeks gestation and slightly over one pound, "great" means that he can hold his head up, that he is beginning to eat solid food, that he is past worries about severe visual difficulties, and that we hope he will be on the chart for height and weight adjusted for his scheduled delivery date soon. My brothers both mean "great" when they say it, but the context, which is mutually shared and understood, is different. Context is critical here, as is information that is understandable and pertinent to the situation. Judgement is required in drawing a conclusion and in understanding it, and it is difficult to do this in the absence of background or a framework.

## Looking at performance at the national level

Generalising from two one-year-olds to a nation of children, many of the same arguments hold: we need to understand context; we need information that is direct and pertinent to the situation; and we need judgement. But what kind of answers can we give to the national question: Yes, but how are our kids doing? Here are some possibilities:

- better than before
- not as well as the Norwegians
- pretty close to being there
- as well as can be expected under the circumstances.

Each of these potential responses reflects an underlying set of beliefs about how such a statement should be made. "Better than before" reflects a model based on growth. "Not as well as the Norwegians" reflects a norm-referenced approach with the world (or at least Norway) as the norming group. "Pretty close to being there" is a criterion-referenced statement that implies a shared understanding of where "there" is. And "as well as can be expected under the circumstances" reflects a compensatory model that suggests that a variety of factors related to achievement (e.g., socioeconomic status) need to be considered in making a judgement.

Various combinations and permutations of the statements above are possible, leading to the potential for statements such as, "Well, we're doing better than before, although our growth is still not as strong as the Norwegians, but all in all, we can say that we are doing as well as can be expected under the circumstances, and we are pretty close to being there." As amusing as the foregoing statement may be, it is indicative of the serious need to think carefully about what our aspirations are for our society, and how the education and development of our children play into

those aspirations. Do we want to become a society that has a strong technological base, or one that is more agricultural in nature, tourism-based, or industrial, or some combination of these? What is the relative importance we want to place on the future of our children as opposed to the present for our children? How much do we know about how learning and development occur, and what role formal education plays in that learning and development? Do we want a grand plan, or are we more comfortable with the transition from childhood to adulthood being left in the hands of individuals and their families/whānau? And how do we take into account the idea that different groups of New Zealanders will look at these issues differently?

## Options for making meaning with assessments

As suggested above, there is a variety of options for helping to make assessments meaningful. These options are not mutually exclusive; indeed, they are often used in combination. But each has a distinctiveness about it that is worth considering. We begin with the notion that a score, in and of itself, basically does not have meaning. A score of 68, for example, tells us nothing without some mechanism for providing a meaning to that score. If it is an exam score at the University of Otago, it would be a B, and, by most accords, a fairly good mark. But a score of 68 percent would be a failing score on the New Zealand Driver's Licence examination. Without a basis for making meaning of the score, it is simply a number. Fortunately, there are a number of options available to us for making that meaning.

### Norm referencing

Norm referencing might be a good place to start to better understand how assessment scores acquire meaning. It is important to note at the outset that it is the referencing, or meaning-making system that is normed here, not necessarily the assessment itself. That is, with the same assessment, we might make it meaningful by providing a norm-referenced interpretation of it, or by providing a criterion-referenced, or domain-referenced interpretation. Basically, norm referencing is the process of giving meaning to a score by comparing that score to some appropriate reference, or norming group. Three norm-referenced statements one might encounter in everyday life are:

- the best Italian restaurant in Ranfurly
- an auto make with a below-average rating in reliability
- in the 60th percentile in height for her age.

We understand how well the person (restaurant, auto, etc.) did in comparison to others of the same class. Basically, it is a rank ordering of the performance. The same is true with assessments. Here are three norm-referenced assessment statements:

- Mary is in the fourth stanine in reading.
- Simon's mark is in the 83rd percentile on his PAT.
- Janine was head girl in her high school.

Norm referencing requires a comparison group of some sort. If the score under consideration is for an individual, then an appropriate norm group would be other individuals in the same year, or the same age, etc. If the score under consideration is an aggregate, say for a school, then the

comparison group must be the same type of aggregate. League tables for reporting NCEA scores is an example of such a norm-referenced comparison.

Norms at the individual level are typically established by administering the assessment in question to a large sample of individuals who are believed to be a cohort of some sort—say all Year 5 students in New Zealand. This is called a norming study and is the basis for developing percentiles (the percentage of children who are below a given score), stanines (where 5 is the mean, and 2 is the standard deviation of the scores), and things like IQ scores (where the mean is 100 and the standard deviation is 15).

But what if the aggregate is the nation as a whole? What would be the appropriate comparison or norm group in that instance? Basically, it would have to be other nations. This becomes somewhat problematical: Should New Zealand be compared to Norway, or Australia, or the United States? For many norm-referenced interpretations of a score, the nature of the comparison group is critical for the proper interpretation of the score. Furthermore, such comparisons rely on the relative equivalence of groups being compared. Different countries start children in school at different ages, and have different approaches to curriculum and different goals for formal education. When norm referencing against other countries, the question of whether such comparisons are appropriate becomes of real concern. This problem, at a macro level, is fundamentally the same problem as deciding who to compare my premature grandnephew to. Answering the question "compared to whom?" is critical in norm referencing.

## Criterion referencing

The problem with norm referencing is that it implies that it is important to understand who is better than whom. It is basically a sophisticated approach to rank ordering people along a continuum. It assumes that the rank ordering will bring meaning to an individual score, and that a continuum exists in the domain or trait under consideration that can be appropriately characterised by a single score (more on that later).

In response to the rank-ordering problem, and with the advent of the taxonomy of educational objectives, the criterion-referenced approach to making sense out of assessment scores began in the 1960s. In criterion referencing, scores are given meaning by referring them to an arbitrary standard that has been decided by an expert or group of experts. Here are some examples of criterion referencing in everyday life:

- That fence will keep the sheep in.
- He's a good enough driver to do Skipper's Canyon.
- That salt substitute tastes as good as real salt.

In each of these situations, the outcome is one that meets an essential criterion. Here are some educational examples of criterion-referenced assessment:

- Harold got an Achieve in his NCEA maths assessment.
- Merimeri passed the entrance examination for the police training academy.
- Bill's reading assessment indicates that he is ready to move up a level.

In a classroom setting, a teacher may decide that a student knows enough about a particular aspect of a lesson to move on to the next aspect. That is a criterion-referenced interpretation of the

student's performance. It is also basically formative in nature (that is, it is an assessment *for* learning, not *of* learning). On the New Zealand Driver's Licence examination, the passing score is set at 32 out of 35 questions (this is a summative assessment, an assessment *of* learning). This standard was set by the people in charge of determining who should and should not have a licence. It doesn't matter where an individual ranks in the country, or how their neighbour did, if they get a 32 or better, they have successfully completed the exam. The NCEA examination system is also a good example of criterion-referenced assessment. The exams are written according to a set of specifications and marked according to them as well. The score necessary for an Achieve does not take into consideration how many children will pass or fail based on that standard.

## Domain referencing

A variant of criterion referencing is called domain referencing. In domain-referenced assessment, a domain of knowledge or ability is mapped out that defines, at some level of specificity, all of the things a person should know and be able to do within an area of interest. This might be recall of math facts, such as all single-digit addition problems, or the multiplication tables. It might be the rules for the formation of all regular verbs in French. The idea is that the entire domain of knowledge and skill is made explicit. An assessment is then developed that samples the domain and provides a score on student ability that can be generalised from the assessment to the entire domain, much in the way that a political poll generalises from the sample of people interviewed to the nation as a whole. Thus, we might be able to say that Marisa has mastered 76 percent of her multiplication facts, or that Jack can conjugate all regular verbs in French. Examples of domain referencing in real life are:

- Bill knows how to drive to 80 percent of the locations that taxis in Auckland drive to frequently.
- Ian can make all of the dessert items on the restaurant's menu.

We see here that domain referencing is quite similar to—one might argue a subset of—criterion referencing, but in domain referencing the notion of a fixed standard is not necessary, and the population of behaviours and knowledge is well understood. These are some examples of domain referencing from education:

- Marisa has mastered six of the major scales on her oboe.
- Jack correctly uses most of the rules of comma usage.

Domain referencing requires a knowledge or skill set that is very well defined, or definable. In a New Zealand setting, we might be interested in knowing if students can explain the three articles of the Treaty of Waitangi, or can identify lakes and volcanoes in New Zealand. Domain referencing is a highly useful tool when and where appropriate, but is often somewhat limited in applicability.

## Growth models

A completely different approach to making meaning out of scores is to use growth or change models. In this approach, scores are understood by looking at the difference between the current status and one from some time in the past. We all know growth models from when we were kids

and still growing (for some of us, that process has begun to reverse!). Here are some examples of growth models from real life:

- She grew four inches since her last birthday.
- I lost four kilograms just by exercising more.
- Housing prices are down from last quarter.

It is interesting to note that often with stock market prices or prices on commodities, such as gold or oil, the change in the price is more important than the absolute level of the item. These are some examples of growth models from education:

- She grew 0.4 standard deviations on her asTTle maths score since last year.
- His reading level has gone from Year 5 to almost Year 7.
- Year 8 students had an effect-size gain of +0.23 over students from four years ago on the NEMP science scores.

Growth models can be powerful models. One of the inherent advantages of the use of portfolios is that they allow for the assessment of growth over time. There don't have to be any numbers involved in such an assessment. One can take a writing sample from last year and simply compare it to a writing sample from this year. This would allow one to make statements such as "Look at how much more complex the sentences are, and the more confident use of rich modifiers."

Growth models are also quite useful at the national level, as hinted at in the list of exemplars above. The NEMP model, which holds back roughly 50 percent of tasks from full release pending administration again four years later, allows for a very strong look at growth at the national level, along with changes by gender and ethnicity. Thus, an approach such as NEMP can confidently answer the question "Yes, but are our kids doing better?", which is often a fairly good substitute for the question "Yes, but how are our kids doing?" (It should be noted at this point that I am co-director of NEMP and not an unbiased source of information about the strengths of the project.)

In looking at growth and change, researchers often use the idea of an *effect size*. An effect size is a method for standardising the difference between two means. If we say the difference between two means is 14.5, we don't really know if that is large or small. On a set of scores that ran from 0 to 40, a difference of 14.5 would probably be considered to be quite large, but on scores that ran from 0 to 4000 it would probably be fairly small. So, how can we know how big a difference any given difference is? Well, we simply bring the standard deviation into the mix to standardise the information. We can't get into the nature of standard deviations too much in this paper, but a quick definition is that the standard deviation is a measure of how much a set of scores spreads out around the mean. In a typical normal distribution (bell-shaped curve), if you go one standard deviation above the mean, and then one below it, and look at all the scores in between, they would comprise about 68 percent of all of the scores. That is, 68 percent of the scores would fall between the mean plus one standard deviation and the mean minus one standard deviation. About 96 percent of the scores would fall between the mean plus two standard deviations, and the mean minus two standard deviations. So a total of four standard deviations takes you from near the bottom to near the top.

The effect size is simply the difference between the means (say, the mean in 2004 and the mean in 2008, or the difference between boys and girls) divided by the standard deviation of the scores. In

our example, the difference between the means was 14.5. If the standard deviation were 52.4, then the effect size would be:

Difference between means/standard deviation = 14.5/52.4 = 0.277

In this case, with the scores spanning about 200 points (roughly four times the standard deviation of 52.4), the mean difference seems to be moderate. The effect size is simply how many standard deviations apart the two means are.

If the standard deviation of the scores were 6.2 (the scores are bunched around the mean much more tightly), the effect size would be 14.5/6.2 = 2.34, a very large difference. Effect sizes are comparable across any numbering system. Effect sizes less than around 0.20 are generally, in the social sciences, considered to be small; around 0.50 they are considered to be moderate, and around 0.80 they are considered to be large. Stanines work on much the same model. A stanine is simply a score that has been transformed to have a mean of 5 and a standard deviation of 2, and then has been rounded to a whole number. Thus, a stanine of 8 is one standard deviation more than a stanine of 6 (because 8 – 6 = 2, and then, 2/2 = 1).

Growth models are particularly useful in that they let us compare where we are now to where we were then. Are we gaining or losing? Are we gaining enough? In which areas are particular strengths or weaknesses? As long as we are willing to make comparisons with ourselves, or our "past selves" as the case may be, growth models have a lot to argue for them. At a national level, the use of growth as an indicator of "how our kids are doing" is appealing because we have a good sense of where we have been as a nation, and perhaps a good sense of where we want to go.

Estimates of growth are not always easy to obtain, however, and when growth is substantial there are often technical difficulties in finding a measuring system that will capture the growth. For example, it can be hard to find a system that will measure the performance of students well at the beginning of Year 2 and also at the end of Year 3.

## Item Response Theory (IRT) models

One approach to solving the problem presented in the previous sentence is to use Item Response Theory, or IRT models. IRT models are a family of approaches to analysing test data that use sophisticated statistics, and often some rather strong assumptions about the assessments one is using. The Rasch model is probably the most widely known and used of these models. Here are some real-life examples of statements that take an "IRT approach" to interpretation:

- The Highlanders have about a 30 percent chance of beating the Chiefs this weekend.
- That boat won't survive a storm like the one we had last winter.

In an IRT approach, the object we are trying to assess is compared to a given situation (beating another team, surviving a bad storm). Here are some examples of an IRT approach to educational assessment:

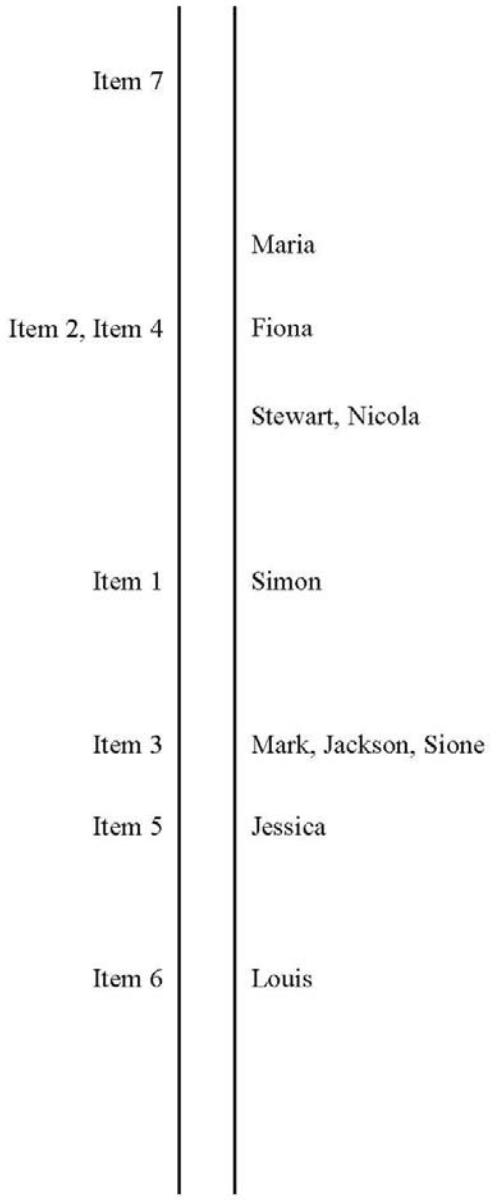
- This is the type of question that Mary has no difficulty with.
- Here we can look at Hamish's growth over the past three years.

IRT models start by placing an assessment item (a test question) on a scale. Each item on the assessment has such a value. Thus we could imagine a scale like a thermometer, and we would be

able to place each item on that thermometer. Some would be higher (the more difficult items) and some would be lower (the easier items) on the scale. The numbering system we use for the thermometer doesn't really matter.

Now, at the same time, by giving our students the assessment and seeing how well they do, we can assign each student a number that will place him or her on the same thermometer scale. Imagine an assessment with seven items that have been scaled using the Rasch model, and 10 students who received the assessment. The results from such an assessment might look like Figure 1.

Figure 1 **An item/person IRT map**



This figure seems to indicate that Item 7 is "harder than Maria is able", and that Item 1 and Simon are somehow the same. It also looks like Item 2 and Item 4 are equally difficult, and that Mark, Jackson, and Sione are equally able. All of these conclusions are exactly correct. In IRT theory, if a person and an item are at the same point on the scale (Simon and Item 1), then the person has a 50 percent chance of getting that item correct. If the item is higher than a person (Item 7 and Maria), then the person has less than a 50 percent chance of getting the item right. If the person is higher than the item, then that person has better than a 50/50 chance on the item. The mathematics involved in IRT are highly complex, but the basic idea is simple. The placement of a person on the scale is compared to the placement of the item, and from that, the probability of the person getting the item right can be estimated.

If Mike joins our class, and we want to assess Mike, we give him some of the items on the assessment and see how he does (we don't have to give him all of them, just some). We might see that Mike gets 7 right out of 10 items. So where should we place Mike on the scale? All we have to do is see what score we would need so that when we figured out the probability of Mike getting each of the 10 items he took correct, the sum of those probabilities would equal his actual obtained score, 7. We just move him up and down the scale until his actual score equals the score our system says he should get.

One aspect of the power of IRT is that, once the system is built, we can link harder items to easier ones (through statistical analysis way beyond the scope of this paper), and then even harder ones, and so on. This allows us to build a vertical scale that lets us put Year 2 students (taking easier items) on the same scale as Year 7 students (taking much harder items). And this lets us look at the issue of growth over a number of years. The perceptive reader may be saying, "this sounds a lot like asTTle". Indeed, asTTle is based on the Rasch model of test analysis. IRT models are very flexible and highly useful. They have their critics as well. For example, some researchers argue that a single construct like mathematics doesn't grow in such a fashion that you can measure Year 2 and Year 9 students on the same scale. Also, there is some limitation on the nature of the items that are useful for IRT models. It is a powerful tool nonetheless, and one that is growing in utility almost daily.

## Inherently communicable assessments

The final stop on our list of possible ways to have assessments have meaning is to think of assessments that are simply inherently communicable. They are assessments that are understandable by their very nature. This may be because they are things we are very familiar with, or that make good sense to us. Here are some examples from everyday life:

- Jack is a good listener.
- Our son is potty trained.
- Susan's temperature is 39ºC.

When we hear these things, we don't really need further explanation. There is a common understanding of what is being said. We don't need to refer to other people, a fixed criterion, or a previous status. Here are some examples that relate to educational attainment:

- Jack can read and understand the *Listener*.
- Our son can recite the alphabet.
- Susan can explain how an internal combustion engine works.

What these statements share is their "understandability". Some of them are milestones or benchmarks that are readily recognisable (potty training, alphabet); others are more endpoints in a process (being a good listener, being able to read the *Listener*).

In an age of increasing specialisation and technical detail, we should strive to present information about how our children are doing in terms that the general public can readily understand, whenever possible. In reading, this might be letting the public know what children can do in everyday terms. It is easy enough to report to the public when children know all of their letters (and can print them), when children can read simple stories, when they can read newspaper stories about topics of interest to them, and when they can read somewhat more sophisticated writing such as is found in the *Listener*. This would represent a series of benchmarks that are, in my mind, highly communicable.

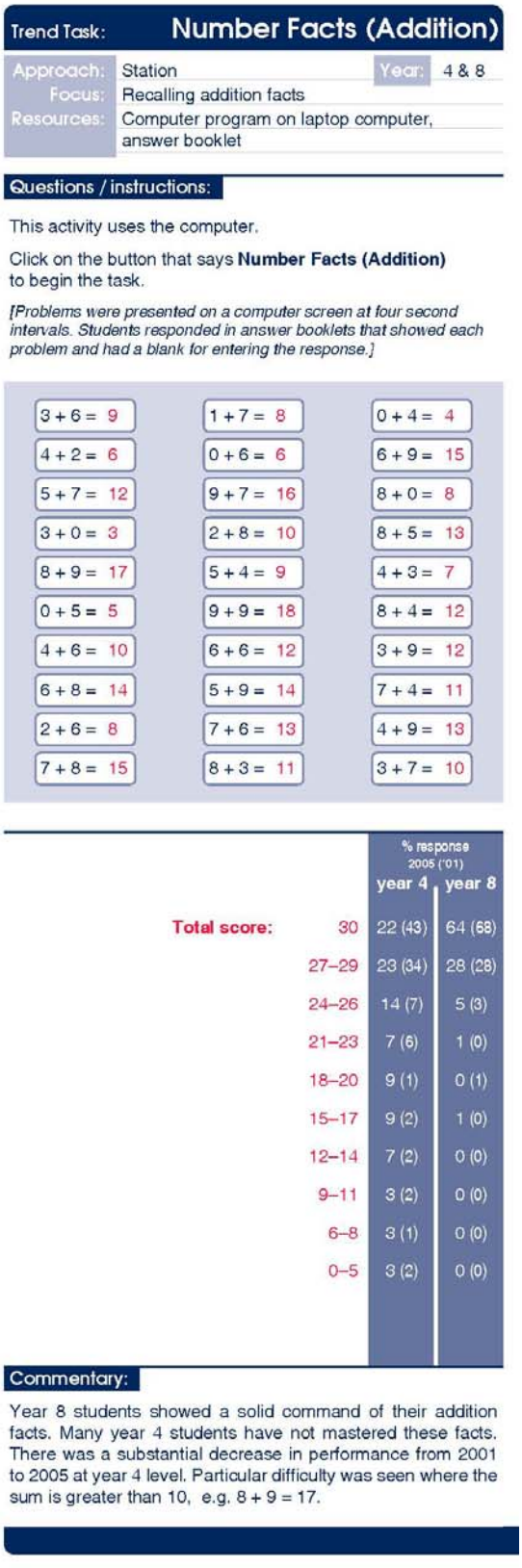
Another example might come from physical education. There is great concern in the nation about children being overweight and poor in athletic skills. We don't really have a good mechanism for reporting to the public in terms they can understand just how serious the problem is (if, indeed, it is serious). But that should not be too hard a problem to solve. We can estimate body mass from height and weight; we can actually get a fairly good idea of whether a person is a little bit overweight, or a lot, simply by looking at them. We know what a press-up or sit-up is, and we can find out how many a child can do, or how fast they can run 40 metres, or throw a cricket ball. These are indicators of the state of affairs related to fitness that are not difficult to collect, and are readily communicable to the public.

The idea here revolves around the notion that we educate children in order that they may successfully participate in life, and that the public has a decent idea of what that means. So can we find measures in all areas of schooling that we can present to the public without having to put too much context on them for the purposes of understanding? This is not always possible (we want some children to become great at mathematics, and most of the public is quite a long way from there), but we should seek such measures out wherever we can find them.

## An example that cuts across various models

As mentioned earlier, it is not the assessment per se that is norm referenced or criterion referenced, it is how it is used or interpreted. An example of a NEMP task illustrates this point well. Consider the NEMP task shown in Figure 2.

Figure 2 **A NEMP mathematics task**



**Trend Task: Number Facts (Addition)**

| | | | |
|---|---|---|---|
| Approach: | Station | Year: | 4 & 8 |
| Focus: | Recalling addition facts | | |
| Resources: | Computer program on laptop computer, answer booklet | | |

**Questions / instructions:**

This activity uses the computer.

Click on the button that says **Number Facts (Addition)** to begin the task.

*[Problems were presented on a computer screen at four second intervals. Students responded in answer booklets that showed each problem and had a blank for entering the response.]*

| | | |
|---|---|---|
| 3 + 6 = 9 | 1 + 7 = 8 | 0 + 4 = 4 |
| 4 + 2 = 6 | 0 + 6 = 6 | 6 + 9 = 15 |
| 5 + 7 = 12 | 9 + 7 = 16 | 8 + 0 = 8 |
| 3 + 0 = 3 | 2 + 8 = 10 | 8 + 5 = 13 |
| 8 + 9 = 17 | 5 + 4 = 9 | 4 + 3 = 7 |
| 0 + 5 = 5 | 9 + 9 = 18 | 8 + 4 = 12 |
| 4 + 6 = 10 | 6 + 6 = 12 | 3 + 9 = 12 |
| 6 + 8 = 14 | 5 + 9 = 14 | 7 + 4 = 11 |
| 2 + 6 = 8 | 7 + 6 = 13 | 4 + 9 = 13 |
| 7 + 8 = 15 | 8 + 3 = 11 | 3 + 7 = 10 |

| | | % response 2005 ('01) year 4 | year 8 |
|---|---|---|---|
| **Total score:** | 30 | 22 (43) | 64 (68) |
| | 27–29 | 23 (34) | 28 (28) |
| | 24–26 | 14 (7) | 5 (3) |
| | 21–23 | 7 (6) | 1 (0) |
| | 18–20 | 9 (1) | 0 (1) |
| | 15–17 | 9 (2) | 1 (0) |
| | 12–14 | 7 (2) | 0 (0) |
| | 9–11 | 3 (2) | 0 (0) |
| | 6–8 | 3 (1) | 0 (0) |
| | 0–5 | 3 (2) | 0 (0) |

**Commentary:**

Year 8 students showed a solid command of their addition facts. Many year 4 students have not mastered these facts. There was a substantial decrease in performance from 2001 to 2005 at year 4 level. Particular difficulty was seen where the sum is greater than 10, e.g. 8 + 9 = 17.

This is a fairly straightforward set of 30 addition facts that had been given to a random sample of Year 4 and Year 8 students as part of a NEMP assessment. Since the sample was nationally representative of New Zealand school children, it provides normative data. If a student new to a Year 4 class were given these items, and got 28 correct, a teacher could say, "Well, Simon is roughly just below the top third of the nation" (22 percent got them all correct, and Simon was in the middle of the next 23 percent, so, with a bit of interpolation, we can say that roughly a third of the children did better than he did). This would give the teacher a fairly good idea of where Simon stood nationally, and if they knew their current students well, they might have a good idea of where he was compared to them also. A principal could give all of the Year 4 students the same tasks to find out about how well the school was doing in this area. This is norm referenced because the scores are given meaning by referring them to a norm group.

The teacher might take a slightly different approach, however. They might say, "I want students to really have their addition facts down cold. No room for mistakes here." In that case, they might conclude that Simon was getting close to the mark, but that he needed a bit more help. This would be the criterion-referenced interpretation of Simon's score. It is given meaning in terms of how it relates to the teacher's required standard.

Now, there are more single-digit addition facts than are contained in this little assessment. For example, 3 + 2 is not among the items. But the set represents a reasonable sample from the "domain" of single-digit addition facts. If Simon got 28 out of 30 correct, we might reasonably conclude that he knows 93 percent (28/30) of his single-digit addition facts. This would be a domain-referenced interpretation of the score.

At a national level, these data provide interesting information about growth. We have information on how well students did at Year 4 and at Year 8. Since we have random samples in both instances, we can draw the inference that the gains we see here are reflective of national trends. It is easy to see that although many students at Year 4 are experiencing some difficulty in math facts, by Year 8 about two-thirds of students are getting them all correct. Additionally, the scores provided not only cover 2005, when the assessment was given, but also 2001, when the exact same assessment was given the first time. Since the assessment was given to random samples of students in both years, we can look at growth from 2001 to 2005. The picture at Year 4 is not very encouraging, as performance seems to have dropped quite a bit from 2001. It should be noted that increases were seen in other areas of the assessment for Year 4 students.

Finally, we can look at this assessment as an example of one that is, to a large degree, inherently meaningful. That is, people generally understanding what addition facts are all about. They did them in school and use them all the time in real life. Thus, when they see results such as these, they have a good sense of what is being discussed and what good, or less than good, performance looks like. It doesn't require a lot of explanation and interpretation.

This example was not analysed using IRT theory so, unfortunately, I cannot extend the example to include that approach!

## Bringing these ideas together

Where does this leave us? We return to where we began, trying to answer the question "Yes, but how are our kids doing?" From my perspective, the approach to answering this question is two-fold. First, the questioner has to accept the notion that "great" simply isn't going to suffice as an

answer, as attractive as it may be. There are just too many Dashiells and Carsons in the world to be able to give answers without context, interpretation, and judgement. On the other hand, at some level, everything is explainable down to the point of being meaningless. We need to strike a happy medium. In my mind, such a medium would include a process that might look something like this:

1. We need to discuss and agree upon what successful schooling and development would look like if it were occurring. What do we think childhood should be, both in terms of the experience of it, as well as what children should be like as they enter their early adult years? To a degree, these ideas might be represented in the new curriculum—or not, but that is where I would look for them first. It's hard to know if we are getting closer to our destination if we don't have a good idea of where we want to go.

2. We should look hard at what some highly communicable benchmarks might be on the way to becoming a successful (happy, productive, good citizen, etc.) young adult. Can we find some inherently communicable milestones that we can report to the public on a regular basis, and that are easy for them to understand?

3. We need assessments (communicable or otherwise) that would let us know if children are getting where we want them to be, so that educational professionals, and others deeply concerned about educational issues, can have strong, valid indicators of progress. Such measures, and the ways that we give meaning to them, might be complicated, and require a degree of explanation, but that does not mean that we should necessarily shy away from such approaches.

4. We need to better understand the progression of children as they move through the school years into adulthood. We know that not all children progress at the same rate, and that there are starts and stops along the way, but we don't yet know enough to know when to be really concerned about children, and when to give them more time to have a growth spurt and catch up.

5. We need to compare our progress in a normative fashion with other countries, because we live in a competitive world that grows smaller and more competitive every day. This is not meant to laud competition over co-operation, but to say that our children need to be successful in a world where jobs can go overseas in a heartbeat.

6. We need to remember that the quality of our lives is not measured solely, or even primarily, by the dollars that we earn or the cars that we drive. We need to remember that our kids are also doing well when they are enjoying this beautiful country, going to an art museum or a rugby match, and looking out for one another.

# Standardised testing: Dilemmas and possibilities

**Charles Darr and Sue McDowall**

*New Zealand Council for Educational Research*

## Introduction

There are dilemmas and possibilities associated with standardised tests, particularly when the tests are used on a national basis and to report against national norms. The dilemmas are political, philosophical, and, to some extent, ethical. Many of them are well known, while others are more obscure. The possibilities, on the other hand, are exciting, particularly when new techniques and technologies are available to make the tests more useful for teaching and learning.

In 2003, NZCER began a process of redeveloping the Progressive Achievement Tests (PATs). We published a new edition of PAT:Mathematics in 2006 and a new edition of PAT:Reading (Comprehension and Vocabulary) in 2008. For the team involved, these redevelopments provided an opportunity to reflect on the dilemmas and possibilities associated with standardised tests.

This paper begins by looking at different perspectives on standardised testing. It then identifies some of the dilemmas involved with testing and considers how the project team responded to these and at the same time worked to produce new possibilities.[1]

## Perspectives on testing

### Testing as tyranny

In 1964, Banesh Hoffman, a respected mathematician and author, wrote a book called *The Tyranny of Testing* (Hoffman, 1964). As the title suggests, Hoffman was deeply upset by the use of testing in education. Among other things, Hoffman argued that the questions contained in

[1] We have drawn some of the material presented in this paper from previous presentations developed by the PAT:Reading team: Charles Darr, Sue McDowall, Hilary Ferral, Juliet Twist, and Verena Watson.

standardised tests, particularly those used in multichoice tests, did not recognise students who were original and deep thinkers.

This kind of viewpoint could be called *testing as tyranny*. From this position most, if not all, testing is seen as tyrannical. Testing and tests are implicated in a "catalogue of crimes". For instance, tests are accused of:

- cultural and gender bias
- damaging students' self-esteem
- having too narrow assessment focuses
- suffering from stereotype threats[2]
- promoting negative effects and behaviours when the tests are used to inform high-stakes decision making[3]
- giving test developers power over the curriculum
- resulting in "teaching for the test".

## Testing as a magic bullet

The testing as tyranny viewpoint is, of course, not the only viewpoint. On the other side of the divide, testing is presented as the antidote to an underperforming education system that favours craft knowledge over evidence-based practice. We could call this point of view *testing as a magic bullet*.

If Banesh Hoffman's book was an example of the first viewpoint, an article from *The Dominion* newspaper written in 2001 (Hunt, 2001) might be an example of the magic bullet perspective. In the article, Geoff Hunt, a member of the Business Roundtable, argues that standardised testing should be implemented in all New Zealand classrooms. Such a scheme, he claims, would identify students in need of remediation, provide principals with a rigorous tool for managing teacher performance, give boards of trustees a key focus for measuring their principal's performance, and allow the Ministry of Education to identify poorly performing schools and measure student improvement.

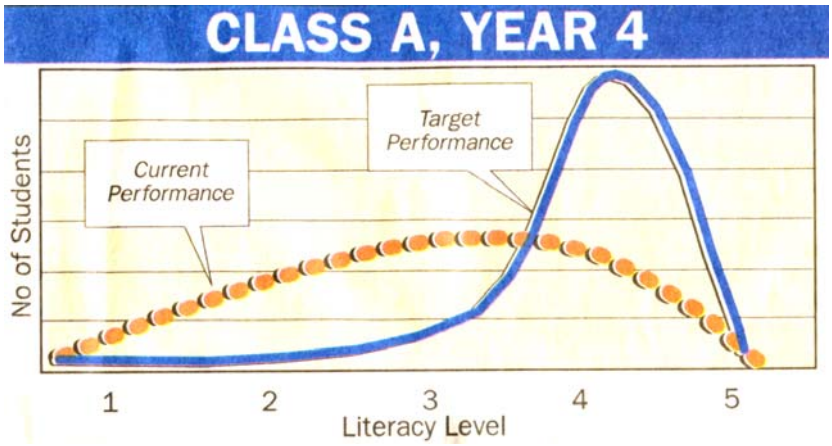
According to Hunt, this approach to performance improvement is included within the principles of what he calls "Total Quality Management". Performance targets would be stated and reported graphically, as in the chart in Figure 1.

---

2 The term *stereotype threat* relates to achievement effects manifested when individuals are told that their performance on the test is related to their membership of particular groups with which they identify, for instance an ethnic group or gender (Aronson et al., 1999).

3 When tests determine high-stakes outcomes they are often associated with perverse effects. For instance, a study of Chicago schools found that approximately 5 percent of teachers doctored answer sheets when tests were used to rate teacher performance (Jacob & Levitt, 2004).

Figure 1 **Suggested reporting of performance targets**




Hunt notes that such an approach to performance management in schools "… would be similar to measuring the performance of a branch operation where the branch manager was equivalent to the teacher" (p. 11).

From the magic bullet perspective, testing is seen as a way of presenting the hard data needed to inform a scientific approach to the problem of education. Business and medical metaphors tend to dominate this discourse.

Another example of this perspective can be seen in the Bush administration's policy of "No child left behind".

## Testing as tool use

Others take what would appear to be a more balanced view regarding testing, exemplified by this quote from a textbook published in 1978 on measurement and evaluation in schools:

> The standardized test is only an educational tool to be used along with other educational techniques. (Karmel & Karmel, 1978, p. 10)

Describing tests as "assessment tools" has become very popular in recent times. It suggests that standardised testing is one of many techniques that can be used to assess students. We can call this perspective *testing as tool use*.

Is this "middle way" the appropriate way to consider testing? Can we really take advantage of what testing has to offer, while avoiding the pitfalls?

As an aside, it is a surprise to find tests described as tools in a book that was published in 1978. Generally in books of this era even the word *assessment* is used infrequently. Instead, books use words such as *testing*, *measurement*, and sometimes *evaluation*. The tests themselves are sometimes referred to as *instruments*.

In today's educational context the term *assessment* encompasses a greater range of activities and concerns than could ever be signalled by words such as *measurement* and *evaluation*. However, the discursive shift from *measurement* to *assessment* and *instrument* to *tool* can also be read as symptomatic of a move to lessen the association of assessment with psychology and scientism.

Historically, standardised testing came out of a move to make the study and measurement of cognitive attributes scientific.

It is interesting, though, to look at how standardised tests have been used in our own history. Recently, I was helping the librarians at NZCER to "weed" a selection of older books on tests and test design. I had to stop and take a second look when I came across one that was entitled *Educating Backward Children in New Zealand* (Winterbourne, 1944). What an amazing title! This book was published in 1944 by the New Zealand Council for Educational Research. What does it have to do with testing? Some of the data used in this book come directly from a national survey of intelligence for 10- and 11-year-olds. The survey involved around eleven thousand 10- and 11-year-olds who were administered the Otis Self-Administering Test of Mental Ability.

The results are interesting. By using cut scores, the author has defined two groups of students. The first had IQ scores below 70 and are designated as "feebleminded". The second have IQs in the range 70 to 85 and are designated as "dull or border-line feebleminded". Table 1 below shows the results of the survey.

Table 1 **Incidence of "feeblemindedness" by gender**

| Gender | Number tested | Feebleminded (%) | Border-zone and dull (%) |
|---|---|---|---|
| Male | 5820 | 2.7 | 14.8 |
| Female | 5558 | 1.4 | 10.6 |

Using tests like this to categorise people and define intelligence fuels much of the perspective we have called testing as tyranny. What is more, this kind of history or tradition of testing doesn't disappear just because a few words are changed—it lingers, even in the language we use. For instance, it is interesting that the word *tool* incorporates the essence of the word *instrument*.

## The "tests as tool" metaphor

Let us return to the testing as tool use perspective. At its heart is the metaphor of "tests as tools". When we usually talk about tools, we are most commonly referring to physical implements, and, in particular, ones used by people who are involved in trades or crafts. A mechanic or a builder, for instance, will usually have access to a large range of tools. These days, of course, tools don't have to be associated with physical work. Calculators and computer software are often described as "tools", as we recognise the potential for these things to give us an advantage when it comes to doing mental or abstract work. Wikipedia picks up on this extended meaning when it defines a tool as: "a device that provides a mechanical or mental advantage in accomplishing a task" ("Tool", 2008).

It is probably not too far a shift to describe the things we do or use to carry out assessment as tools. Our assessment tools certainly do provide us with a mental advantage. They give us the means to collect and process data in an efficient and systematic way.

From a perspective of possibilities, the idea of the teacher or assessor as a "tool user" can be quite powerful. It evokes that idea of a skilled craftsperson or professional choosing and applying what is needed to solve a problem. But does the assessment as tool metaphor have its blind spots?

## Autonomous technology

In our "everyday" or common sense ways of understanding, tools are purely instrumental—that is, they are simply a means to an end. They make our lives easier, but are themselves intrinsically neutral. They are things we control—rather than things that control us. But are they?

Some philosophers who enquire into tool use or, more precisely, technology, argue against this. Langdon Winner, who writes about the political consequences of technology, certainly questions this assumption. In an article entitled "Do artifacts have politics?" (Winner, 1980, p. 125) he notes that this instrumental way of thinking, "do[es] not stop to inquire whether a given device might have been designed and built in such a way that it produces a set of consequences logically and temporally *prior* [italics original] to any of its professed uses".

Winner provides several examples of tools or technologies that affect us in ways well beyond their intended use. Perhaps the most poignant example concerns a system of overpasses that regularly cross the landscaped highways of Long Island, New York. Built to careful specifications, these bridges are remarkably low, sometimes only clearing the curb by nine feet.

When they were built, the overpasses represented a real innovation; as a new piece of road technology they did away with the need for multiple intersections. However, this is not the extent of the reasoning behind them. Their design reflects the racial and social biases of their builder, city designer Robert Moses. His arrangement meant that while cars owned by what he referred to as "whites of the upper and comfortable middle classes" would be able to use their cars on the new highways, poor people, and particularly blacks, who predominantly relied on public transport, would be kept off the roads and therefore from the beaches and public parks the roads provided access to. This was because the 12-foot-tall public buses could not get through the overpasses. One of the particular results of Moses's innovation was that access to an acclaimed Park, Jones Beach, was severely limited for racial minorities and low-income groups.

In this case, engineered into the overpasses was a systematic way of promoting certain social relationships. Although the overpasses themselves are apparently neutral, and certainly on the surface simply a tool or technology for facilitating the movement of people, they perpetuate a systematic social inequality:

> Consciously or not, deliberately or inadvertently, societies choose structures for technologies that influence how people are going to work, communicate, travel, consume and so forth over a very long time. In the process … different people are differently situated and possess unequal degrees of power as well as unequal levels of awareness. (Winner, 1980, p. 127)

According to Winner, all tools or techniques have these structuring effects. By consciously or unconsciously promoting or defining particular ways of interacting with the world and with others, people are variously positioned in relation to power. While some of this structuring is obvious, at times some of these structuring effects are hard to observe—they look natural or at least become natural—like Moses's overpasses they are simply part of the landscape.

Like any tool, standardised testing has structuring effects, side effects, and unintended consequences. Some of these have already been mentioned as dilemmas. Here are two more.

## The construction of subject areas

Although it is not necessarily intended, tests construct what it means to know a subject. Take mathematics for instance. By selecting material to assess and using particular question types, a standardised test presents a version of mathematics. Often this is a mathematics where:

- one correct answer is favoured over multiple perspectives
- speed and recall is more to the forefront than exploration and inquiry
- application of learnt procedures dominates over problem solving
- mathematics is presented as abstract and solitary rather than connected, embodied, and social.

Mathematics does not have to be presented like this. Mathematics is a human endeavour, strongly rooted in the way humans experience the world.

Portraying mathematics in a particular way influences the relationship learners have with the subject matter. Some argue that our failures in mathematics education relate to a failure to help students develop an appropriate identity as mathematicians.

In developing the new PAT:Mathematics test we tried hard to confront this dilemma. For instance, we worked hard to include questions that required thought and application of mathematical principles, rather than just the use of learnt processes. We had to accept, however, that there are definite limits to how mathematics can be presented using a paper-and-pencil-based assessment format.

The same is true for reading comprehension. If standardised tests were all we used, we'd have a limited view of reading as something which occurs:

- in the English language (and of a particular form)
- with texts from the "canon" of academic disciplines
- in print (as opposed to online)
- in isolation from social action and connection.

There is the risk that reading would be seen as:

- looking for the one "correct" interpretation (as opposed to being aware of a range of possible readings/interpretations)
- using printed words in isolation from the visual and aural
- examining text content (rather than construction)
- finding information and inferring (but not examining the points of view represented and silenced by particular texts).

There are two main reasons reading comprehension is constructed in these ways. One is to do with conceptions of what reading is. And this is certainly not fixed. The second is to do with the limits and possibilities afforded by standardised tests.

Our challenge was to work with the constraints and harness the possibilities of standardised testing to create a test that came as close as possible to reflecting our beliefs about what it means to be a reader. So how did we deal with these limits when redeveloping PAT:Reading?

We began by looking at the features of the old PAT:Reading and thinking about what we were prepared to keep and what we wanted to change. Our first dilemma related to the multiple-choice format.

In our view the best way to assess comprehension is through conversation around text. And the next best way is through very broad, open-ended questions. And so we considered moving away from the traditional PAT multiple-choice format. But our brief was to develop a standardised and easy-to-mark test, and the sort of open-ended questions you can develop within these parameters are narrow—as narrow or even narrower, we came to believe, than what could be elicited through cleverly constructed multiple-choice questions. So we went with multiple-choice.

You may recall that the original PAT:Reading provides students with five possible answers to choose from. It is difficult for students, especially for younger students, to carry in mind the question and five possible answers while also negotiating their way through the text. Further, the ability or lack of ability to do so is not necessarily related to comprehension ability. For this reason we decided to reduce the options to four, as a way of decreasing the cognitive load.

Our next dilemma related to ideas about text difficulty. In the old PAT:Reading the texts get increasingly difficult as determined by a measure of word frequency. People often assume that the challenge of reading resides in the text and that reading growth involves making meaning from increasingly difficult texts, as determined by readability measures based on word frequency and sentence length. Margaret Forster (2003) shows that, in fact, text difficulty does not help to explain the difficulty of reading items, demonstrating that while both weak and strong readers can correctly answer simple questions of complex text, only strong readers can correctly answer complex questions of simple text. Rather, she argues, it is the *relationship* between text and task that determines the difficulty of items. This thinking informed our use of readability measures. We were careful not to include texts at particular year levels that would be much too difficult for students to read, because our purpose was to test comprehension, not decoding. But we were comfortable about including some texts that would be considered easy, as determined by readability measures, if we judged their content and the questions asked of them to be suitably complex, and if our trialling data confirmed this.

Our third dilemma was about the sorts of questions we asked. The old PAT:Reading categorised questions as inference or factual—a seemingly clear-cut division. But we found that nearly all questions worth asking require at least some level of inference. These are those that do not involve direct retrieval of information—the type of questions I call *hunt and peck*.

Here is an example of a hunt and peck question.

"Last Saturday, Steve spent all day working at home. **Straight after breakfast** he **helped his father wash the car**. After lunch he mowed the lawns, weeded the garden and swept the paths…"

**Q** What did Steve do **straight after breakfast**?

**A** Mowed the lawns.
**B** Swept the paths.
**C** **Helped wash the car.**
**D** Brought in some firewood.

There are two problems with these sorts of questions. One is that it is possible to answer them correctly without fully comprehending the text being read. You do this by finding words in the text which match words in the question—in this example, "straight after breakfast"—and then matching the words in the text with the words in the answer—in this case, "helped wash the car". For this reason such questions are not necessarily a good measure of comprehension.

The other problem is the "who cares?" factor. Often hunt and peck questions focus in on information that is not related to the main idea or plot of the text. They are the details of a text that readers do not often remember or think about, because they are not important. In this story about Steve it is unlikely that what Steve did straight after breakfast is going to be an important or memorable part of the story.

Now there is a place for retrieval questions. One of our texts is a recipe and it would be a worry if using recipes required inference. The questions we ask of our procedural texts are retrieval questions. But with other texts we have tried to avoid the type of questions illustrated above. We have focused instead on developing questions requiring students to infer.

We categorised the inference questions into two types: local inference and global inference. We define local inference questions as those requiring comprehension of implied information from specific parts of the text. Examples include questions such as "Why did Glen shrug?", which involves inferring a character's intention from implied information across six sentences, or "What does the poet compare herself with in the last verse?", which requires unpacking a simile. We define global inference questions as those which require comprehension of implied information across the text as a whole. Examples include questions such as: "What is the main purpose of this text?" or "What impression of the brain does this text create?"

We believe that all students of all ages and achievement levels need opportunities to demonstrate and practise their abilities to infer. Not all people hold this view of reading comprehension. Some people believe showing achievement in reading is only about retrieving information from increasingly complex texts. This was brought home to us at the end of last year when we collected

the tests from the printer. He thought there was a mistake because there was no answer in the text which exactly matched the wording of the first question he read. He was expecting all the questions to be like the example I've just shown you. This expectation may have come from his own experiences of sitting standardised tests.

Historically, inference has not always been a big part of reading fiction. This is often illustrated by comparing the murder scenes in *Oliver Twist* written in 1838 (Dickens, 1838/1992) and *Tess of the d'Urbervilles* (Hardy, 1891/2007), written over 50 years later. In *Oliver Twist*, Bill Sykes' murder of Nancy is described in great detail. Dickens gives us a blow-by-blow account. He does the work. The reader is passive. There is nothing to infer because it's all there on the page.

When Tess murders d'Urberville the reader needs to do the work to create their own image. Hardy sets us up to do this. The first clue is when the landlady looks up from her sewing to see a damp patch in the ceiling that she hadn't noticed before. As she watches it slowly spreads. She stands on a chair, reaches up to touch the ceiling and finds she has blood on her hand. She goes upstairs and listens outside the door of the room above hers. At this point Hardy (1891/2007, p. 466) provides just three words: "Drip, drip, drip." When the landlady eventually enters the room all we are told is that the carving knife, which she had seen on the table earlier that day, is gone. The reader is active and is required to put something of him or herself into the reading. This is what makes reading engaging.

We found that our capacity to write complex inferential questions was highly text dependent. Texts with dialogue were particularly good for asking inferential questions. So were poems. This is because ideas are shown, not told.

This leads me to our last important decision—text topic and type. One of the things we wanted to model in the new PAT:Reading was a broad range of text topics and types. Some people think it is important to provide students with topics they are already familiar with. So, for example, if you have a group of students keen on skateboards you provide texts on this topic. Or if your school has a high Somali population you show cognisance of this when selecting texts.

All students must be able to see themselves, as well as other groups, in texts. But reading is also about broadening horizons beyond the parameters of your current world. It is about engaging with new ideas and new ways of using language. I think we are doing students a disservice if we only expose them to texts which reflect their world. Also, I would question our capacity to really know what those worlds are. Texts of any topic can be engaging if they require the reader to infer and if students are given the necessary support to do so.

## The construction of norms

So far we have been looking at how tests structure what it means to know and achieve in a subject area, and four dilemmas related to this in the context of comprehension. Tests as assessment tools have another structuring effect. In measuring students they construct norms.

We talk about assessment being like a "snapshot" because it captures achievement at a particular moment. And it's easy to forget that's all it is—how a student performed with a particular task at one point in time. Sometimes the marks we give students can take on a life of their own. The meaning of a grade can shift to describe, not the work, but the student. You can see this when students are referred to as their grade: "She is an 'A' student", or, "He is level three."

Gallagher describes this as a process by which we turn children into subjects "in every sense of that word" (Gallagher, 1992, cited in Davis, 1996, p. 255). Davis describes it as "a naming, an objectification of one's fluid self" (Davis, 1996, p. 247). The danger of this is that it can prevent teachers from observing and listening to additional and subsequent evidence of students' capabilities. And as Davis says, "the artifacts of our assessments should provide *occasions* [italics added] for our listening, not obstructions" (1996, p. 240).

## A new approach to measurement

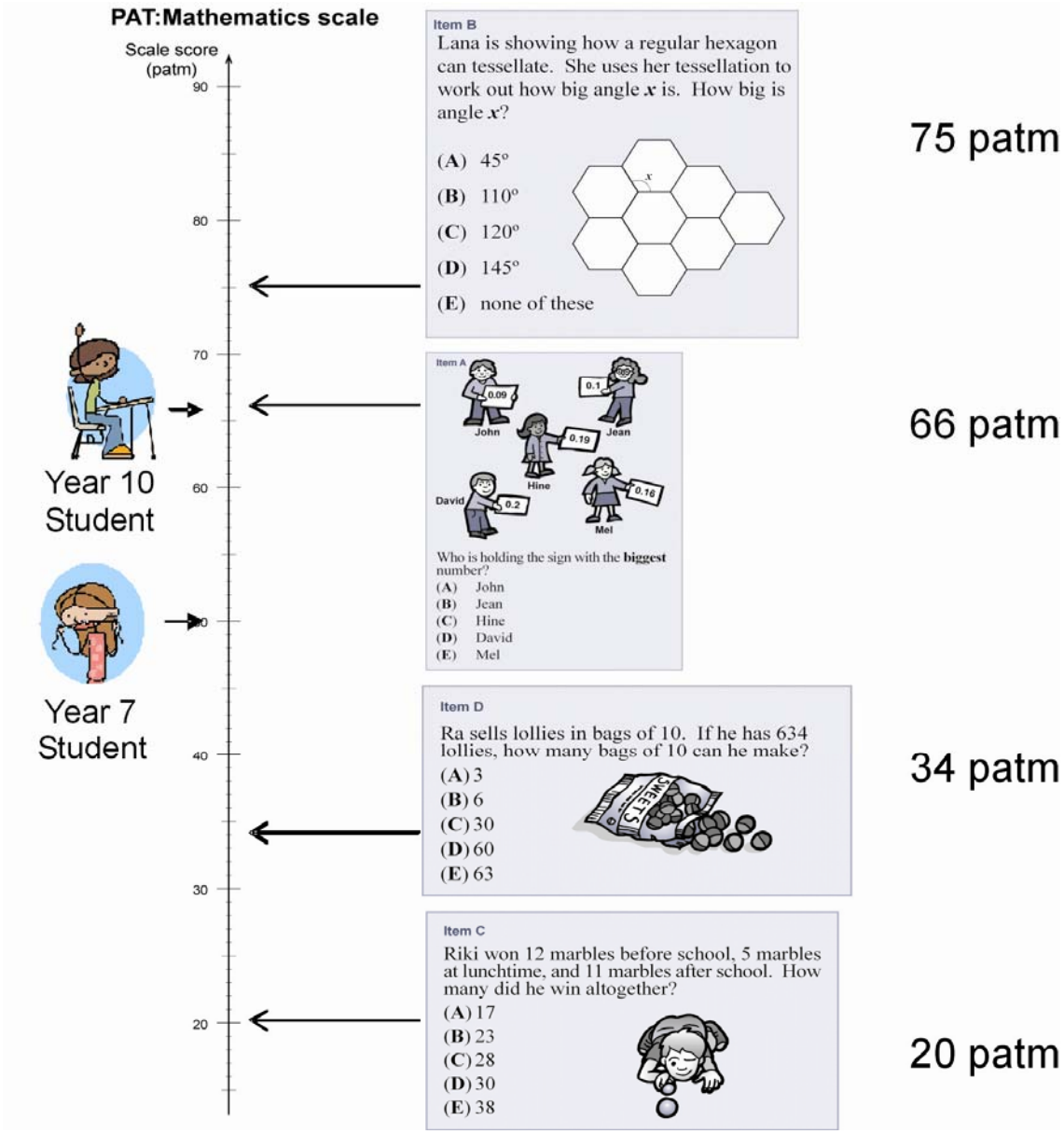
In New Zealand, achievement on standardised tests has traditionally been reported almost exclusively in terms of student norms. This type of reporting ranks a student's achievement against the achievement of a group of peers; for instance, against a national sample of students in the same year level. The new PATs, however, employ an approach to measurement that does not only rely on reporting achievement against norms.

At the heart of each new PAT (Mathematics; Comprehension and Vocabulary) is a Rasch Measurement scale. Rasch Measurement (RM) was developed by the Danish mathematician Georg Rasch in the 1950s (Rasch, 1960/1980) and has been used in the construction of interval scales in education and other fields since the 1960s. An RM scale allows us to locate the relative achievement of students and the relative difficulty of questions on the same interval scale. Each location on the scale is said to represent a certain amount of the attribute (such as reading comprehension) being measured. In constructing the scale, RM makes no assumptions about the distribution of student achievement. A student's location on an RM scale is a measurement of how much skill and knowledge has been demonstrated by the performance on a test and does not depend on who else has been tested or what particular questions have been used to do the testing.

Raw scores on any of the new PAT tests can be converted to locations on the appropriate RM scale. The conversion takes into account the difficulty of the questions used in the test. The use of a common scale for each subject area allows student progress to be tracked from test to test. It also allows teachers to choose the level of test appropriate for their students.

Figure 2 shows a section of the PAT:Mathematics RM scale, along with the locations of a selection of questions from the PAT:Mathematics tests. The portion of the scale shown is numbered from 20 to 90 PAT:Mathematics units (patm units). These numbers are arbitrary, and represent increasing levels of knowledge and skill. Each unit, however, represents the same amount of change. An increase from 70 to 80 patm units, for example, represents the same amount of change as an increase from 20 to 30 patm units. Also shown in the figure is the mean location on the scale for Year 7 and Year 10 students.

Figure 2 **The PAT:Mathematics scale with sample question locations**

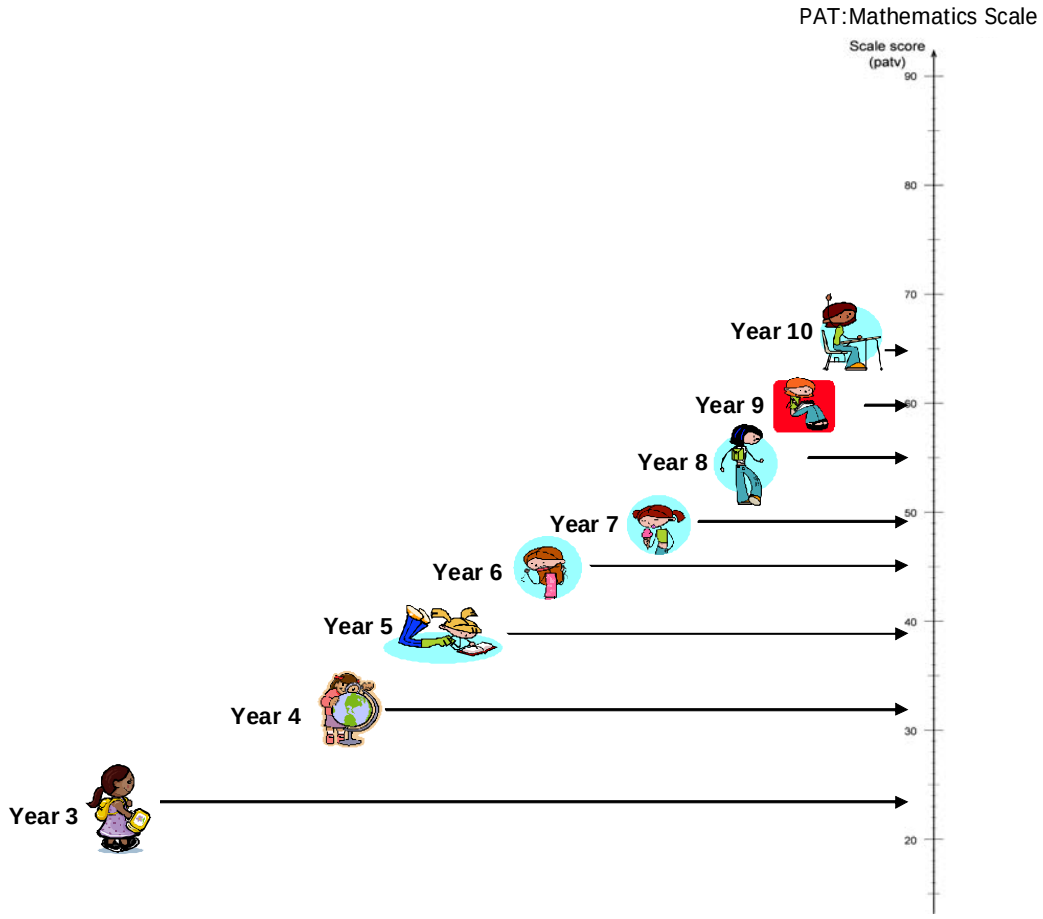



Once a student's scale location is known, comparisons can be made between the level of a student's achievement on the scale and questions that are located above, at, and below this location. Typically, students are more likely to answer questions below their scale location correctly and less likely to answer questions above their scale location correctly. Students will generally answer correctly about half of the questions that are located at the same level as their own achievement. In Figure 2 the "average" Year 7 student would be likely to find Item C straightforward, and Item B quite difficult.

Student norms are still available and are useful when interpreting progress on the scale. Figure 3 shows the mean achievement level for national reference groups in mathematics from Year 3 to

Year 10. Once a student's achievement on the scale is known it becomes possible to compare their achievement with norms at multiple year levels.

Figure 3 **"Normal" progression on PAT:Mathematics**




## Reporting on the new PATs

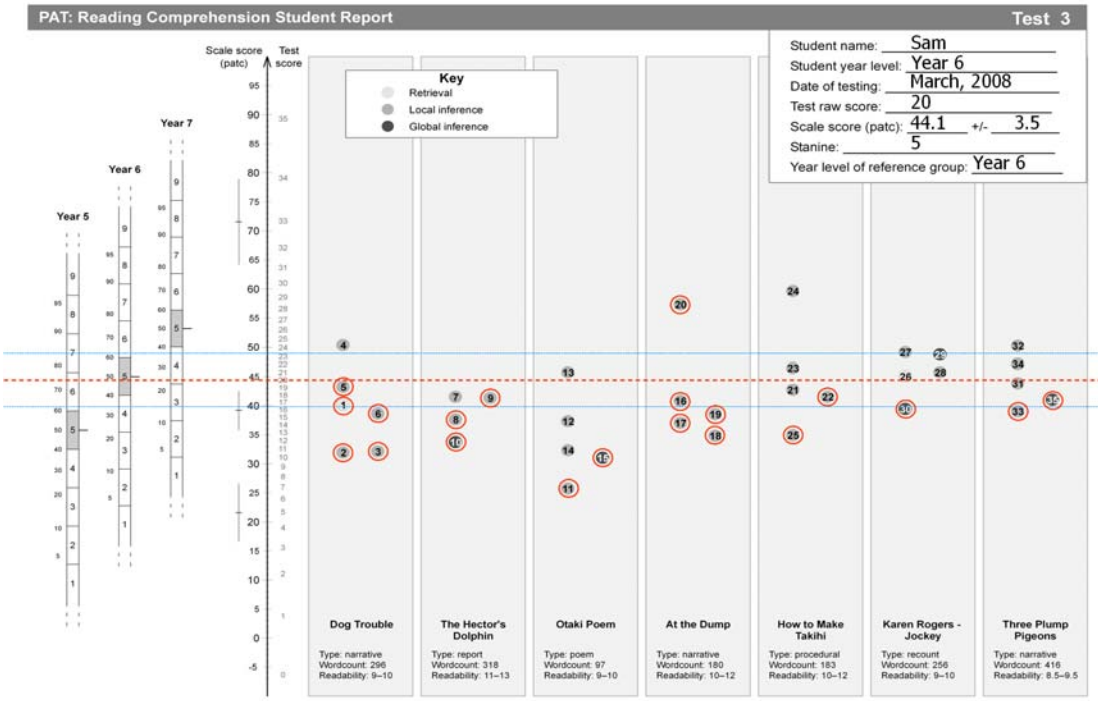
We can pull all this information into an informative report. From looking at the big picture of all the comprehension tests, we can look at just one. Figure 4 shows a report for one student who sat PAT:Reading Test 3, which is aimed at Year 6.

Now you can see the vertical text bands, which show there are seven texts in this test, and the questions (shown as round markers) within each text band as they are located on the PAT comprehension (patc) scale. Each task (or question) has a location on the scale.

This test has been marked to give the results for one student. The circles around individual questions indicate a correct response. The middle horizontal line is the student's score of 44.1 on the patc scale. The lines above and below it indicate a margin of error. We can expect this student to get 50 percent of questions around these lines correct, more than 50 percent of questions below these lines correct, and less than 50 percent of those above.

Looking at the year bands to the left, we can see that because this student is in Year 6, her patc score gives her a stanine of 5. If she were in Year 5, her stanine would be 6.

Figure 4 **The PAT:Reading individual report for Test 3**




## Describing the scale

As mentioned, each question in the new PAT tests has a location on the scale. By analysing the tasks we can describe the kinds of competencies associated with different points on the scale.

As you can imagine, this was a more straightforward task for maths than reading. With reading it is difficult to demonstrate in general terms the difference between what a reader at the top, middle, and bottom of the scale does. Both good readers and not-so-good readers can read a range of text types; locate information; make inferences; analyse, synthesise, and evaluate information; and so forth. All readers do these things and as they become better readers they do them in more sophisticated ways and with more complex texts—so when you try to describe progressions you end up saying similar things at each level.

This will be familiar to those of you who have used the English statements in both the old and new curriculum documents. Let me give you an example from the new curriculum. At Level 1 students are expected to "recognise and identify ideas within and across texts"; at Level 2, "show some understanding"; at Level 3, a "developing understanding"; at Level 4, "an increasing understanding"; at Level 5, "an understanding", at Level 6, "a developed understanding"; at Level 7, "a discriminating understanding"; and at Level 8 "a discriminating and insightful understanding" (Ministry of Education, 2007, English fold-out charts of achievement objectives by level). This is unavoidable when you try to describe literacy progressions in general terms.

The difference between what a good reader and a not-so-good reader does can best be illustrated with specific examples of text and task. So what we decided to do was to develop examples. We looked at all the questions across all the tests in order of difficulty shown on the scale and considered the difference between the easy and more difficult ones. We found that as you moved

up the scale there was an increasing requirement for students to be able to do each of seven things. Students were increasingly required to use:

- abstract information
- separated information
- multiple pieces of information
- implied information.

They were increasingly required to:

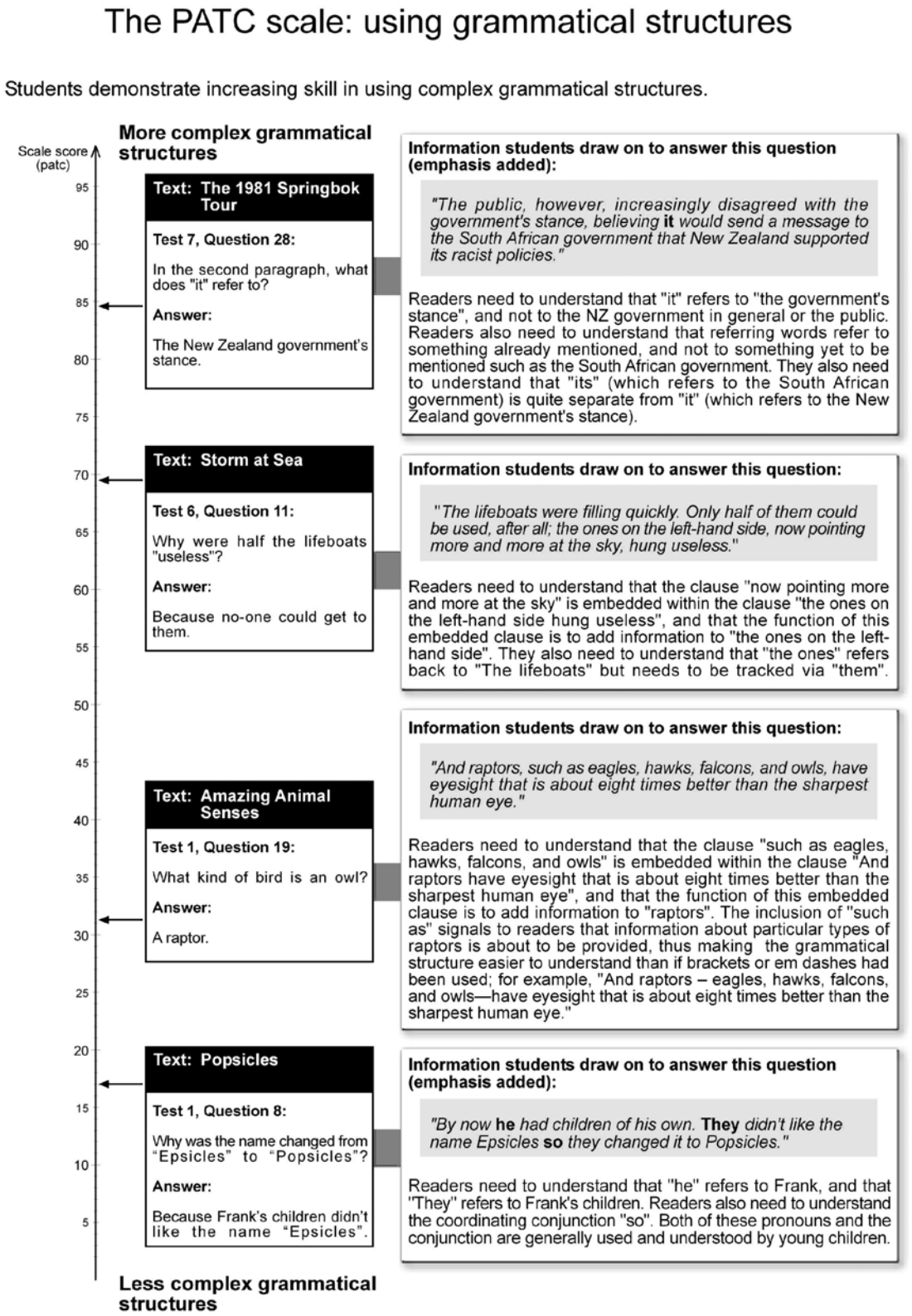
- reject competing information
- make meaning from increasingly complex vocabulary
- make meaning from increasingly complex grammatical structures.

For each of these seven areas we provided examples of four questions spread across the scale from least difficult at the bottom to most difficult at the top. The scale description for using grammatical structures is shown in Figure 5.

You will see we have provided the answer for each of the four questions in the boxes on the left-hand side. At the top of the boxes on the right we have provided the segment of text in which the answer is located. Below this we unpack what students need to be able to do to work out the answer. The arrows show where the questions are located on the scale.

The seven scale descriptions help illustrate what progression in reading comprehension looks like by focusing on the interplay between text and task. Each scale description can be focused on separately to consider what progression in using abstract information, separated information, multiple pieces of information, implied information, and so forth, might look like. If the scale descriptions are placed side by side, they can be read horizontally to see the range of skills crossing all seven areas students within a particular area of the scale demonstrate when faced with particular texts and tasks. Teachers can also look at the examples above this level to gain some idea of the skills these students may need to develop next.

Figure 5 **The scale description for "using grammatical structures"**



## Making too much of the data

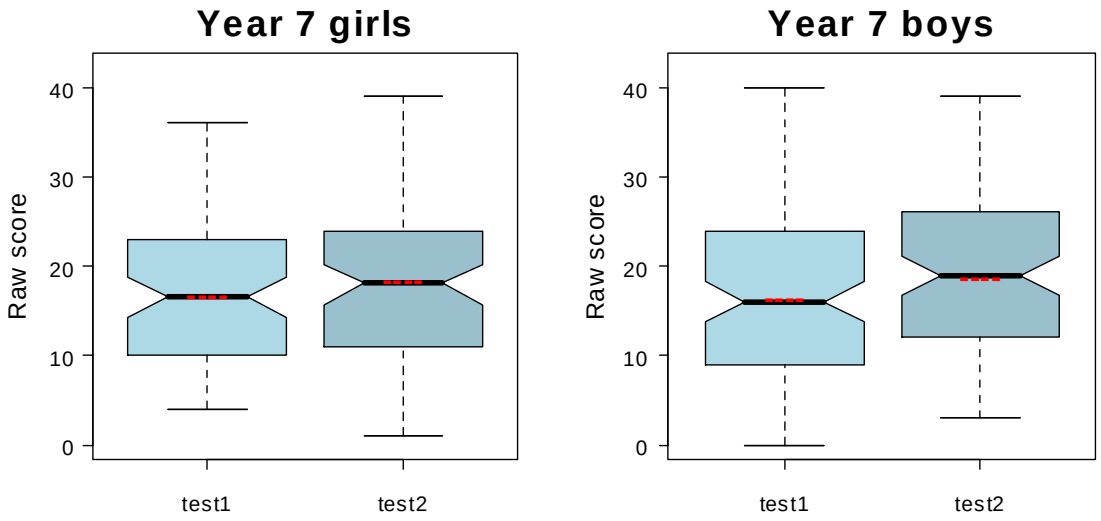
The reporting and scale descriptions available for the new PAT tests open up a wealth of possibilities. They allow teachers to examine what is being tested and measure how students are progressing. There are, however, dilemmas associated with providing a tool that can "dig down" like this. One of these involves the temptation to make too much of the data associated with a single testing point.

A test provides a "snapshot" of what students can do. No test is completely reliable, and on another occasion students could perform differently. Making valid inferences about what an individual student can and cannot do, on the basis of one test result, can therefore be difficult. This is particularly salient when inferences about a student's understanding are based on the response to a single test question.

Recently, we had the opportunity to look at what happens when students are asked to do a test twice. A school decided that their students would repeat the tests after all the answer sheets went missing before the tests had been marked. The original test papers were then found, but only after all the students had done the test again.

When both set of tests were marked, the overall achievement picture for each administration was very similar. Figure 6 shows the distribution of achievement from both tests broken down by gender.

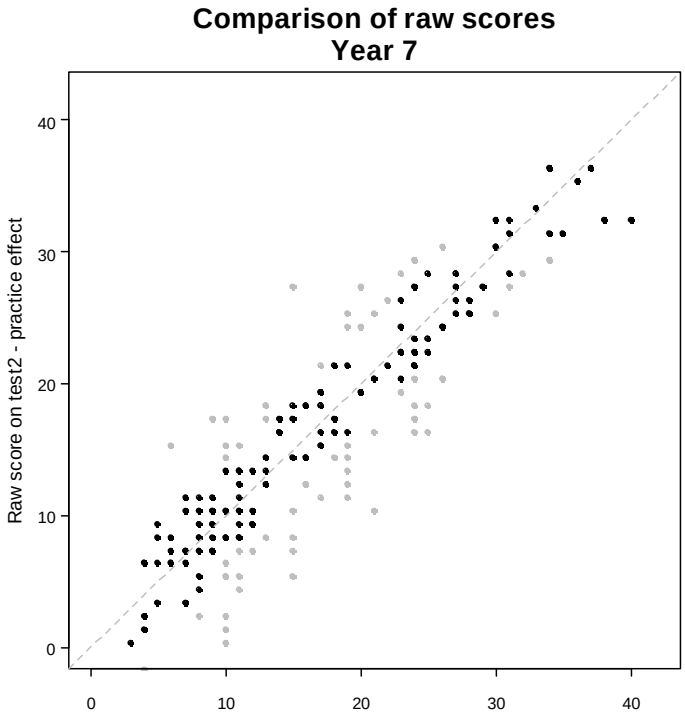
Figure 6 **Achievement for repeated administrations of the same test**




While there is evidence that the group of students as a whole gained a little from a practice effect, the overall distribution of achievement as measured on the first testing occasion is very similar to the overall distribution as measured on the second. But what about at the student level?

The graph shown in Figure 7 plots each student's score on the first testing occasion against their score on the second occasion. If the results were perfectly consistent we would expect the dots to lie on a straight line.

Figure 7 **Comparison of test scores for each student**




The graph shows a strong linear relationship between scores for the two tests, but it also indicates that for some individual students the two scores vary considerably.

All tests involve measurement error; that is, random noise that masks the true achievement of the students. To indicate the extent of this noise, tests often report student achievement as a range within which we can be reasonably sure the student's true score actually lies. By "reasonably sure" we mean that if the test was to be repeated, then about 66 percent of the time the score will lie within this interval.

The grey dots in Figure 7 have been used to show the students whose second score fell outside the confidence interval associated with the first score. As can be seen, about a third of the dots are shaded grey. This is about what we would expect.

Overall it seems that most students came up with a similar result on the second testing occasion, but certainly not all of them. This shows that it would be disingenuous to claim we knew exactly where an individual student was achieving on the basis of one test result.

As we have moved from the cohort to the individual level we have seen a resulting drop in consistency. What happens when we look at the individual-question level?

If the students were perfectly consistent we would expect them to respond to each question in exactly the same way on both testing occasions. When we look at students' response behaviour, however, we find this is not the case. Students are surprisingly willing to change their answers.

Each of the students had two opportunities to respond to each of the questions in the test. Over the two test administrations they can make five different responses to each question. These are listed in Figure 8.

Figure 8 **Possible responses to duplicate testing**

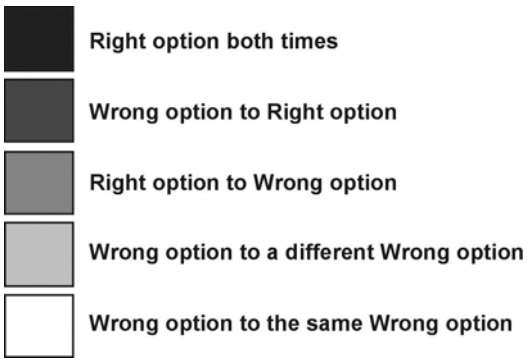
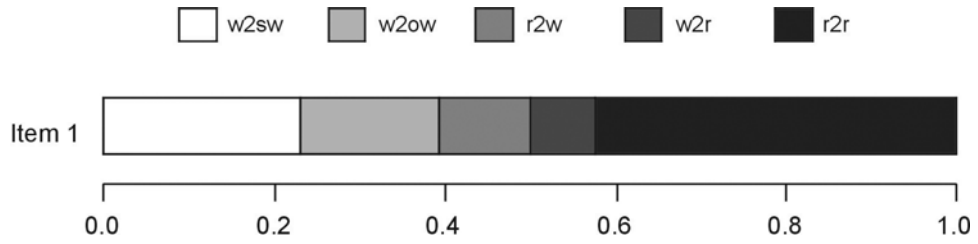



Figure 9 shows the way the students responded to the first question over the two test administrations.

Figure 9 **Response pattern for question 1**




We can see here that about 20 percent of the students who chose the correct option on the first testing occasion were prepared to change their answer to a wrong option on the second. A similar percentage of the students whose first response was wrong chose the correct option on a second attempt.

This kind of variability in responses indicates that it would be unwise to use a response to this question to diagnose whether an individual student understands the concepts involved. As demonstrated, a sizeable proportion of the students will change their mind given a second opportunity to respond.

This kind of pattern is repeated for all the questions. It can be seen as pretty natural behaviour. Some questions challenge the limits of what we know or can apply. Given a second opportunity to answer a question we try a different tack or have a moment of insight and choose a different answer.

Reports for the new PAT tests do allow teachers to look at how students have responded to particular questions. Figure 10 shows a report which displays the actual question answered and lists which option each student has chosen. A feature of this report is the use of colour coding to indicate whether students' responses are highly expected or unexpected. For instance, the names highlighted in red (showing here as black) indicate those students who scored very highly in the

test but have answered this particular question incorrectly when a correct answer was expected. Similarly, the names highlighted in green (showing here as grey) indicate students who have chosen the correct answer and who would be expected to do so given their overall performance on the test. It is not surprising that these students have answered correctly. The names that are not highlighted indicate that given these students' overall results it is hard to make a probabilistic statement about what kind of response could be expected. In other words, it would not be particularly surprising for these students to have chosen different answers.

Teachers can use this kind of report to look at what is being tested and get an indication of how their students are responding. The colour coding will indicate when responses are expected or unexpected. The report also reminds teachers that by itself the question is not necessarily a reliable test of whether a student understands the concept or skill being assessed.

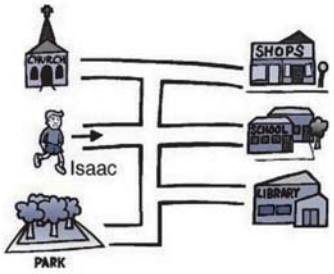
Figure 10 **The individual item report for PAT:Mathematics**

## Individual Test Item Report

**Progressive Achievement Test of Mathematics**
**Test 1 – Question 11 (Geometry/Measurement)**

Previous ◄ 6 7 8 9 10 11 12 13 14 15 16 ► Next

**11.** Isaac is walking along the road. He turns left, walks, and then he turns right. Where does he get to?

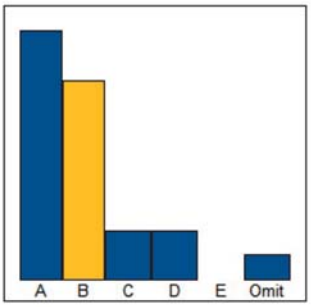



(A) the church
(B) the shops
(C) the school
(D) the library
(E) the park

**Distractor information**

***Demo School Year Class Room1***
***(Reference Group Used: Year 4, Number of students: 23)***

| A (43.5%) | B (34.8%) | C (8.7%) | D (8.7%) | E (0%) | Omitted (4.3%) |
|---|---|---|---|---|---|
| Kyle Mason | Michael Hersch | Lana Gunter | Charlotte Waters | | Shah Abhishek |
| Matthew Ryle | Deborah Din | Yentl French | John Carrcun | | |
| Anna Brown | Eric Park | | | | |
| Tyler Simmonds | Ali Drury | | | | |
| Hannah Farlap | Madi Bhardwaj | | | | |
| Jen Cortes | Jess Pascoe | | | | |
| Kole Tein | Amy Cho | | | | |
| Razan Ostel | Yune Sang | | | | |
| Reuben Stine | Young | | | | |
| Jessie Ambags | | | | | |
| **10 students** | **8 students** | **2 students** | **2 students** | **0 students** | **1 students** |



## Conclusion

At the start of the session we contrasted three perspectives on testing: testing as tyranny; testing as a magic bullet; and testing as tool use. Many of us probably feel more comfortable adhering to the "testing as tool use" viewpoint.

The new PAT represent very powerful tools that can enhance what we do as educators. The careful test construction, new ways of measuring progress, and the use of technology to dig deeper and report performance on the tests open up a whole lot of new possibilities.

To extend the tests as tools metaphor further, as a tool the new PATs provide leverage and give us the opportunity to magnify our efforts.

The new PATs are not a magic bullet. Neither do they have to be associated with the tyranny of testing. However, we can't escape from the dilemmas of testing simply by calling the PATs a tool. As we have seen, tools have unintended consequences and unconscious effects. As soon as we pick up and use a tool we are implicated in these effects.

Assessment tools have to be used with wisdom. We need to be assessment literate, and take responsibility for our decisions and their consequences. We have to have what Neil Postman called "crap detectors" (Postman, 1969).

Neil Postman wrote a lot about technology. One thing he warned against (Postman, 1979) was something he called "technicalisation". According to Postman, technicalisation is something that happens when we forget there are alternative methods and ways of doing things and reduce all possible ways of doing something to one way only. He describes it as converting "a method" into "the method".

It is easy to become seduced by impressive techniques, especially when they involve computer technology and fancy reporting. Perhaps the biggest dilemma would be to forget that the PATs, or any other national assessment tool for that matter, are only one way of assessing student progress. To do that would be to default to one way of defining what it means to learn and achieve. It would catapult us towards the dilemmas associated with testing as tyranny.

## References


Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46.

Davis, B. (1996). *Teaching mathematics: Towards a sound alternative*. New York: Garland Publishing.

Dickens, C. (1838/1992). *Oliver Twist*. Ware: Wordsworth Classics. (First published in 1838 by R. Bentley.)

Forster, M. (2003, May). *Text difficulty and the measurement of reading growth*. Presentation at Just Lexellent? Motivational reading systems to support school programmes, National Library of New Zealand, Wellington.

Hardy, T. (1891/2007). *Tess of the d'Urbervilles*. London: Penguin Books. (First published in 1891 by James R. Osgood, McIlvaine.)

Hoffman, B. (1964). *The tyranny of testing*. New York: Collier.

Hunt, G. (2001, April 17). Closing the education gap is step no. 1. *The Dominion*, p. 11.

Jacob, B., & Levitt, S. (2004). To catch a cheat. *Education next*, *4*(1). Retrieved 1 March 2008, from http://www.hoover.org/publications/ednext/3345096.html

Karmel, L., & Karmel, M. (1978). *Measurement and evaluation in the schools*. New York: Macmillan.

Ministry of Education. (2007). *The New Zealand curriculum for English-medium teaching and learning in years 1–13*. Wellington: Learning Media.

Postman, N. (1969, November). *Bullshit and the art of crap detection*. Paper presented at the national convention for the Teachers of English, Washington, DC.

Postman, N. (1979). *Teaching as a conserving activity*. New York: Delacorte Press.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago: The University of Chicago Press.

Tool. (2008). In *Wikipedia, the free encyclopedia*. Retrieved 1 March 2008, from http://en.wikipedia.org/w/index.php?title=Tool&oldid=205892862

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*, 121–136.

Winterbourne, R. (1944). *Educating backward children in New Zealand*. Wellington: New Zealand Council for Educational Research.

# Māui, Tawhaki, kaupapa Māori, and … "progress"?

**Garrick Cooper (Ngāti Karaua and Te Pirirākau)**

*New Zealand Council for Educational Research*

*Kaua a tātou tamariki e tukuna ki te kura, kei hoki mai ki te patu i a tātou.*[1]

*Do not let our children go to school, lest they return to hurt us [with the ways of the Pākehā].*

*—Te Kooti Rikirangi, founder of the Ringatū movement*

## Introduction

The purpose of this paper is to think about the usefulness of current ideas of progress within mainstream education—in particular I want to provide a Māori lens on these ideas of progress. Gilbert's paper in this volume argues that in the postmodern era we need new ways of thinking about progress that enable us to move beyond universalism to embrace diversity. I want to take up this challenge of "new ways of thinking" by drawing from Māori epistemologies. In doing so I am wanting to build on the idea suggested by Gilbert that if we are to build a 21st century education system we need to redesign "the system so that it focuses, not on sorting out who 'has what it takes' to go on to higher education, but on finding ways to ensure that everyone achieves" (2008, p. 71).

My initial idea for this paper was to draw from kaupapa Māori theory, specifically the theory/theories of progress that were being practised in kura kaupapa Māori. However, as I examined this idea more closely, I felt that the ideas of progress practised in kura kaupapa Māori replicate, or are similar to, those practised in English-medium education. As kura kaupapa Māori aspire to be informed and driven by Māori epistemologies, it seemed to me that Māori traditions would be a rich source of thought and wisdom to help us look at ideas of progress in education in "new" ways.

---

[1] Personal communication, Te Awanuiārangi Black, he tohunga nō te hāhi Ringatū (9 August 2007).

In drawing from some Māori traditions I want to highlight how they provide insights into a dynamic and ongoing review process that provides a mechanism for challenging and, if necessary, changing social institutions and beliefs. This mechanism provides a socially sanctioned way of challenging and critiquing the status quo. Two series of traditions are particularly useful for this purpose—the Tawhaki and Māui traditions. These traditions are an epistemological manifestation and expression of a safeguard against oppression and hegemony that has been collectively theorised by Māori and Polynesian ancestors over many, many generations. Māui is a Māori and Polynesian equivalent of an "out" clause if you like. These traditions anchor social practice in the known and stable while also making space and allowing for new frontiers to be established, and the unknown and unstable to be explored. Over time these in turn become the known and the stable. This constant examination and renewal is critical to any society, but, in my view, has been lacking when thinking about the dominant notions of progress in our education system.

Before expanding on these ideas there are a number of assumptions and beliefs that I want to make visible. Firstly, I start from a position that Māori epistemologies, and the traditions and beliefs that are a part of these epistemologies, are as relevant and applicable today as they were in the times of famous ancestors like Tawhaki and Māui. Secondly, I believe that dominant ideas of progress deployed in "mainstream" education are not particularly helpful for some students, and can potentially have materially harmful effects for students—particularly Māori and Pasifika students,[2] and perhaps others. Doing well at school, "progressing", or learning does not occur in a social vacuum. There are particular socialisation practices that operate in schools. In order to "do well" and "progress" in schooling a student needs to know these practices, processes, and language, or to have particular forms of what Bourdieu (1973) calls "cultural capital", *and* conform to these. Socialisation processes in themselves are not inherently a "bad" thing; however, in the colonial (and multicultural) context, socialisation processes are, or have the potential to be, assimilationist. This is what Te Kooti warns us about in his *kupu whakaari* or prophecy. Furthermore, schools and other cultural institutions in Aotearoa New Zealand now work to avoid assimilation practices. We need to guard against reverting to these practices.

## Tawhaki and Māui: Māori epistemologies

As I have already suggested, the Tawhaki and Māui traditions provide ideas that could be used to build new ways of thinking about progress; ideas that are informed by Māori epistemologies. Before developing this idea further, and using it as a lens to consider the limitations of English-

---

[2] Evidence of this is found in what is most often referred to as the yardstick of student progress in English-medium education, student achievement. So, for example, 2007 NCEA results show that between 14.8 and 25.3 percent fewer Māori and Pasifika than Pākehā students achieved NCEA Level 1 to Level 3 qualifications (New Zealand Qualifications Authority, 2008). What makes this more concerning is that only 55 percent of Māori and 61.8 percent of Pasifika students attempted these qualifications, compared to 76.4 percent of Pākehā students. When I talk about materially harmful effects here, I am referring to the types of "psychological violence" that Fanon (1963) talks about in the colonial context and the "symbolic violence" that Bourdieu (1971, 1973, 1974) talks about in the schooling context.

medium education notions of progress,[3] I need to provide more detail about the Tawhaki and Māui traditions.[4]

Māui is more widely known in Aotearoa than Tawhaki; however, Tawhaki is no less important in Māori traditions. Tawhaki was born of chiefly lineage. He was a direct descendant of the gods. The chiefly lineage of Tawhaki, in some traditions, was from his female genealogical lines. For example, his grandmother was Whaitiri, or *thunder*. Like Māui, Tawhaki was smart and a strategiser, and he used these skills wisely when it came to overcoming challenges. He was blessed with mighty powers, befitting someone of his status. Māui was the *pōtiki*, the last-born child. Pōtiki are known to be cheeky, intelligent, challenging, and brave—sometimes too brave, as in the events that led to his death! Māui *is* the embodiment of all these dispositions and traits.

Let us now look at a few of the Māui and Tawhaki traditions. There are three traditions that I want to very briefly retell.

1. Māui set about obtaining the knowledge of fire from his kuia, Mahuika. He first had to placate her in order to get her attention. Through cunning means he then tricked her into giving him her fingernails, where the power of fire resided, until she had only one left. When he went back for the last fingernail she threw it at him in frustration and it lodged in the kaikōmako tree. The branches of the kaikōmako are still used to make fire.

2. In the time of Māui the days were too short due to the sun (Tama-nui-i-te-rā) travelling too fast across the sky, and people were not able to complete their work. Māui took the magic jawbone of his grandmother and, with his brothers, embarked on an expedition to capture and slow the sun down. Māui's expedition was successful and hence the days are now long enough for us to complete our work.

3. After a series of feats, Tawhaki's fame spread to the heavens and captured the attention of the celestial maiden, Tangotango. She visited him at nights and in due course became pregnant. When conducting the tohi ritual at the birth of their child, Tawhaki commented that the child smelt. Tangotango was insulted and returned, with their newborn child, to the heavens. Tawhaki, with his brother Karihi and their slaves, went to the heavens in search of his wife. On the journey the slaves were told not to look at the sacred citadel of Tongameha. The slaves did not listen and perished. When Tawhaki made it to the heavens he changed his appearance into that of an old man. Tawhaki was made to perform the duties of a slave by the family of Tangotango. This was his way of humbling himself to Tangotango and her family in order to make up for his insult.

---

[3] I want to acknowledge the work of Tongan academic Dr 'Okusitino Mahina of The University of Auckland, whose work and ideas around Māui-kisikisi and Tahaki in the context of historical and contemporary Tongan politics has influenced these ideas about what Māui and Tawhaki represent in Aotearoa.

[4] I should note that there are many tribal variations of these traditions. The variations are significant for the different tribes and their contexts. In describing these traditions I have drawn extensively from *Ka Whawhai Tonu Mātou: Struggle Without End* (Walker, 2004). In turn, Walker drew most of the Māui and Tawhaki traditions from *Ngā Mahi ā Ngā Tūpuna* (Grey, 1953), whose main informant was Wiremu Maihi Te Rangikaheke of Te Arawa. Most of what we have come to know as the "orthodox" versions of these traditions can be traced back to Te Rangikaheke via George Grey's *Ngā Mahi a Ngā Tūpuna*.

What can we read into these traditions with regards to social institutions in Māori society?

Both the Māui and Tawhaki traditions show us where social "norms" lie. For example, the commands of superiors need to be respected and the ultimate sanction can be death, as in the case of the slaves of Tawhaki and Karihi, if these commands are not respected and observed. Another example is that the insult that Tawhaki made to his wife needed to be addressed. To make that type of comment with regard to something as natural as giving birth is socially unacceptable. Furthermore, Tawhaki subordinates himself to these social norms, seeking redemption from Tangotango and humbling himself to her and her family, despite his high status. The Tawhaki stories have an underlying theme of observing social mores and edicts—of respecting tradition.

On the other hand, while the Māui traditions show us where social norms are, they also tell us where they once were and, more importantly, how they have been changed. For example, women are repositories of knowledge, and the acquisition of new knowledge requires commitment and strategy. One cannot expect that this knowledge is a birthright. The person trying to acquire the knowledge must be strategic—as the story about the acquisition of the knowledge of fire from Mahuika shows. Māui first needed to placate Mahuika. This tells us that holders of knowledge are to be respected. However, Māui goes further—by using a form of deceit to acquire the knowledge, he "transgresses" the norm of being respectful to holders of knowledge. The deceit was justified by need and the benefits that such a feat would bring to people. The same could be said of the violent act of beating Tama-nui-i-te-rā into submission. Finally, all of these challenges to social norms require careful consideration of the consequences, both positive and negative; strategy; planning; and, most of all, courage and guile. For Māui then, social norms are not set in concrete and are not static. The Māui traditions have an underlying theme of—when the end is justified—transgressing social norms and tradition.

How can we respect each of these traditions if Tawhaki tells us to observe social norms and Māui is challenging and transgressing them? The Tawhaki and Māui traditions represent a type of binary—something that Hanson and Hanson, in their poststructuralist study *Counterpoint in Māori Culture* (1983), have suggested that Māori culture has in abundance. These traditions could be interpreted as paradoxical if we are to "listen" to and respect each one. However, importantly, paradoxes sit comfortably in Māori epistemologies. This is in contrast to some Western epistemologies which have a tendency to try to find a "truth", and attempt to reconcile tensions and conflicts. If these messages are apparently conflicting then how does one navigate through this? There are two things that I would say: firstly, the traditions need to be seen together and not as separate, competing narratives; and secondly, I would argue that it is context that then becomes the arbitrator of appropriate action and ways forward, rather than strict adherence to what either Tawhaki or Māui perhaps represent.

What can be said about these traditions, then, when we put them alongside each other? At one level, the Tawhaki traditions anchor us in stability, the known, and adherence to current social norms. Tawhaki represents respect of, and subordination to, cultural norms, and shows the way to do this. In contrast, the Māui traditions push and pull us into the unknown, into instability, and new frontiers. Māui represents particular dispositions and skills required to assess, challenge, and transgress social norms and create new ones. Together, Māui and Tawhaki represent a mechanism for review and renewal. It is a socially sanctioned way of ensuring that there are social constants for stability *and* the ability to critique and change social practices as needed.

## Some implications for education

What messages can we take from these traditions and apply in our analysis of schooling and education? At a surface level these traditions, along with a range of other traditions, provide us with the social norms, values, and beliefs that Māori communities practise. In education, kura kaupapa Māori aspire to practise these in everyday activities. These inform the ways in which kura operate, from concepts of *manaaki i te tangata* (hospitality and providing for one another) to *ako* (learning as a two-way process). These values and beliefs inform the types of socialisation processes that kura kaupapa Māori practise.

Tradition, or perhaps more accurately, the known, is very important in Māori society, but so is the ability, capacity, and imperative to transgress social norms and create new ones. The Māui and Tawhaki traditions provide for socially informed ways of changing and transforming social norms. These traditions—and again, along with a whole range of other traditions—tell us about the types of skills, dispositions, qualities, and characteristics that are valued in Māori society. This is perhaps an important message for those who are responsible for developing kura kaupapa Māori curriculum. This range of skills and dispositions is, or at least could be, the basis of what kura kaupapa Māori could try to develop in their students. These could inform the types of "key competencies", if you will.

Finally, we need the full range of these skills, and others, in our society. Māui and Tawhaki belong to the same set of traditions and need to be read alongside each other. Together, they speak to the types of skills that are valued in Māori society. Māui and Tawhaki are quite different; however, they are equally important, and perform equally important roles. In Māori communities you will often hear children who are displaying the qualities of Māui being referred to as *nanakia*, or *hianga*. Some children have a "natural" inclination to Māui-like qualities, others tend to be more reserved and observe tradition and the known. So using this message for schools, education is not about producing a whole lot of "little Māui" or a whole lot of "little Tawhaki"; we need both of these and everything in between.

I need to state why I think these traditions can be used in the modern context to think about and critique educational practices. As I have mentioned, these traditions are not just "stories" about the fanciful adventures of some mythical ancestors. These traditions contain the collective wisdom and thought of many, many generations of ancestors, projected back onto historical narratives and ancestors. They are devices used to store knowledge and wisdom about the full range of social practices in Māori societies. We can draw from these traditions this accumulated knowledge and wisdom and apply these insights to analyses of current social phenomenon. In this specific case I am looking at how they might be used to critique a particular educational construct, that of "progress".

In the next section, I use ideas that I have drawn from these Tawhaki and Māui traditions—those of review and renewal and the range of skills and dispositions that are valued—and use them to suggest that we need to move beyond thinking of progress as being prescribed and recognised in one way to imagine new, and potentially important, other ways.

## Rethinking notions of progress

In our current education system the predominant theory of progress is based on an idea that progress is linear in nature and that there is a prescribed "normal" pathway. Progress is measured against a set of "norms" and "benchmarks" which are derived from sometimes quite sophisticated statistical analyses and equations (Gilbert, 2005). The concept of "norms", as we have seen from the Māui and Tawhaki traditions, is also something that is present in Māori epistemologies. The context is different, but the concept is similar. At a micro level, norms in education are deployed in achievement tests and assessments. In schools, achievement test results are used to track students' progress against a prescribed and "known" pathway of progress. Achievement results tend to get used in ways which have the effect of trying to make students conform. So, for example, if a student does not achieve a particular norm in a given test, the student is given remedial work or lessons so that they can be "brought up to speed" or, less delicately, "fixed". At a macro level, in "bringing students up to speed" we are attempting to make them conform to a particular model of academic and social achievement—one that is consistent with and informed by the dominant forms of socialisation that Te Kooti warns us about. It is from this discourse that we get terms like the "long tail of underachievement" and "closing the gaps". Not surprisingly we are not, and may never be, successful at making all students conform; nor, in my view, is this always desirable.

There are a number of things, then, that we could say about the long tail of underachievement. The standard interpretation of these data is that there is a group of students—Māori and Pasifika students are disproportionately represented in this group—for whom current pedagogies and teacher practices are not working. There have been a number of interventions that have been designed with the aim of changing and improving teacher practice so that there is a better "fit" for these students. Te Kotahitanga and Strengthening Education in Māngere and Ōtara are two examples. This is a shift away from earlier interventions, which would have interpreted the "long tail" as a sign that it was the students, and *not* the system, who were "deficient", and that the students just needed more of the same. These recent types of approach could be described as a part of the wider "schooling improvement" agenda. This may be a productive approach, and useful for some students, but certainly not sufficient. To simply keep trying to improve what we are already doing, I think, will not resolve the issue.

The Māui and Tawhaki traditions together show us that although there are norms, sometimes these need to be challenged and, at times, transgressed. So if we look at the long tail of underachievement, for example, using the notion of transgressing norms, we need to look seriously at the so-called norms against which the "long tail" is located. Let us return to Māui for a moment. When Māui was challenging norms, wider society frowned upon him as he was challenging the "known", or tradition. However, ultimately what he was doing was showing a pathway forward that would benefit all people, it was just that at the time not everybody could see it. Perhaps this group, or these groups, of students are also doing this?

Another reading of the long tail of underachievement could be that our current ideas about and measures of progress are inadequate for these students, and therefore need to be challenged and transgressed. Rather than pathologising and deficit-theorising this group of children ad infinitum, and trying endlessly to make them conform to so-called "norms", we should be—and this is important—looking carefully at what they *are* doing and doing well. If we are careful in our observations of what they are actually doing, we will be introduced to knowledge, skills, and competencies that could be beneficial for all people—the good things that they are doing could

become the “norm”, as Māui saw before everyone else. They are telling us something, and I think we need to listen carefully.

Finally, we can make a link here with another key message from the Māui and Tawhaki traditions, and that is that we as a society need a range of skills, dispositions, and characteristics. There is no one correct set or combination of these, but a range of different sets and combinations. This sits comfortably with the idea that there is no abnormal range or abnormal combinations either. If there are no abnormal sets of skills, dispositions, and characteristics or combination of those, then any set or combination can then be normal. In a sense, abnormal can be and has the potential to be normal. In essence we are moving to the idea that there is no single prescribed knowledge and skill set, but a range of knowledge and skill sets, and a range of combinations (Gilbert, 2005).

## Concluding remarks

What message(s) should we take to think about a new theory or theories of progress? Māui and Tawhaki tell us that both maintaining norms *and* transgressing norms is normal. In theorising progress, then, we could say that being “not normal” can be “normal”. A new theory of progress needs to accommodate the “normal” *and* the “not normal”. In fact it needs to be able to not just strengthen the normal (the known, tradition), but also accommodate and make space for the abnormal (the unknown and new frontiers). Instead of pathologising these students we need to be critically comfortable with both norm conformers and norm trangressors, and look for new “frontiers” of achievement. We should listen and pay careful attention to what students are telling us (through dialogue, and engagement with children, and via current assessment methods) and not what we think they *ought* to be telling us. We need to think about progress not in a narrowly defined, prescribed band of achievement, as in the linear theory of progress, but in terms of strengthening and affirming the known (what we know of now as knowledge) whilst simultaneously challenging and questioning it (moving to the unknown). Our theory has to accommodate a range of ways of progressing, as I think the students in the “long tail” are “telling” us. We may even, or should, stop using pathologising discourse like the explicit “normal” and the implicit “not normal” because, whilst based on historical data, assigning labels such as these to students is a projection into the future, a future that we do not and cannot know. Perhaps we should also look at what *is*, and from there, what *could* be. Finally, I think, we need to allow the messages of Māui to find expression in our new theories of progress.

## References


Bourdieu, P. (1971). Systems of education and systems of thought. In M. F. D. Young (Ed.), *Knowledge and control: New directions in the sociology of education* (pp. 189–207). London: Collier Macmillan.

Bourdieu, P. (1973). Cultural reproduction and social reproduction. In R. Brown (Ed.), *Knowledge education and social change* (pp. 71–112). London: Tavistock.

Bourdieu, P. (1974). The school as a conservative force: Scholastic and cultural inequalities. In J. Eggleston (Ed.), *Contemporary research in the sociology of education* (pp. 32–46). London: Methuen.

Fanon, F. (1963). *Wretched of the earth.* New York: Grove Press.

Gilbert, J. (2005). *Catching the knowledge wave?: The knowledge society and the future of education*. Wellington: NZCER Press.

Gilbert, J. (2008). "Progress" in 21st century education? In *Making Progress—Measuring Progress* (pp. 63–73). Wellington: NZCER Press.

Grey, G. (1953). *Ngā mahi a ngā tūpuna.* Wellington: Māori Purposes Fund Board.

Hanson, F. A., & Hanson, L. (1983). *Counterpoint in Māori culture*. London: Routledge & Kegan Paul.

New Zealand Qualifications Authority. (2008). *National Qualifications Framework statistics 2007: Learning area results distribution by ethnicity.* Retrieved 15 May 2008, from http://www.nzqa.govt.nz/qualifications/ssq/statistics/nqf-stats.do?comparison=2&ch=3020&year=2007&nqfLevel=0&st=2&cg=0&decile=N&eg=e&eg=m&eg=p&eg=a&eg=n

Walker, R. (2004). *Ka whawhai tonu mātou: Struggle without end* (Rev. ed.). Auckland: Penguin Books.

# “Progress” in 21st century education?

**Jane Gilbert**

*New Zealand Council for Educational Research*

*Time flies like an arrow, fruit flies like a banana.*

*—Groucho Marx*

The term *progress*, as it is usually used, means moving in a particular direction in order to arrive somewhere. In educational contexts, students “make progress” by passing through—and successfully completing—a series of developmental stages. Step by step, they come to know and be able to do more—and harder—things. This concept of progress is, however, the product of a particular period in history, and a particular set of cultural understandings. This paper unpacks some of the assumptions, values, and goals that underlie the concept of progress, with a particular focus on its use in educational thinking. It argues that these assumptions, values, and goals are 20th century (or, in some cases, 19th century) ideas, and that we need to think about them differently if we want to build a 21st century education system. The paper has three parts. The first outlines how and why the idea of “progress” is important in modern thought. The second looks at some postmodern critiques of progress—in particular, the role it plays in organising our thinking about education. The third part explores some alternative ways of thinking about learning—about building it and measuring it—that could be more appropriate as a framework for developing a 21st century education system.

## “Progress” in modern thought

Most of the “big ideas” that frame our current institutions (and our thinking in general) are modern ideas: that is, they developed in, and are characteristic of, a time in (European) history known as the modern period. Beginning in the mid to late 18th century, around the time of the American and French Revolutions, and the Industrial Revolution in Britain, this period was a time of major social, political, and economic change. It saw the development of industrialisation, capitalism, nation states, and science, as well as a major expansion of European interests into the rest of the world. The concept of progress was very important to all this. Underpinning it—and the other ideas that made these developments possible—are a number of “big ideas”. Very briefly, these are as follows. First is the notion of people as rational, autonomous individuals, or “selves”,

who think and act independently of other individuals. This idea (which seems obvious and "natural" to those of us who have been enculturated in the Western European tradition) underpins all modern political, social, and economic thinking, including educational thinking. This idea is, however, a construct, and it is a construct that has some very important material effects. One of these effects, relevant to the topic of this paper, is that at the conceptual level it excludes a great many people. Another effect is that emphasising individuals as separate, independently acting entities means that the relationships and connections between individuals are *de*-emphasised—to the point of invisibility.

There is a very large academic literature (mainly in political science) on this point. For example, 20 or more years of feminist political scholarship has shown us that, while all individuals are people, in modern Western thought, not all people are individuals.[1] Several major groups of people are not—and have never been—included in the modern concept of individuality. When this concept was first developed, the "individual" was an independent head of household who governed all those who came under his household's jurisdiction as its dependants. This individual—the social, political, and economic "actor" in modern thought—thus "stood for" a whole range of others—women, children, and non-property-owning men, for example—who were seen as being part of the domestic—or nonpublic—sphere. These "others" did not participate in the political, economic, and intellectual activities that are the basis of public life, but, importantly for later developments, they were not *conceptually* part of public life. They were not seen as autonomous, independent thinkers, and they were not free to live their lives as they chose. They did not have equal access to knowledge, and, because they were not rational, they could not be the subjects—or creators—of knowledge. These people were not equal, they could not vote, and they did not have the same human rights as those who were "fully individualised".[2]

While this way of thinking is of course no longer legitimate—all adults (unless they are in a prison or psychiatric institution) can now participate in electing their representatives in government, and, in theory, all have equal rights—its origins leave their traces, in the form of some serious, and as yet unresolved, conceptual problems. One problem relevant to this discussion is the problem of equality. In modern thought, all individuals are, in theory, equal. They are seen as having an equal capacity for rationality, and as having certain basic human rights. In the 20th century it was generally acknowledged that many people would require assistance to realise these basic rights and capacities, and, as a result, access to education, health care, and some form of social welfare "safety net" became a basic "social right". However, these systems, because they are underpinned by the same conceptual framework that made them necessary in the first place, cannot actually produce equal opportunity—or "reduce disparity" as we say these days—in the way they were set up to do. As I will argue later in this paper, the conceptual frameworks on which these systems are based lead them to employ a range of technologies that, in effect, reproduce the very inequalities they were set up to reduce. First, however, I need to introduce the second "big idea" of modern thought.

If the "individual" is one of modern thought's key ideas, the other essential idea is the emphasis on reason and knowledge, particularly scientific knowledge, as the route to human freedom and happiness. In modern thought, science and progress are intimately linked. Through science we

---

1 See, in particular, the work of Carole Pateman (1988, 1989) and Anna Yeatman (1994). For a more detailed summary of these ideas, see Chapter 5 of Gilbert (2005).

2 This phrase comes from Yeatman (1994, p. 78).

will achieve enlightenment: that is, ultimate truth, the point at which all is known. As Stephen Hawking puts it, science's goal is to logically articulate the universe's rational organisation; it is to develop a single, all-encompassing theory that describes and accounts for everything in the universe. This goal, once achieved, will allow us to "know the mind of God", and science will no longer be necessary (Hawking, 1988, pp. 13–14, 185). While this goal has not, as yet, been realised, scientists have provided the kind of knowledge needed to explain, predict, and control the natural world. This knowledge has in turn been used to produce the enormous growth in wealth, the colonial expansion, and the technological developments that were a feature of the modern period.

However, science's progress rests on two important assumptions. One is the idea (articulated by Hawking) of the universe as a system with its own independent rationality and logic, a system that is organised by some sort of plan, and into which everything "fits". Science's job is to articulate this system, to discover its logic. Once we understand the system, we can find ways to help everything (and everyone) fit it, and everything and everyone will be happy in its rightful place. The second assumption is the idea that scientific knowledge can only be produced by a certain kind of "knower". This knower is the rational, autonomous "individual" of modern thought that, in its fully developed (that is, educated) form, has the "real oil": the attributes and dispositions (and the training) of a scientist (for example, the ability to think rationally, to split thinking from feeling, observations from value positions, and so on). As we have seen, this concept of individuality excludes certain major classes of person: however, scientific knowing, because it requires this individuality in its fully developed form, is even more exclusive.[3]

Producing science requires particular kinds of individuals. These individuals, in order to *be* those individuals, need scientific knowledge. They are formed in—and constructed by—the same set of assumptions that make science possible. Thus these two ideas—the "individual" self, and scientific forms of knowledge—are essential to each other. Each depends on and enables the other, and each in turn depends on and enables the modern world view (with all its successes and failures).

Modern education systems as they have evolved have a key role to play in all this. Education systems are, on the one hand, designed to produce the particular kinds of individuals needed by this system. On the other hand, however, they are also designed to organise, sort, and screen out those who don't seem to have the required attributes. The "progress" metaphor plays an important role in making this possible. The education "sciences" are framed by a model of learning that mirrors the scientific model of knowledge building. Important concepts are built up, step by step, one on top of—or next to—the other, in a logical, sequential, linear way. These concepts are thought of as being like things, concrete objects that can build on, and be added to, other objects, like bricks in a wall. This building process eventually leads to the development of some sort of whole—complete understanding of a discipline, for example (or the universe). Along the way, educationists measure the extent to which learners have in fact managed to build these concepts, the extent to which they "know", or "have" these concepts, or can "do" certain things requiring an understanding of them. This information is then used to assess the learners' overall progress—and, by implication, their general "ability"—relative to certain norms. Some learners "measure up", and are allowed to go on to "higher" forms of education designed to further develop—or

---

[3] There is also a very large literature in this area. See, in particular, Bleier (1984), Bordo (1987), Haraway (1988), Harding (1986), Keller (1985), and Tuana (1989).

"bring out"—their capacity to be fully rational, autonomous individuals. The others who don't "have what it takes" are screened out, and, in the process, pathologised because they do not—and cannot ever—become the fully functioning individual of modern thought.

The British educational psychologist Valerie Walkerdine has written extensively on these issues, in a series of books and papers published in the 1980s. For example, her 1988 book *The Mastery of Reason* shows how mathematics education is a cultural practice which not only produces the kinds of knowers needed in modern scientific society, but also, through its emphasis on rationality and reason, maintains the fantasy of control over a "calculable" universe that is a necessary part of the modern sociopolitical order. In her 1989 book *Counting Girls Out*, she explores how these practices legitimate the exclusion of girls as not "having what it takes" for "real" success in mathematics. In a 1984 article on developmental psychology (Walkerdine, 1984/1998), she shows how the discourse of child development as a series of ages and stages that children pass through, because it is accepted as scientific fact, and taught to teachers and parents as such, has the effect of ensuring that most children do in fact go through the prescribed stages. In other words, children's development is constructed to ensure that it fits with the scientific model provided by developmental psychologists. Children who don't appear to be following this pattern are put into remedial programmes to "correct" their development so that it does fit the model. Schools, Walkerdine argues, are an "apparatus of regulation and classification … founded in science" (p. 164). While they are, in theory, designed to "bring out" each and every child's natural capacity for rationality, to liberate children by facilitating their "natural" development, they are in fact a major agent of normalisation. Furthermore, she argues, following Foucault, the rise of mass education, alongside that of "scientific" forms of administration, has made modern forms of government—which rely on covert regulation rather than overt coercion—possible. These forms of government (and their "apparatuses"—such as education) produce the normal "subject" (or individual) of modern society. At the same time, however, they also produce this subject's abnormal, pathologised "other", a category that is necessary to the system in that we can only know what normality is by knowing what it is not.

Walkerdine's work is a major challenge to many of the ideas underpinning our current public education system, a system which is taxpayer-funded because it is supposed to provide the foundations on which the "level playing field" of equal opportunity is possible. She shows how the modern notions of reason, progress, natural development, and liberation work to legitimate the classification, control, regulation, and normalisation of people that is essential for modern forms of government to "work". Her work, the result of a series of projects carried out in the UK in the 1980s, has, it is fair to say, had very little effect on day-to-day practices in schools, possibly because it was ahead of its time. I refer to it in this paper, written more than two decades later, because I think it is possible that its time has now come. As is now widely acknowledged, we are moving out of the modern period, into a time widely known as the postmodern period. There is widespread awareness that our education system, created as it was to serve the needs of modern society, is in need of some rethinking if it is to serve the needs of postmodern 21st century society. In the next section of this paper, I outline some of the "big ideas" in postmodern thinking, focusing in particular on the critique of "progress", and then, in the last part of the paper, I look at how current educational thinking might be reoriented so that it can work with—not against—these ideas.

## The postmodern critique of "progress"

Postmodernism is characterised by what Zygmunt Bauman (2007) calls the passage from "solid" to "liquid" times—the end of traditional social structures and institutions, and the end of what Jean-François Lyotard (1984) calls the "grand narratives"—the overarching, one-size-fits-all "stories" or world views that organise, explain, and account for everything and everyone—that structured modern thought. Some theorists argue that modernism's emphasis on universalism, rationality, and progress has produced intolerance and violence (since it implies the suppression of everyone who disagrees—or doesn't fit—with this picture). Others (see Horkheimer & Adorno, 1972, for example) go even further, arguing that modern thought produced the separation of science and morality which led to some of the 20th century's worst excesses—two world wars, Hiroshima, and Auschwitz, for example.

Postmodern thought is also characterised by a loss of faith in the idea of progress. If there is no longer one single "true" picture of the world, then the idea that, by filling in more and more of the pieces of this picture, we will eventually have the "full" picture, is difficult to sustain. Instead, there is an emphasis on multiplicity and plurality: the idea that a number of different perspectives are possible, each of which has its own internal "rules of the game", that these can sit side by side, and that it is possible to move between these different perspectives.[4] While change is inevitable, this is not seen in terms of linear progress toward some known endpoint, but rather as a series of "networks and flows",[5] a series of connections and reconnections that can't be predicted in advance, and that, because they are constantly forming and re-forming, don't last long enough to, as Bauman puts it, "solidify". Thus, instead of emphasising direction, order, coherence, simplicity, and control, postmodern thought emphasises fragmentation, diversity, discontinuity, contingency, and pragmatism. In addition, and importantly for the purposes of this paper, there is an emphasis on process and relationships, on what happens in the spaces *between* entities, not on the entities themselves.

This shift came about at least in part as a result of various developments in science. As more and more theories, concepts, and models are developed it is impossible for any one scientist to keep themselves informed about all of them, even in their own discipline. Writing about this in the 1960s, Thomas Kuhn used the term *incommensurability* to describe the way in which even apparently basic terms like *time* and *space* are used in completely different ways in the different disciplines of science, with the result that translation—and communication—between the different disciplines of science has become impossible (see Kuhn, 1970). This inability of the different disciplines to "talk to" each other calls into question the belief in scientific progress as the gradual completion of a true, generally agreed on, universally applicable picture of the world. If such a picture existed, the different theories would function as the pieces of a jigsaw puzzle, the parts gradually fitting together to form a whole. The incommensurabilty of the pieces means, however, that they simply don't fit: each is a fragment, one part of its own independent system, not part of some overall whole.

The same kind of fragmentation can be seen in the wider culture. For all sorts of reasons (the Internet, other media, increased possibilities for travel, and so on), people in the late 20th and

---

[4] This idea does not, as some commentators have argued, imply that postmodern thought is necessarily relativist: that is, that one perspective is as good as another, that "anything goes", and nothing matters. For a fuller discussion of this issue, see Chapter 6 of Gilbert (2005).

[5] This term comes from Castells (2000).

early 21st centuries are exposed to an enormous number of different ideas, cultures, religions, styles, ideologies, and so on. They are not necessarily "part of" just one culture, set of ideas, or style, nor do they choose to identify themselves with just one: they mix and match elements from many, identifying themselves via their own particular "pastiche" (see, for example, Jameson, 1997; or Harvey, 1990). Some commentators have argued that one of the defining features of postmodernism is that nothing is actually new any more: all of its so-called inventions—DVDs, cyberspace, the new biotechnologies, for example—simply parasitise and repackage the originality of past ideas. Others characterise postmodernism as a period of ever-increasing uncertainty, complexity, "not knowing", and confusion.

If we accept all this, there are major implications for how we think about—and what we need from—our education system. This shift away from order, certainty, rules, and frameworks requires that we are all able to think for ourselves—at a far higher level of intellectual abstraction than was probably necessary in modernism. Knowing the "rules of the game"—how to think, behave, follow the rules, and do things in the (old) mainstream system is no longer enough. The 21st century citizen needs to be able to think and work in—and across—many different systems, often at the same time. To do this, they need certain intellectual skills—the ability to assess and evaluate information; to analyse, synthesise, and create new knowledge, for example—in a highly developed form. Developing these abilities is, of course, an important goal of the current system, but in this system they are (at least in theory) developed *implicitly*, through the study of the traditional disciplines (mathematics, science, history, English, and so on). A 21st century education system needs to *foreground* the development of these abilities. The shift into postmodernism also has implications for how we think about learning: what it is; what aspects of it are important; what metaphors we use to think about it; and how we assess whether or not it has taken place. This, I argue in the last section of this paper, has implications for how we might think about "progress". However, before turning to this, we need to look at another major aspect of postmodern thought, one that is highly significant for educationists because it forces us to think differently about how and why we might go about achieving one of public education's major goals, the goal of "reducing disparity".

In the first part of this paper, I referred to the way modern individuality excludes certain classes of person, and the related problem of equality. The latter part of the 20th century saw various attempts to deal with this problem, which were fraught with difficulty. The early days of the various 20th century liberation movements (black liberation, women's liberation, gay liberation, and so on) were characterised by demands for equality and equal rights. While these demands eventually achieved formal equal rights, in practice very little changed. Because, as outlined earlier, Western political thought is underpinned by the one-size-fits-all model of individuality, equality effectively means "sameness". Achieving equality thus means becoming the same as—or assimilating to—the one-size-fits-all model. It is not possible to be the same *and* different: difference is not recognised *as* difference, but as deficiency, lack, not "the real thing", not "having what it takes". The recognition of this produced forms of political activism that focused, not on equal rights and sameness, but on arguments that claimed that the groups they represented were different from—and better than—the "standard" individual of modern political thought. One result of this was the proliferation of a whole range of new and different "identity categories"—women, Māori, gay people, differently abled, and so on—each of which then had to make a case for inclusion in competition with the representatives of a whole range of other identity categories. There were arguments about which oppression "came first" or which was the worst. Alliances between these groups weakened, and separatism became the main political strategy. Those who

were critical of it called this approach "identity politics" or "essentialism" (see, for example, Fuss, 1989).

This work was important because it made it obvious that the standard one-size-fits-all model of individuality does not in fact fit everyone, but actively excludes a significant proportion of the population. However, it did not solve the problem of how to argue against injustice from a position of difference. It is logically impossible, in modern political thought, to claim equality while at the same time also claiming difference. Postmodern political thought developed out of this difficulty. Postmodern political theorists are interested in the concept of difference. They are interested in thinking about it, not as the antithesis of equality, but as an interdependent term. Emerging out of this work are some interesting new ways of thinking about individuality and identity, and some new ways of thinking about politics and political activism. This work, it seems to me, is highly relevant to educationists.

Postmodern political theorists say we should move away from the one-size-fits-all model of individuality and equality. They think we should look for new and different ways of thinking about individuality, ways that allow difference to be expressed—*as* difference, not as deficiency, lack, or exclusion. This approach is known as the "politics of difference"—to distinguish it from the "identity politics" it rejects. Postmodern scholars do not, as some claim, say that we should abandon egalitarianism and individualism. They argue that we can't just reject these ideas because we don't, as yet, have any other ideas to replace them. Moreover, these ideas have structured our thinking for so long that it is not easy to think without them. However, this doesn't mean that we have to treat these ideas as givens: we can, as one theorist puts it, put them "under erasure" (to signify their inadequacy), or we can deconstruct them.[6] We can dig up the assumptions that underlie them, have a good look at them, and, where necessary, change bits of them. Some scholars (see Fuss, 1989, for example) argue that the act of doing this will, on its own, be enough to produce change. They say that simply bringing the assumptions that underlie an idea out into the open changes the way the idea "works".

Postmodern political theorists want to think about egalitarianism and individuality, not as absolute ideals that can be defined ahead of time and applied to everyone, but as concepts that form a starting point for discussion and debate. They focus on developing contexts and processes that allow debate and provide space for previously excluded groups to have a voice. Their goal is to find ways to make difference part of the debate, not so that it can be assimilated (or made "the same"), but so it can just "be"; so it can express itself, not in relation to existing norms, but on its own terms. This, they say, will allow previously excluded groups to work out their own different conception of individuality. Thus the goal is to rethink equality and individuality in ways that allow us to recognise different "ways of being", and different ways of knowing and learning. Instead of creating new categories of difference that then have to compete with each other, postmodern political activism aims, as one theorist puts it, "to work difference together",[7] to build relationships that allow the partners to acknowledge and genuinely recognise each other's differences as difference, not deficiency.

So: if we are in fact seeing a shift away from the "universalist" notion of equality to an emphasis on plurality and difference, on multiple and diverse "ways of being" coexisting alongside each

---

6 The terms *under erasure* and *deconstruction* originate in the work of Jacques Derrida. See Norris (1982) or Sarup (1988) for introductory accounts of these concepts.

7 Yeatman (1994, p. vii). See also Yeatman and Wilson (1995) and Young (1990).

other, what does this mean for a public education system which was founded on—and which is, at least in theory, legitimated through—the goal of equal opportunity? For reasons that I hope are now clear, we can't just abandon this goal, replacing it with an emphasis on diversity. The equality goal has organised our thinking for so long that we can't easily think outside it. We can, however, put it "under erasure". We can draw attention to its problematic status: we can look for ways to think around, past, and beyond it, for ways to see the education landscape through a different lens with different things in focus. We need to do this because, as I hope I have shown, modern education systems do not—and cannot—actually produce equality. Rather, they screen and sort people according to their likely future employment, and the "progress" metaphor plays an important role in this process.

How then should we begin the process of moving around and beyond these 20th century ideas to lay the foundations for a 21st century education system? In the last section of the paper I suggest four ideas that, taken together, could help us to start thinking beyond the 20th century one-size-fits-all model of education. None of these ideas are new and all are extensively elaborated elsewhere. There are of course many other ideas that could be used to do this: however, I have chosen these particular ideas because they are already "out there". Two are strongly signalled in the recently released *New Zealand Curriculum* document (Ministry of Education, 2007) and so are—or soon will be—well known to teachers. However, it has to be said, this familiarity is a strength but it is also a trap. If these ideas are simply added—and subsumed—into the existing system, there will be no real change. Their meaning will be subverted and diluted and, eventually, in a few years, people will be saying that the new ideas didn't work, that they were meaningless jargon, and we will be hearing calls for a return to "the basics"—the traditional subjects and the traditional ways of teaching them. If these ideas are to help us rethink our education system for the 21st century, they need to be understood alongside—and in the context of—all the other ideas that go with 21st century thinking.

## Learning and "progress": How are they different in the 21st century?

The first of the four ideas I want to look at here is the notion of building learning "capacity"—not so that people can know more stuff, but to give people the ability to go on learning—in a variety of different situations, by themselves or with others, for the rest of their lives—when there are no longer teachers (or parents) to chase them.

While educationists have long talked about developing people's capacity to learn, we have not been especially successful in actually doing this. Helping students "learn how to learn" has usually involved giving them hints and tips for efficiently organising, retaining, and retrieving information, and "successful" learners have been those who do well in examinations. This approach does not, of course, build learning capacity and/or the ability to be a lifelong learner (in fact, as is well documented, it usually has the opposite effect).

In addressing this question, Guy Claxton proposes a model based, not on learning "skills", but on what he calls learning "dispositions": building the ability and the disposition to learn—that is, being ready, willing, and able to make use of that ability. As Claxton points out, teaching young people to be "disposed to" learning requires a significant culture change in schools. It requires a different orientation: different kinds of activities; a different language; and different forms of assessment. It requires us to put learning in the foreground (and knowledge acquisition in the

background), something that, he cautions, will be more difficult than it looks (see, in particular, Claxton, 2007; but also Claxton, 1999, 2002). In particular, the assessing of students' progress in building their learning capacity requires new models of assessment capable of measuring *learning* building, not knowledge building.

The second idea is that of a curriculum that is designed, not to "fill students up" with certain highly specific forms of knowledge, but to build certain core "competencies" (of which the "disposition to learn" could be one). This idea is an important foundation for building a 21st century education system, firstly, because working out what these competencies should be forces us to discuss—and agree on—what things we think *everyone* needs, the things we think *everyone* should take away from their time at school. We can then focus on building these things, in many different ways and contexts, in learning programmes that can be personalised for individual students and their needs. Secondly, this idea allows us to rethink the relationship between traditional disciplinary knowledge and the school curriculum: in particular, the traditional focus on this kind of knowledge as both an end in itself and a means of screening and sorting people. Assessing the extent to which competencies are developed requires new ways of thinking about "progress", not as a linear, step-by-step process of concept building that eventually produces the "full picture", but (probably) as the development of deeper and richer networks and links between different aspects of the competency (whatever it is); a process which is achievable via many different pathways.

The third (related) idea is that of "thinking together"—the importance of building people's capacity to work, think, and *be* together productively in ways that build the cognitive capacities of all those involved, both individually and collectively. This idea is not at all new: cognitive scientists have long seen learning as primarily a collective activity that involves making connections and processing information in connected systems.[8] However, its influence has yet to be seen in the day-to-day practices of school classrooms, largely (I think) because of the predominance of the 20th century notion of learning as an entirely individual matter. Scaffolding the development of collaborative teams of learners requires a completely different emphasis from the traditional approach (which was designed to—eventually—produce independent, autonomous thinkers). Assessing student learning in this context requires us to completely rethink our ideas about what we are looking for and why.

The fourth idea I want to suggest here is the idea that developing a 21st century education system involves redesigning the system so that it focuses, not on sorting out who "has what it takes" to go on to higher education, but on finding ways to ensure that everyone achieves. This idea requires us to abandon many core 20th century ideas—including the notion of general intelligence or "ability" as being made up primarily of the traditional mathematical–linguistic abilities. It requires us to go back to basics in terms of our ideas about "achievement". What does this term really mean? What do we want it to mean? What is progress? How do we know if someone is moving forward, getting better, continuing to grow and develop? Do we *want* it to mean that they know (or appear to know) more—and harder—stuff; that they "have" certain concepts or can do certain things that show they have them; and/or that they have more, less, or about the same kinds of

---

[8] These theorists commonly use metaphors derived from computer science: for example, some talk about *distributed* learning, *parallel distributed processing*, or *cluster matrix* models, while others use terms like *connectionism*, *situated cognition*, and *learning communities*. (See, for example, Brown, Collins, & Duguid, 1989; Brown & Duguid, 2000; Lave, 1988; Lave & Wenger, 1991; Rogoff, 1990; Rogoff & Lave, 1984; Rumelhart, 1989; Rumelhart, McClelland, & the PDP Research Group, 1986.)

concepts as others at the same age and stage? What does it mean to measure progress in this way? Alternatively, could we see (and perhaps measure) "progress" in terms of the level of complexity of what a person is able to *do* with those concepts in real-world situations where those concepts really matter. Rethinking our ideas about progress might require us to use different metaphors. For example, the cumulative, construction-oriented "bricks in a wall" metaphor of knowledge building could be replaced by one involving increasingly richer and more complex networks and flows—a spider's web perhaps, or nerve cells in the brain, something dynamic and constantly changing as it responds to new inputs and connections.

These ideas are of course just ideas. I am by no means an assessment expert, so will not attempt to speculate beyond this. My point here is not to suggest how we might go about assessing the extent to which students have acquired whatever it is we decide is valuable and necessary in 21st century learners. Rather it is to map out some of the background against which we need to begin the work of deciding what these things should be. However, whatever it is we decide to measure, I'm not at all sure that we should continue to use the term *progress*. As I hope I have shown here, it is a term that is very deeply imbued with the modern preoccupation with moving forward in a particular direction, towards a particular, known-in-advance goal. However, because we don't, as yet, have another, better word (and we probably won't until we try to do some of these things), in the meantime all we can do is to signal its problematic status.

## References


Bauman, Z. (2007). *Liquid times: Living in an age of uncertainty*. Cambridge, UK: Polity Press.

Bleier, R. (Ed.). (1984). *Feminist approaches to science*. New York: Teachers College Press.

Bordo, S. (1987). *The flight to objectivity: Essays on Cartesianism and culture*. Albany, NY: State University of New York Press.

Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Brown, J., & Duguid, P. (2000). *The social life of information*. Boston: Harvard Business School Press.

Castells, M. (2000). *The rise of the network society* (2nd ed.). Oxford: Blackwell.

Claxton, G. (1999). *Wise up: Learning to live the learning life*. London: Bloomsbury.

Claxton, G. (2002). *Building learning power: How to help young people become better learners*. Bristol, UK: TLO.

Claxton, G. (2007). Expanding young people's capacity to learn. *British Journal of Educational Studies*, *55*(2), 115–134.

Fuss, D. (1989). *Essentially speaking: Feminism, nature and difference*. New York: Routledge.

Gilbert, J. (2005). *Catching the knowledge wave?: The knowledge society and the future of education*. Wellington: NZCER Press.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575–599.

Harding, S. (1986). *The science question in feminism*. Ithaca, NY: Cornell University Press.

Harvey, D. (1990). *The condition of postmodernity*. Oxford: Blackwell.

Hawking, S. (1988). *A brief history of time*. London: Bantam.

Horkheimer, M., & Adorno, T. (1972). *The dialectic of enlightenment*. New York: Herder & Herder.

Jameson, F. (1997). *Postmodernism: Or, the cultural logic of late capitalism*. Durham, NC: Duke University Press.

Keller, E. F. (1985). *Reflections on gender and science*. New Haven, CT: Yale University Press.

Kuhn, T. (1970). *The structure of scientific revolutions*. Chicago: Chicago University Press.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, MA: Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lyotard, J-F. (1984). *The postmodern condition: A report on knowledge*. Manchester: Manchester University Press.

Ministry of Education. (2007). *The New Zealand curriculum for English-medium teaching and learning in years 1–13*. Wellington: Learning Media.

Norris, C. (1982). *Deconstruction: Theory and practice*. London: Methuen.

Pateman, C. (1988). *The sexual contract*. Stanford, CA: Stanford University Press.

Pateman, C. (1989). *The disorder of women: Democracy, feminism, and political theory*. Cambridge, UK: Polity Press.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in a social context*. Cambridge, UK: Cambridge University Press.

Rogoff, B., & Lave, E. (1984). *Everyday cognition: Development in social context*. Cambridge, MA: Harvard University Press.

Rumelhart, D. (1989). The architecture of mind: A connectionist approach. In M. Posner (Ed.), *Foundations of cognitive science* (pp. 133–159). Cambridge, MA: MIT Press.

Rumelhart, D., McClelland, J., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. (Vol. 1: Foundations).* Cambridge, MA: MIT Press.

Sarup, M. (1988). *An introductory guide to post-structuralism and postmodernism*. New York: Harvester Wheatsheaf.

Tuana, N. (1989). *Feminism and science*. Bloomington, IN: Indiana University Press.

Walkerdine, V. (1988). *The mastery of reason; cognitive development and the production of rationality*. London: Routledge.

Walkerdine, V. (1989). *Counting girls out*. London: Virago.

Walkerdine, V. (1984/1998). Developmental psychology and the child-centred pedagogy: The insertion of Piaget into early education. In J. Henriques, W. Hollway, C. Urwin, C. Venn, & V. Walkerdine (Eds.), *Changing the subject: Psychology, social regulation and subjectivity* (pp. 153–202). London: Routledge. (First published in 1984 by Methuen.)

Yeatman, A. (1994). *Postmodern revisionings of the political*. New York: Routledge.

Yeatman, A., & Wilson, M. (Eds.). (1995). *Justice and identity; antipodean practices*. Sydney: Allen & Unwin.

Young, I. M. (1990). *Justice and the politics of difference*. Princeton, NJ: Princeton University Press.

# Participants' perspectives

Participants were asked to get together in groups to discuss the implications of the day's discussion for practice, policy, and research. This section seeks to convey the key points the participants noted down, and to capture the overall flavour of the discussion.

A strong common theme across practice, policy, and research was the need for improved connections between what is going on in the early childhood sector, primary and secondary schools, and the tertiary sector, so that opportunities are created for the different sectors to learn from each other.

Time for reflection and learning also came up repeatedly—not just for teachers but for the whole community, so everyone can be better informed about assessment measures, issues around notions of progress, and future directions in education.

## Implications for practice

### Time, resources, and training

Most groups noted the need for teachers to have the time, resources, and training to help them better understand the various assessment tools and the issues involved in measuring progress, and to enable them to carefully plan which instruments to use for the defined purposes or goals that the school is interested in. Several groups stressed the need for practical, school-based training. They saw it as important for the credibility and morale of the profession for teachers to have a better grasp of the complexities surrounding assessment. One of the reasons given was that teachers would then be informed enough to keep questioning the purpose of doing more assessment.

Participants said that increased reflection time would allow teachers to deeply consider the purpose of assessment tools. This needed to include thinking about what schools would do with the information they gathered and how to weigh up the time spent testing versus teaching. "How do we develop a vision of progress?" asked one participant. Another posed the question: "What do these forms of assessment objectify as the end outcome of learning?" Some also felt each school needed a dedicated person to co-ordinate assessment data collection and use.

### Involving students and the whole community

Several groups brought up the importance of hearing from all communities within society about what is considered important in terms of measuring progress. That would allow schools to make changes that reflected their own learning community. Schools needed a shared vision with stakeholders—students, teachers, the board, etc.—about what assessment data represent and how they can be used. One group made the comment that the community needed to be educated about

what pedagogical approaches contribute to progress. Another group stated that testing analyses should only be used to inform teaching.

### Individualised assessment

"Can I provide individualised learning programmes for 30 students?" asked one participant. "I can, but not well."

Several groups raised the place of the student voice and the concept of "co-construction" when it comes to making choices about assessment. One group mentioned that the use of portfolios is very effective at the tertiary level to encourage students to reflect on their own progress.

Other groups mentioned portfolios as providing powerful evidence of learning, but one made the comment that they were time-consuming. There was interest in exploring Margaret Carr's quadrant model and the relationship between and among the quadrants in a way that linked into children's lives.

## Implications for policy

A number of groups had a firm message for policy makers and politicians: no national testing. Participants felt a national testing regime would be narrow and simplistic, and that it would have the potential to obscure what is really going on in schools. It would inhibit what the new curriculum was all about, one group commented.

A group commented: "We don't want a culture of testing for the sake of testing. It must be purposeful and meaningful."

Several groups felt it was up to the teaching profession to articulate what are informative and useful data.

A number of groups felt the discussion raised policy implications for teacher training, with one suggesting that the graduate diploma needed to be longer than a year in order for trainees to get to grips with the complexities of assessment and measurements of progress. More professional development for teachers in this area was also mentioned, with one group commenting that this was particularly necessary at secondary school level.

### The value of autonomy

A strong thread throughout most comments from participants was for policy to reflect the ability of schools to create their own needs-based curriculum. One group wanted to see policies that supported the individuality and diversity in the curriculum—"underpinned by the very sound pedagogical theories that are found in the new document".

Teachers needed to "retain autonomy as professional educators to decide on/develop processes that suit our specific learning communities". Schools needed suitable resourcing to facilitate rich assessment practices.

### 21st century learning

Participants felt policy makers needed to inform the government and parent community about 21st century learning priorities and practices. Significant professional development would be needed to strengthen the key new elements in the curriculum that reflect the future paradigm.

## Implications for research

Most of the suggestions here were for actual research projects that people felt would be useful. There were a number of comments about approaches to research and a call for more research funding. Several made the overall comment that research needed to draw better connections between the early childhood, primary, and secondary school sectors. There was also interest in research that included the student voice—"bring it back and make it fashionable". Student-voice research could investigate how students "make explicit their own learning tools".

One group wanted to see more projects involving researchers being partnered with teachers to look at new approaches to curriculum, learning, or teaching. Data-driven thinking should not be allowed to subvert the potential of the new ideas in the curriculum, one group commented.

### Suggestions and questions that groups put forward as material for research:

- The development of benchmarks for the key competencies. How do we measure progress? How can we assess the things that really count: resourcefulness, resilience, thinking skills, and strategies etc?
- What are other cultures doing in terms of use of assessment data?
- The effect on teaching and learning when teachers use assessment tools such as PATs or asTTle.
- How are schools using assessment tools to further children's learning?
- Research into success stories about what has shifted struggling students.
- The new curriculum and its implementation: Is the timetable achievable? Are schools being given enough resources? How will we know when we get there? How is the professional development being provided?
- More pilot studies of progress models in schools to build the evidence base.
- Research into how schools are utilising the new literacy progressions. Are they using them to inform teaching and learning?
- Focus on the needs of tomorrow's communities—what do we need to change in our current practice?
- Research on educating communities, which could include resources for schools, such as a DVD on future challenges.

## A final comment

A couple of groups urged NZCER to repeat the seminar in other parts of New Zealand, so that more teachers had access to the messages. And they wanted us to keep doing what we do: "Continue to be at the cutting edge—being provocative, challenging us."