REVIEW PAPER 1: Assessing the hard to measure

Abstract

This paper redefines the "hard to measure" as the "hard to define". Using a heuristic developed by Margaret Carr for NZCER's 2008 Making Progress, Measuring Progress one-day conference, it scopes the various types of curriculum outcomes signaled as important by The New Zealand Curriculum, and discusses the nature of curriculum uncertainties, particularly in relation to describing progression, that should be addressed before methods of assessment of these outcomes are debated. Drawing the threads of some very diverse curriculum outcomes together in the context of "thinking" as a key competency, the paper makes some recommendations about constructive entry points to ongoing curriculum and assessment conversations, and uses potential changes to the NCEA as a specific example of what could be achieved if construction of rich descriptors of progress is undertaken by teachers and others with curriculum expertise working in collaboration.

Introduction

What makes something hard to measure? Does the difficulty lie in the actual *process* of measuring, in which case the problems to be addressed are essentially technical, even if complex? Or does the difficulty lie in conceptual uncertainty about the nature of assessment targets? In this case the challenges entail:

- curriculum interpretation;

- clarification of the anticipated outcomes of learning; and

- subsequent description of the nature of evidence that would indicate the extent to which these outcomes have been met.

This paper takes the position that conceptual issues lie at the base of the uncertainties, and that these manifest as *curriculum* questions rather than technical questions of how to assess. This view is supported by recent commentary from an ongoing systematic OECD investigation *Understanding the Social Outcomes of Learning*:

> ..there are debates to be had over what a public service such as education is intended to achieve; logically this should be prior to, and shape, the measures to be used in ensuring accountability (OECD, 2007a, p22).

Some very big questions are encapsulated here. There is a debate to be had about how we balance and deliver on the mix national educational priorities signaled by The New Zealand Curriculum (NZC) (Ministry of Education, 2007a). How do we seek information about less familiar outcomes

(less familiar that is, in terms of what we have had experience of assessing in the past) as well as those that we already have the tools to assess? What are the implications for the assessment methods we promote if we want to compare achievement related to these very different curriculum priorities? (We might want to do this to sense of overall patterns of achievement, and how the various priorities inter-relate.) How do we assess for accountability in ways that allow for variation in ways schools meet the national priorities (assuming we take seriously the mandate to design a curriculum to meet local needs, as signaled clearly by NZC)?

Following logically from these questions is another set. For what *purposes* do we intend to gather assessment data, and how does purpose impact on the type of data to be gathered? Much has been written in the past decade about the power of using assessment to inform ongoing learning decisions. This is commonly called assessment-for-learning, in contrast to assessment-of-learning, although we should heed the caution that the two are not as distinct as sometimes portrayed (Harlen, 2007; Matters, 2006). In the context of the topic set for this paper, it seems worthwhile to ask whether the intention to gather assessment data that can inform ongoing learning (as opposed to summative reporting of overall achievement) will impact on measurement difficulties, and if so how? A related challenge concerns ownership of the assessment process. For example, there are clear curriculum signals that fostering life-long learning is an outcome to be valued. In that case inclusion of the learner at all stages of the assessment process becomes an important consideration (Aikenhead, 1997; Claxton, 2006).

The UK EPPI Centre recently undertook a systematic comparison of formal tests and assessment methods that rely on teacher judgement. Research evidence concerning the validity, reliability, impact of, and resources needed for these two broad assessment approaches was evaluated. The research team concluded that the overall balance of benefit falls on using methods that involve teachers in making judgements of their own students' work, for all but the criterion of reliability (Harlen, 2007). They also cited evidence that reliability can be enhanced when task design is kept relatively open but criteria used to judge the tasks clearly specify what evidence of progress would look like. That is, the specificity should relate to evidence of learning, not to the actual task used to generate that evidence. This recommendation leads squarely to the third overarching question to be considered in this paper.
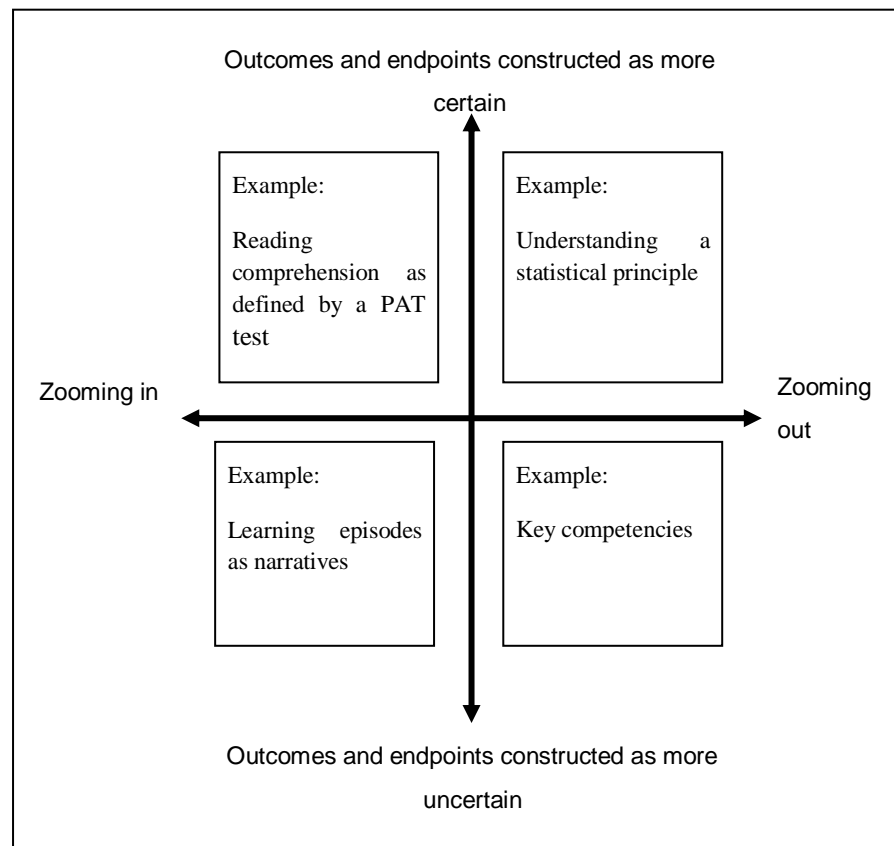
What does "making progress" look like when the focus is on aspects of learning that have not traditionally been assessed? This is the question fore-grounded in the main body of the paper, which scopes the nature of uncertainties to be addressed when seeking to describe the nature of progress. Aspects of the first two questions are integrated into this discussion as relevant. Two specific examples are then used to illustrate and integrate the issues raised. The first example takes "thinking" – one of five key competencies described in NZC – to discuss curriculum and assessment challenges inherent when intending to educate learners in ways that develop this competency. Arguably this is the most familiar of the key competencies in terms of what teachers would say they already do, and many ways of measuring it have already been devised. Any challenges raised here could expect to be multiplied when the less familiar competencies are considered. The second example takes the NCEA – the meeting point for assessment and

curriculum in the senior secondary school – to consider whether the possible approaches suggested by the analysis outlined in earlier sections could inform persistent assessment issues for the qualification.

The nature of making progress in learning

A framework devised by Margaret Carr for reflecting on the nature of progress has been used to organise the main part of the paper (Carr, 2008). The framework is useful because it explicitly identifies uncertainty of endpoints and pathways as an issue to be addressed when considering what evidence of progress could look like. It also takes into account differences of perspective when looking closely at focus. Relatively well defined and bounded learning outcomes are contrasted with "bigger picture" learning achievements whose boundaries are much more uncertain, and where there is less likely to be a shared understanding of what making progress can look like. These dimensions are organised as juxtaposed continua to make four quadrants as shown in the figure that follows. The examples included on the diagram are those identified by Carr. Other examples were identified during the development of this paper and are discussed in the sections that follow. This paper is mostly concerned with all *but* the top left hand quadrant (Quadrant One), although that is briefly discussed first to set the background for the discussion of the other three quadrants that, for different reasons, signal higher levels of uncertainty about the meaning of progress and hence the nature of assessment evidence that could be used to describe it.

Figure 1 **Outcomes and pathways: four quadrants (from Carr, 2008)**

What contributes to uncertainties in making progress?

Quadrant One: More certain outcomes/zooming in

In this quadrant, the focus tends to be on the individual learner and, typically, on behavioural and/or cognitive outcomes. This is a more familiar way of thinking about what evidence of learning can look like. A key assumption is that the construct being measured is well defined and can be measured relatively independently of context, providing the assessment tool is valid and sufficiently robust, and the construct can be reliably sampled (Delandshere & Petrosky, 1998). Additionally, the link between learning and outcomes tends to be framed within a more straightforward model where lines of accountability for students' progress are more evident (OECD, 2007a). The benefit of this clarity of focus is that existing theoretical knowledge is able to inform the types of outcomes that can be anticipated and hence assessed – there is a shared understanding of what evidence of learning can look like. Psychometric measures can pay attention to correlations between items, checking for internal reliability. Interactions between different components of the identified progression may also be made visible.

There can, however, still be uncertainty in this quadrant. Any progression is a *construction* – what it demonstrates will depend on how the "ruler" used to measure progress-as-defined is constructed (Darr & McDowall, 2008). No tool is totally reliable (Harlen, 2007). And, although well documented, teachers may not be as familiar with the subtleties of the progression as researchers, and may not have the necessary level of "assessment literacy" to read appropriate meaning into data generated by the relevant assessment tools.[1] This, along with a lack of appropriate pedagogical content knowledge may limit their ability to use the assessment results to inform next learning steps, which will be important if assessment for learning is intended. It follows that the *meaning* of assessment results will likely require translation for students and parents,[2] if they are to be more actively involved in co-construction of strategies for making progress.

Quadrant Two: More certain outcomes/zooming out

Some aspects of certainty are carried over from Quadrant One. These include the focus on the individual learner, on mainly cognitive outcomes, and on assumptions of linear connections between learning inputs and demonstrated outcomes. This is not to say that other types of broader outcomes cannot be assessed using relatively conventional tools – provided the construct of interest is well defined. For example, NZCER has recently developed an assessment tool to measure engagement with school in a generic sense. The construct of engagement is well theorized, meeting this quadrant's specification of more certain outcomes. However, as in Quadrant One, uncertainties do arise here when working out the *meaning* of the data patterns generated. Does the considerable difference in patterns of engagement of different classes in the same school (as found during trials) reflect primarily on teacher differences, or are there contextual differences in play that should be more carefully investigated? What does a drop in engagement from year 8 to year 9 mean?[3] What should schools do about this? Again the challenges are not in creating the assessment as much as in making meaningful and productive use of the information gained.

"Zooming out" to broader outcomes also creates new challenges for describing progression, and hence for assessment. The following discussion suggests that the certainties identified here are tempered by a range of current uncertainties, where questions need to be framed for ongoing analysis and debate. Overarching the discussion of the nature of progress from the perspective of this quadrant is the idea of *transfer* – that separate acts of learning assume richness and depth when linked to existing understandings in meaningful, productive ways. For example, seeing *relationships* between ideas, skills and learning experiences – what Anat Zohar calls "connectedness" (Zohar, 2006) - is one dimension of zooming out where progress could be

---

[1] The topic of teachers' assessment literacy is covered in another of the papers commissioned for this review.

[2] Again, this is addressed by another of the commissioned papers for the review.

[3] The everyday notion of progress tends to imply a growth or forward movement – here we see an indicator moving *backwards*, which challenges thinking about progress as getting more of something and suggests that progress can be about changes in a total configuration – so some things diminishing.

described. Indeed, Carr suggests that progress here could be seeing new patterns and order where previously there were seemingly unrelated pieces of knowledge (Carr, 2008).

The SOLO[4] taxonomy theorises progression in these terms, providing a useful foundation for the development of assessment tools (Biggs & Collis, 1982). An "extended abstract" response at the most developed level of SOLO captures the essence of deeper understanding and flexible use of pieces of knowledge, but translating the *generic* levels of SOLO to descriptions of progress in understanding specific sets of concepts in one or more specific learning areas requires careful empirical investigation. To cite just one example of this need, recent experiences with unreliability of NCEA achievement standards, whose different levels express the idea of deeper or more flexible thinking in generic terms that typically borrow from Bloom's taxonomy, would suggest that it is not self-evident what evidence of such "deeper thinking" can look like in different curriculum learning areas and at different curriculum levels.

A related issue, identified by both Zohar and Carr, is that acts of relating or linking are dependent on the learner also zooming out to the bigger questions within disciplinary and topic-based *contexts*:

> Grasping what a concept or principle means depends in considerable part on recognizing how it functions in its broader context. The context can be that of a topic, an entire discipline, or even multidisciplinary. In addition it is essential to develop a sense of how the discipline works as a system of thought. For example all disciplines have ways of testing claims and mustering proof-but the way that is done is often quite different from discipline to discipline (Zohar, 2006, p.1587).

This way of framing the bigger picture of progress adds considerably to the uncertainties already identified. The key competencies of NZC can be seen as having a meta-dimension that draws attention to disciplinary differences in ways of thinking. For example, the shaping of the "Nature of Science" integrating strand of the science curriculum was explicitly intended to model elements of key competencies as these relate to learning in science (Barker, Hipkins, & Bartholomew, 2004), but it is by no means clear that teachers understand the intent and potential of this strand – or that they know how to teach for the unfamiliar nature of science (NOS) outcomes specified there. Similar challenges have arisen when evaluating mathematics teachers' use of algebraic calculators (CAS) with year 9 and 10 students. CAS technology can help students make higher-order connections between algebraic processes and concepts by removing the lengthy computational process that could hinder such pattern formation (Kutzler, 2003). However, before this potential can be realised, teachers need to have this purpose in mind, and to see such patterns and connections in the "nature of mathematical thinking" for themselves (Neill & Maguire, 2006).

These uncertainties of curriculum interpretation raise a vexed question. Is it reasonable to attempt to describe and then look for evidence of progress in what is not yet being widely taught? Teachers' lack of familiarity with teaching about disciplines as systems of thought means that

---

[4] Structured Observations of Learning Outcomes

their practical wisdom is less likely to be available as one potential source for collectively building understandings of what progress might look like.[5] While, in theory, evidence of "deeper understanding" and "rich connections" is empirically investigable, there are questions of curriculum interpretation to be addressed first.

Evidence of transfer can also be seen when students adapt what they already know and can do to respond to learning challenges. The OECD DeSeCo project described *taking appropriate action* in personally meaningful contexts as an important dimension of competency (Rychen & Salganik, 2003) and so this provides another strong frame for thinking about what could be observed when "zooming out". This type of zooming out is described in the Health and Physical Education learning area as "action competence" but research in this area shows it is the curriculum aspect most likely to be ignored, at least in classroom teaching (Robertson, 2005).

Carr has developed the idea of knowledge-in-use to describe progression as a "strengthening" of key competencies, demonstrated by their use in a wider range of increasingly unfamiliar contexts (Carr, 2006). More recently this has become the B (= breadth) in her ABCD descriptor of dimensions of strength. By definition, assessment here would require students to *do* something, so traditional pencil-and-paper-based assessment of conceptual knowledge would be an inadequate means of capturing evidence of achievement and/or progress. However standards-based instruments could describe dimensions of evidence of practical know-how, ideally in combination with indications of the disposition to act on that knowledge, provided these were clearly understood. Again the challenge would seem lie in the construction of sufficiently informative descriptors.

A different uncertainty arises at the point where both types of zooming out (making rich connections, acting in new contexts) are brought together in a more complex whole. Recently, experimental assessment items of this type have been constructed for the Assessment Resource Banks (ARBs). Each ARB item is, in effect, a small research project, as student responses are analysed and coded to take account of the sense they have made of the task. Illustrating the complexities this can reveal, a task for students in years 6-8 requiring a combination of drawing completion and simple single-sentence responses demonstrated that contextual knowledge can interact with concept development when learning about taking action to protect ecosystems. This suggests that a richer understanding of contexts *per se* should not be overlooked as a dimension of making conceptual progress (Hipkins, Bull, & Joyce, 2008).

Quadrant Three: More uncertain outcomes/zooming in

Moving below the horizontal line of Carr's heuristic, there is an important shift in the nature of the learning theories that underpin the discussion. Dimensions of socio-cultural learning theory are to the fore in both Quadrants Three and Four, and this contributes to uncertainties for

---

[5] Such input richly informed the development of the MOE's recently published literacy progressions for example (Ministry of Education, 2007a). Teachers' expertise in teaching for literacy *was* useful for this task because it is at the forefront of their practice and professional focus.

describing outcomes for which learning progress might be demonstrated. This type of framing requires analysis of progress to take account of the interplay between all the aspects of the context over which the learning is "stretched" (Carr, 2008). This is the D = Dimensions of her ABCD model of assessment. These dimensions include: the attributes (knowledge, skills, attitudes, values) the learner brings; interactions with other people; the artifacts available to support learning; and the context in which the learning is set. All can be either enabling or constraining, and can work together or work against each other.

Carr's suggestion is that we think of progress here as taking place in "pieces" – each one a contained episode that takes account of the interactions between all the dimensions just outlined. (We could similarly think of the ARB items mentioned above as pieces of progress within a learning area – pieces that take account of concepts and context). One metaphor Carr suggests is to think of progress as building a "dictionary of experiences" (Carr, 2008). One advantage of assembling progress as pieces of learning is that 'life-wide' learning can be taken into account in a way that is not possible with more traditional assessment tools that tend to focus on more formal aspects of schooling (OECD, 2007a). A challenge that arises, however, lies in ensuring that the pieces thus assembled do add to some more coherent whole. Issues of uncertainty here would seem to lie in constructing an overall *framework* within which pieces can be assembled, so that their *significance* as evidence of making progress can be evaluated. (In what ways does this episode contribute to an overall picture of progress? What is the meaning of the learning demonstrated – i.e. what are the implications for ongoing learning?). Thus, the questions of Quadrants 1 and 2 do not go away. Rather additional layers of uncertainty are added.

Like Quadrant One, Quadrant Three is intended to zoom in, so details of the intended curriculum should be used to construct the framework within which pieces of learning are assembled. Given the necessary resources, the challenge of assessing to describe progress that combines various pieces within a competencies framing can be met quantitatively. The PISA assessment approach is a good example. The specific intent of PISA is to assess literacy in its broadest sense, as demonstrated in the following "science literacy" frame work (OECD, 2003).
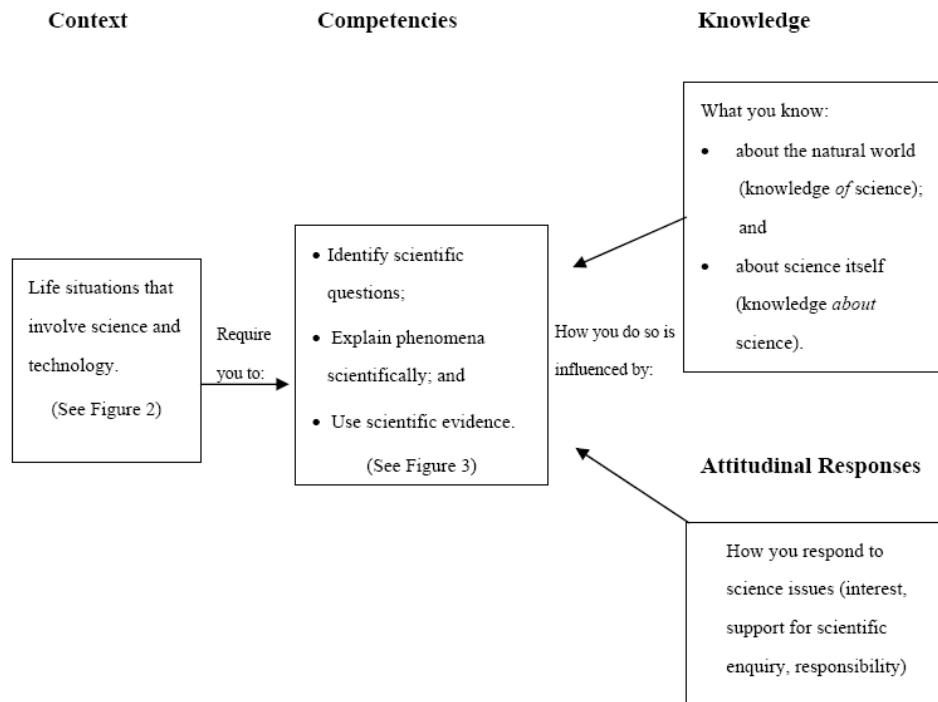
Figure 2 **The framework used for the 2006 PISA assessment of scientific literacy**

Again we see that the framework is somewhat generic, although it specifies a wider range of types of learning outcomes than are traditionally assessed by any one assessment tool located in Quadrant One (use of evidence, problem solving, investigation skills) and includes some dimensions already discussed in Quadrant Two (deeper understanding as "scientific explanation"). It also addresses some aspects of "being", as opposed to "knowing", that will be further discussed in Quadrant Four (interest, taking personal responsibility etc). All the uncertainties discussed so far about what making progress can mean continue to apply, although the details of the types of evidence being sought change in some respects. In that case, developing detailed assessment questions requires further curriculum guidance. This paper has already noted teacher uncertainties about what teaching for "Nature of Science" understandings means (here described as "knowledge *about* science" in the top right hand box of the framework). Identification and description of evidence of progress are more likely to be achieved by those with wider science education expertise than by classroom teachers. Indeed that was the case in PISA, where both the framework and the items were developed by international teams of science education experts.

Using mostly closed questions and well established item-response assessment techniques, PISA integrates the various conceptual/knowledge-based dimensions of the framework into a six-level scientific literacy scale. Each level has a multi-dimensional descriptor and there was no attempt to describe progress for each component separately. Rather, the ways the various dimensions of the

framework come together are modeled as one entity. Because the 2006 PISA also included a minor focus on literacy and numeracy, validity issues such as the impact of basic literacy levels on ability to demonstrate scientific literacy could be evaluated. Attitudinal components were modeled as a separate scale.

Nationally standardised assessment tools that fit Quadrant One dimensions of progress tend to eliminate context as a variable as far as possible, in the interests of validity. Any such tools developed for Quadrant Three face the challenge of putting contexts back, without compromising validity in the process. This was an issue for PISA, where many potential items were eliminated because different nations objected to contexts on the grounds of likely unfamiliarity for their students.[6] Reviewing research of the use of more formal (test-type) assessments of complex outcomes, Harlen comments:

> It is legitimate to ask at this point whether it is indeed possible to create test items that assess application, problem solving, critical thinking, and so on. Perhaps surprisingly the answer is positive – but with a caveat (Harlen, 2007, p.33).

Harlen cites PISA as one example of what is possible. Here in New Zealand, NEMP tasks could be another example. But Harlen's caveat is that the vagaries of the *context* used can create reliability challenges. She cites UK research that shows little correlation between an individual student's demonstrated abilities for seemingly the same capability, when set in different contexts. She concludes that "scores of individual students in these surveys are not relevant and become meaningful only when combined with those of other students in the sample" (Harlen, 2007, p.33). Since the act of "zooming in" in this quadrant must, by its socio-cultural framing, take contexts into account, the implication is that patterns of progress for populations of students could be monitored with more traditional test-type assessment instruments, but progress for individual students ought not to be assessed in this way. This draws attention to questions of purpose. Assessment for accountability is possible but assessment to be used with individual students to inform ongoing learning will need to use different assessment methods.

How then, might an individual student's progress be tracked? Carr recommends the use of assessment strategies such as learning stories or portfolios. To those we would add the use of tools such as individual Assessment Resource Bank items (ARBs). Each one constitutes a mini-research project, signaling to teachers the likely difficulty of the task while anchoring it firmly in the context of intended outcomes and next learning steps. All these approaches are able to document separate "pieces" to a wider learning framework, each piece complete with all its contextual detail, and so progression could be described as "progress in pieces" (Carr, 2008). Separate pieces are unlikely, however, to make sense of overall progress if the framework within which they are organised is not clear. The challenge of describing progress shifts in shape but does not go away.

---

[6] Personal conversation with the ACER-based item writing team, April 2007.

Research from several curriculum implementation projects has documented the development of school-based rubrics as frameworks to describe progress in a more contextually nuanced manner (see for example Boyd & Watson, 2006). In the absence of a widely shared conversation about what such progress could look like, schools have been doing so with more or less support and sophistication. It would seem sensible to coordinate this work to build on the more insightful descriptions of progress and to illustrate wider and/or deeper possibilities to replace instances where the parameters of achievement have been set too narrowly or at too low a level of expectation. This would require comprehensive analysis to develop national descriptors or benchmarks. These could even be organised as "progress maps" rather than as simple rubrics. This challenge is addressed by another of the commissioned papers for this review.

So far the discussion for this quadrant has continued to address assessment of the achievement of individual students. Within a socio-cultural framing, resources for learning include other learners. Future-focused studies emphasise the importance of learning to work with others as a skill set that is essential for working in so-called "knowledge societies" (see for example Gilbert, 2005). This imperative draws attention to the issue of assessment of work that is produced collectively, e.g. group work. This highlights questions about what constitutes evidence of learning; not necessarily new questions but rather questions that are particularly acute in this context. For example, if the focus is only on the product, and students contribute differentially, is it fair that all receive the same grade or mark? In any case, if the task design is sufficiently sophisticated, it may become impossible to tell who contributed what to the final product (for an example that discusses this in the context of fostering students' creativity see Sumara & Davis, 2006). This is important because the creation of knowledge "in the spaces" between people is a desired attribute of knowledge workers (Gilbert, 2005) that highlights the academic, not just social, possibilities of the key competency "relating to others". A focus on *process* instead of product will not necessarily ameliorate this issue, since students could be busily getting along with each other yet produce low quality work. Foreshadowing the discussion in Quadrant Four, we could also ask what making progress looks like in the context of group interactions. Again the challenge seems to lie in defining what it is we would seek to measure, before discussing how to do so.

Quadrant Four: More uncertain outcomes/zooming out

This quadrant returns the focus to "zooming out", and to outcomes that most people would say they value, yet that have seldom been the focus of specific assessment and documentation. These are outcomes related to students' "being" in the world, as opposed to what they know and can do.[7] For example, the OECD group commissioned to document relationships between education and social outcomes noted that, notwithstanding its demonstrable success and impact:

> PISA deals with only one age band, and it focuses on skills and competencies, not on what happens as a consequence of the learning (OECD, 2007a, p.21).

---

[7] This is not a new idea – in the mid 1990s UNESCO's widely cited Delors report identified "learning to be" as one of four pillars of learning for adult life in the 21st century (Delors, 1996).

Complexity thinkers would say the outcomes of learning are *emergent* and therefore, at least to some extent, not able to be fully predicted in advance (see for example Davis, Sumara, & Luce-Kapler, 2008). Here then, the uncertainty is not just a function of the not-yet-described. Uncertainty is inherent in the framing and must be allowed for when documenting progress. There are obvious challenges for assessment policy and practice.

The OECD authors cite *tolerance* as one example of a value that is often seen as a desirable outcome of schooling because of its contribution to social cohesion. But they also note there are debates about what tolerance actually looks like when enacted (OECD, 2007a, p.23).[8] These comments highlight several important new dimensions of uncertainty to be taken into account when considering issues of progress related to Quadrant Four. The first is that the hard-to-measure here may be the hardest to describe in the first place. It is not just that we have less experience of assessing such outcomes, although that is certainly the case. The greater challenge could be a lack of consensus about what aspects of the broadly specified outcomes we should seek to document, if any.

A related challenge is to describe the antecedents of tolerance as enacted in adult life. What does tolerance look like in a five year old, compared to say, a fifteen year old? Does its manifestation change or develop? Robert Kegan's theory of development in each individual's meaning-making system suggests that it is not necessarily the *observed* act that changes, but the underpinning motivation and reasoning (Kegan, 1994). Following Kegan's theory, our in-house conversations have led us to question the evidence provided by observed behaviours, if those are taken in isolation from other aspects of the context. Depending on the meaning-making system being used, the same act could be a behavioural/compliance response, or it could be indicative of growing competency to self-regulate one's actions and decisions. There is a deep conversation to be had here, but we have barely begun to ask the relevant questions.

Yet another challenge concerns the assumption that development necessarily implies some sort of future-focused upwards trajectory – i.e. something gets better, or there is more of it, or it happens more often. But if development is viewed as "flexible and appropriate adaptation" to the *immediate* situation, then:

> Different images might be more appropriate to describe development, ones that imply recursive cycles and feedback loops. One simple possibility is a cyclist moving over a varied terrain. Depending on the demands of the moment, the cyclist will shift gears - in effect selecting the most contextually appropriate manner of dealing with a particular landscape. Similarly we readily "shift" from on mode of thinking/acting to another in response to new or difficult situations (Davis, Sumara, & Luce-Kapler, 2008, p.51)

---

[8] Illustrating this, messages about tolerance could be read into several of the values specified in NZC, for example valuing: diversity as found in our different cultures, languages and heritages; equity, through fairness and social justice; and respect for self, others and human rights (p.10).

Changing to a lower gear here could be seen as "going backwards" if evidence of an ever-upward trajectory is sought. Here in New Zealand, the longitudinal Competent Children, Competent Learner study has drawn on detailed empirical data about individual students, tracked over their years of schooling (Rivers et al., 2006; Wylie, Hipkins, & Hodgen, 2008), with the implication that this developmental dynamic could be seen as more like a climbing frame than a ladder.

In a similar vein, the Competent Children, Competent Learners project has demonstrated strong correlations between enjoyment of reading and many other markers of academic success (Rivers et al., 2006). The challenge of "zooming out" here is not so much in specifying the assessment tool, as in determining whether *progress* is necessarily an appropriate way to think about desirable outcomes. Some students will be challenged to *maintain* an early love of reading when what is read becomes more intellectually demanding. (The same comment could be made about strengthening any of the key competencies).

Taking a future-focus on *prospective* outcomes is another challenging source of uncertainty. What will students actually do with the learning experiences they have been offered as they move forward in their lives? How will their learning benefit both them personally and society more generally? There are important considerations here for policy developers but is it realistic to even consider future-focused outcomes as potentially assessable while students are still at school? Carr addresses this challenge via two dimensions of her ABCD "dimensions of strength" assessment model: A = Agency; and C = continuity. Continuity, she says, links past, present and possible futures, and is "jointly constructed by teachers and learners, and families have a role in this as well" (Carr, 2008, p.12). The challenge here is to ascertain how students see their own futures, at least in terms of their sense of identity as a learner. This is an important dimension to be developed as part of "learning to learn" (see for example Claxton, 2006). The importance of this outcome is clearly signaled in the vision statement of NZC (Ministry of Education, 2007a, p.8), and by its inclusion as one of eight principles intended to apply across the whole curriculum (p.9). An obvious implication is that assessment must take careful account of the learner's views and perspectives on themselves as learners, not just of *what* they have learned.

Carr describes *agency* as entailing being able to take "an increasingly critical and innovative perspective on a key competency" and suggests it is "important to investigate the students' own perceptions of similar and different learning tasks and learning experiences" (Carr, 2008, p.11). There are elements of the "connected thinking" discussed for Quadrant Two" and "critical thinking" which might be seen as fitting Quadrant Three, but here they are combined with the sense of "being" that is a hallmark of Quadrant Four. This suggests yet another uncertainty - variables can be "reciprocally determined" and progress in several domains can be both cumulative and interactive:

> Empirically, little is known about the cumulative and interactive aspects of learning that occur in multiple contexts (life-wide learning) over the lifespan (lifelong learning) (OECD, 2007a, p.37).

A recent example of an assessment tool that tackles these challenges is the "Effective Lifelong Learning Inventory" (ELLI) developed by a UK team based at Bristol University. The tool describes seven empirically-determined dimensions that are predictive of students' "learning power". Each dimension has a positive pole and a negative one (in brackets): changing and learning (vs. being stuck); critical curiosity (vs. passivity); meaning making (vs. data accumulation); creativity (vs. rule bound); learning relationships (vs. isolation); strategic awareness (vs. robotic); and resilience (vs. dependent).[9] This is a self-report instrument and the seven dimensions are combined to produce a profile intended to help learners identify areas they may need to strengthen.

Figure 3 shows an example of a profile, taken from a Power Point used by the research team to make the point that this student could be challenged to work on their resilience. The intent here is clearly assessment for learning, not assessment for accountability, although presumably results can be combined to create a class or school-wide profile. However one of the creators of the tool has cautioned that it may suffer validity issues if students do not sustain a "conscientious engagement" through the long series of questions (Claxton, 2006, p.14). Obviously, checking with the student in person could ameliorate this when the tool is used for one-to-one feedback.
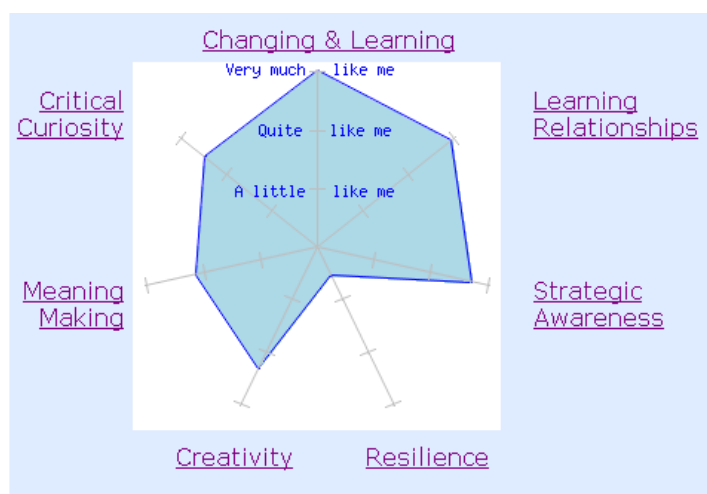
Example 2: Here it is resilience

Figure 3 **Example of ELLI report for an individual student**

---

[9] Slide sourced from http://www.antidote.org.uk/offer/LearningPower.ppt#274,19,Slide 19

If such an instrument was to be used for accountability purposes, for example demonstrating what the school had done to help students strengthen their "learning to learn" knowledge, skills and dispositions over time, an obvious implication is that it would need to be capable of demonstrating progress. The creators of the tool certainly believe that such progress can in principle be defined and measured. As a consequence of many small action research studies with teachers they have:

> … realised that learning capacity can be expanded hugely over time, and that therefore we need some way of talking about progression. If Julie Green's ten-year-olds can begin to think productively about what makes a good scientific question, where can they go next? How can they strengthen, broaden and deepen that exploration even more? (Claxton, 2006, p.12).

The framing of progression in this example is congruent with Carr's ABCD dimensions of strength, but the question is essentially a *curriculum* question, with the NOS overtones outlined in the discussion of Quadrants Two and Three. Claxton advocates working with teachers and students to develop a sense of what can progress. Teachers can bring their practical knowledge of teaching students of different ages to bear. Students can look back on previous years – for example by being invited to shape advice for students younger than them for developing their learning capacity, based on their own growth experiences. He says, however, that such work is only just beginning, but the process advocated suggests a structure for an assessment tool such as the MOE benchmarks for literacy that were developed in this way (Ministry of Education, 2007b). The key difference, as noted earlier, is that this area of learning *about* learning is less familiar to most teachers.

The research of the ELLI team has also identified the "epistemic culture' of the school as an important determinant of whether or not students make progress in strengthening their learning competencies. They suggest the following as indicators of such a culture: a shared language for talking about learning[10]; teacher modeling of learning dispositions; activities are purposefully framed to stretch an aspect of learning capacity and indicators of progression are described; "split screen" conversations allow students to talk about the act of learning, not just what is learned; the contexts selected for learning genuinely engage students; the goal of stretching learning is made transparent to students; they are actively involved in conversations about how to learn more effectively; and transfer thinking is continually supported and modeled (Claxton, 2006).

Evidence for the presence of all of these factors can be described and documented. The ELLI team has designed three such tools: the actual ELLI tool described above; a tool that assesses students' perceptions of their teachers' learner-centred practices; and one that measures students' perception of their schools as emotionally literate places. The researchers sought patterns of inter-relationships between these measures and students' achievements in national assessments of

---

[10] The necessity for a shared language if key competencies are to be strengthened was a common theme in early-adopter schools when implementing NZC. We found this in two unrelated research projects (Boyd & Watson, 2006; Hipkins, Roberts, & Bolstad, 2007).

English, mathematics and science. They describe a "complex ecology" of learning that impacts on progress students can make (Crick, McCombs, Haddon, Broadfoot, & Tew, 2007).

The New Zealand-based longitudinal Competent Learners study has identified clear differences in the frequency of occurrence of many of the factors discussed in the ELLI studies when *both* students and teachers responded to similar item sets about students' self-nominated favourite and least favourite classes (Wylie, Hipkins, & Hodgen, 2008). The patterns found in the quantitative analysis suggested the possibility that *"opportunities to learn"* could be used as an empirical basis for accountability measures of these more complex outcomes. Factors developed from the individual items used in both teacher and student questionnaires closely reflected dimensions of key competencies and could be used as a starting point for school and teacher self-review instruments here. This is a different way of thinking about assessment for accountability, and widens the onus for making progress to all the dimensions of the learning setting, as is appropriate to the socio-cultural framing of this quadrant.

Illustrating the potential of taking a multi-faceted approach to opportunities to learn, the authors of the OECD study of social outcomes of learning describe dimensions that could be measured to gauge the likelihood of civic participation as an outcome of schooling. These include: links made between learning experiences in different curriculum areas, and between curricular and extra-curricular and out-of-school learning (i.e. Quadrant Two "connected" thinking); use of pedagogies that provide opportunities for students to work together in groups and develop effective team work strategies (i.e. aspects of Quadrant Three outcomes), and developing competencies for effective social interaction during experiences such as free and open discussion of political events, sharing of knowledge, and debating issues (i.e. Quadrant Four outcomes) (OECD, 2007a, p.63). The authors say these aspects collectively point to the importance of what they call the "ethos" of a school where students are confident to participate, feel their voice is valued, where strong citizenship norms are modeled and where school policies are congruent with these norms (e.g. the canteen policy models respect for healthy eating). (OECD, 2007a, p.64).

This analysis is an important reminder that the sum can be more than the parts, and ways of combining information about progress of the type described in the various quadrants need to be brought together into a meaningful whole. Since, practically speaking, not everything can be measured, an important question to ask is whether some factors are more likely to also be productive of others – in which case they should logically be the focus of attention. This is illustrated for the key competency "thinking" in the next section of the paper.

Making progress in thinking competencies

This section draws largely on a recent meta-analysis of approaches to teaching thinking (Harpaz, 2007). Harpaz cites 91 references in a systematic attempt to map the competency's field. Some of these would be familiar or at least known to many classroom teachers (De Bono, Costa, Gardner, Perkins, Dewey) while others are likely to be would be more familiar to researchers (Rogoff,

Sternberg, Postman, Bereiter, Lakoff and so on.). Harpaz describes three different approaches to teaching thinking, each aiming for quite different sorts of outcomes, as illustrated by the metaphor selected. These broadly map to Carr's quadrants as follows.

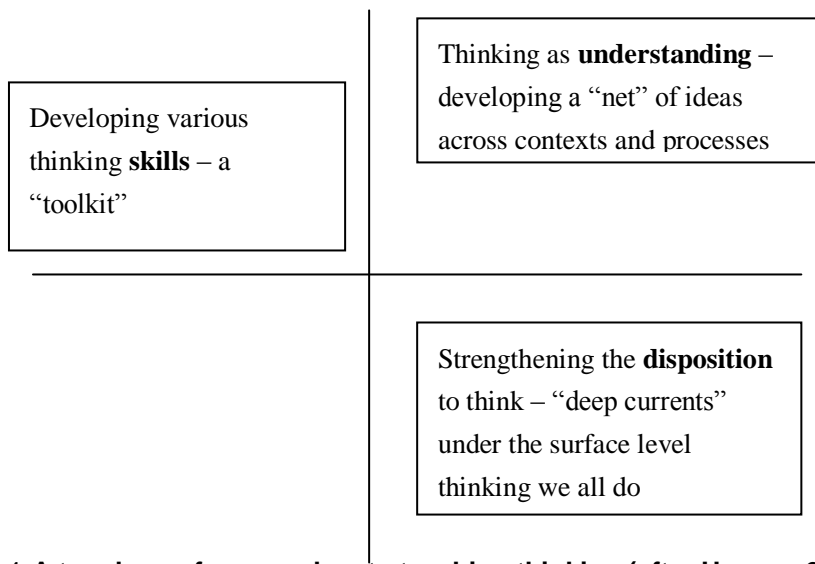| Developing various thinking **skills** – a "toolkit"	Thinking as **understanding** – developing a "net" of ideas across contexts and processes
	Strengthening the **disposition** to think – "deep currents" under the surface level thinking we all do

Figure 4 **A typology of approaches to teaching thinking (after Harpaz, 2007)**

Of these three possibilities, Harpaz locates teaching thinking for understanding (Quadrant Two) as the preferable approach on which to focus, while noting that it is the least likely focus currently. For him "thinking and understanding are inseparable" (Harpaz, 2007, p.1854). He reasons that skills have to be used for some curriculum purpose – you have to think about something and demonstrating skills in one context does not, on its own, imply that they can be transferred flexibly to other contexts. By contrast, the act of thinking to make connections that build a web of understandings is an act of individual cognition and therefore "student centered" to the extent that the web constructed is personally shaped and meaningful. The dispositions approach, he further argues, includes a metacognitive dimension as central to the intellectual character of a good thinker – a part of their way of "being". However:

> The understanding approach reasons that metacognition is possible, or at least generative, when it is equipped with a new understanding, through which former understandings are seen corrected, and improved. Metacognition, like cognition, cannot be an empty activity; it is always and necessarily bound with certain content and is valuable only when this content is understood (Harpaz, 2007, p.1859).

The clear implication here is that, for this key competency, making progress in linking concepts within and across disciplinary thinking frames would be an appropriate focus for assessment attention. Implementing this suggestion would however, be first and foremost a curriculum inquiry, because the nature of such connections and their relative ease or difficulty of construction would need to be described. A different issue is that the current structure of high stakes assessment (e.g. separate and stand-alone standards for NCEA assessment of any one course) does not easily allow for identification of webs of connections.

Harpaz proposes a distinction between "neutral" skills that everyone uses and that make thinking more *efficient* (identify, focus, classify, grade, discriminate, compare, select, generalize, summarize, ask, choose, assume, conclude, solve, decide) and "normative" skills that create or mould new ways of thinking. These include breaking conventional thinking patterns, devising problems, exposing basic premises, and discovering biases, especially in one's own thinking. It is the latter group of skills, he says, that actually differentiate "higher order" thinking and need to be moulded because they are culturally bound, while hierarchies that differentiate amongst neutral skills (such as Bloom's taxonomy) are misleading because skills are used in combination and are intertwined (Harpaz, 2007, p.1848). Although not the primary site of any of his three approaches, Quadrant Three characteristics and challenges are implicated here because an identification of the cultural features of normative thinking patterns becomes an aspect to be described before progress in moulding thinking could be measured.

Can NCEA be expected to reliably assess more uncertain outcomes?

Drawing the various threads of the discussion together, this final section of the paper considers implications for the NCEA. The standards-based assessment model that underpins NCEA was explicitly intended to widen the types of learning outcomes it would be possible to assess. However, in practice, this has proved to be easier said than done. The assessment focus for suites of achievement standards in many subjects has remained firmly in the "traditional academic" mould (Hipkins, Vaughan, Beals, & Ferral, 2004). In this situation, diversification of outcomes for different students has relied mostly on the addition of unit standards to the qualifications mix, but this has led to a hardening of existing academic/vocational binary thinking, with the result that the rigour of the qualification has been brought into question (Hipkins, Vaughan, with Beals, Ferral, & Gardiner, 2005). Can the analysis outlined in the earlier sections of this paper shed new light on this complex and vexed situation?

The first observation to be made is that topic-based achievement standards, developed in many subjects from previous examinations prescriptions, continue to assess Quadrant One knowledge outcomes - outcomes that we have had much more experience of reliably assessing using traditional tools of item collation, item analysis and, where necessary, making adjustments by scaling. These tools needed to be adapted for standards-based items, and NZQA continues to work on such system developments.

However the problematic pattern of unreliable judgments associated with some achievement standards is also likely to be a consequence of the way the standards were written in the first place. In the absence of the necessary *curriculum* conversations about what evidence of making progress would look like (the focus of this paper) descriptions were devised to differentiate progress-as-depth-of thinking in naïve Bloomian terms (description for achieve, compared to explanation for merit, compared to discussion for excellence). As we have seen Harpaz critiques this differentiation as naïve and misleading. For him, the neutral skills implied here are

intertwined and seldom used in isolation from each other. Additionally the most flexible thinkers will choose the most expedient tool for the job –which may be of a so-called "lower order" than the skills a less experienced thinker might attempt to use, at more personal effort, and with less successful results! If we accept his point, there was always going to be a *structural* difficulty with trying to differentiate progress in deeper understanding on these terms.

Elsewhere, responding to the critique of a specific NCEA assessment by a subject expert, I have identified an issue that arises when a problematic mismatch occurs between the overly general "content" typically taught in school science and the necessity to set assessment questions in specific contexts if the task is to allow for meaningful explanation or discussion (Hipkins, 2007). If any students are disadvantaged by such a mismatch (which may or may not be the case) they will be those who have the deepest knowledge of the subject area, as demonstrated by being able to make rich contextual links in the first place, rather than simply repeating what has been taught. The concept/context interactions discussed for Quadrant Three also come into play here, as does Harlen's identification of context as source of unreliability in written assessments.

In any case, the exemplification of the descriptors in the tasks devised to assess the standards then tended to be misinterpreted as more traditional acquisition of a greater amount of content, some of it relatively more difficult for the age level concerned. In this case, what was intended to be a *qualitative* difference in achievement levels has actually been described in pseudo-quantitative terms, with the resultant well documented difficulties in reliably drawing boundaries between levels.

Could the curriculum questions about progression, as outlined above for Quadrants 2-4, help reframe ways of thinking about standards, so that more reliable descriptors and benchmarks could be constructed? In the light of this question, Harpaz's argument for fostering the development of thinking skills and dispositions by teaching for "deeper understanding" seems a promising place to begin. Thinking is one of the five key competencies and this would be the least controversial starting point for development in cases where most teachers in a learning area hold strongly to traditional academic views of desirable outcomes in their discipline area. (Science and mathematics spring to mind here.) In that case, all the questions about progression outlined in Quadrant Two could be brought to bear.

However the atomized nature of achievement standards, *if these continue to be topic based*, would be a barrier to making the sorts of rich connections that both Harpaz and Zohar describe as the hall mark of deeper and more flexible understanding. This is not insurmountable because individual standards could be reshaped in ways that assess ability to make rich and deep connections, but in that case, the first conversation to be had is a *curriculum* one. In science for example, new standards could be developed to assess outcomes suggested by the NOS strand achievement objectives, drawing on contexts of the science discipline(s) as appropriate. Such a process suggests many possibilities but they would need to be carefully worked through and widely understood by the teachers who would need to implement the standards so devised.

Development of some newer ARB items also suggests that evidence of richer and more flexible connections could be sought if the key competency "using language, symbols and texts" is interpreted within a multi-literacies framework, especially where the discipline-specific features of texts, mentioned as important tools for deeper thinking by both Harpaz and Zohar, are the focus of the assessment. For example, ARB trials at year 10 suggest a range of demanding task types where progression might be described. These include: converting one text type to another (written text to a flow chart say); interpreting the key conceptual message in an unfamiliar text (a cartoon intended to convey a science idea for example); comparing both intended and unintended messages in two different visual texts of the same concept and so on. Trials have demonstrated that conventionally sought conceptual understandings are very clearly revealed by such questions, which could help ameliorate teacher concerns that "content" is being neglected. Supporting the findings of our more limited samples, the 2006 PISA assessment, aimed at 15 year olds, placed drawing a specific aspect of meaning from across two different but complementary graphs as one of the most difficult competencies on its science scale (OECD, 2007b).

In some learning areas standards shaped to assess deeper understanding do already exist. My hypothesis is that such standards can already be identified by their more reliable performance, as this is now able to be measured by NZQA. For example, a standard that assesses ability to *use* a language in conversation in a real context has proved to be consistently reliable in discriminating between progress at achieve, merit and excellence levels, ever since its introduction.[11] The act of communicating fluently and persuasively requires the speaker to draw on many connections - of vocabulary, grammatical conventions, idiom and so on – and to synthesise these into a meaningful whole in the moment of speaking. A progression of comparative levels of success in doing so can be readily constructed and understood. A systematic analysis of the *curriculum* features of other standards that also perform reliably would test my hypothesis, and also point towards new outcome possibilities for subject areas that need to redevelop some or all standards.

Bearing in mind the intention to widen the types of outcomes assessed, Quadrant Three also offers many possibilities. Again achievement standards in some subjects already assess elements of the rich "pieces" of learning that can be described here, and some learning areas already use methods of combining such pieces into a unified framework of achievement – the portfolios developed by visual arts students spring to mind. There are however, issues of validity to address in some cases. For example, in the interests of establishing the "authenticity" of students' work, and in the absence of a wider conversation about assessment of group outcomes, evidence of action competence may be reduced to writing about the process after-the-fact, within an academic framing, under examination conditions, notwithstanding the existence of concrete evidence of the ability to act successfully to meet planned targets. This seems regrettable, given the close match between the active and personally meaningful nature of learning for action competence and the key competency "participating and contributing". The so-called "intellectualisation" of some

---

[11] Personal conversation with NZQA analysts.

subjects that come under the umbrella of the technology learning area signals a similar set of issues.

Currently, much of the "research" that students undertake for NCEA is a somewhat hollow generic exercise in information retrieval and repackaging, and achievement standards that assess this aspect of curriculum are likely to be amongst the first that teachers drop from their courses as they seek to address time pressures (Hipkins, 2006). Arguably, success in developing better models for assessing the actual carrying out of action competence projects, or indeed any other model of genuine inquiry,[12] would deliver less tangible Quadrant Four outcomes as well as the benefits for greater validity and reliability already described. Students who experience success in addressing issues are more likely to develop a view of themselves as people who are not powerless to take constructive action later on – i.e. there is a link to developing the disposition for active citizenship in adult life, identified as so important by the OECD (OECD, 2007a). Given the necessary conversations about what making progress could look like, this sort of learning could be documented. For example the framing of our "identities" as the stories we tell about ourselves to ourselves and to others suggests immediate possibilities for self assessment (as suggested by the ELLI tool for example) which could be backed up with documented evidence. Again arts portfolios provide a potential model here.

Summing up

None of the potential benefits outlined in this paper will be realised in the absence of a deep curriculum conversation about outcomes of learning, about links between desired outcomes and actual teaching and learning, and about the nature of evidence of deep learning. In NZC we potentially have the curriculum tools we need – for all levels of learning including the senior secondary school. Given the strong signals of what is actually valued that undoubtedly teachers get from assessment systems (Harlen, 2007), the primary challenge for the New Zealand Assessment Framework going forward would seem to be to make a space for these deep conversations about the meaning and potential of the curriculum framework, and to establish collaborative inquiries in to the nature of progression in a wider range of outcomes, perhaps following the model of the recently completed literacy progressions work where teachers and experts work together. If the timeframe allowed, such inquiry could evolve from the current NCEA standards review. While this is currently the most high-stakes assessment carried out in New Zealand, conversations about *curriculum* are needed to develop new tools, and to continue the evolution of existing tools, at all levels of the school system.

---

[12] Addressing what Harpaz would characterise as 'fertile questions" (Harpaz, 2005). Features of such questions were recently described by Claxton, in the content of discussing experiences likely to foster "learning to learn" (Claxton, 2006).

References

Aikenhead, G. (1997). *A framework for reflecting on assessment and evaluation.* Paper presented at the Globalization of Science Education: International Conference on Science Education., Seoul, Korea. May 26-30.

Barker, M., Hipkins, R., & Bartholomew, R. (2004). *Reframing the essential skills: Implications for and from the science curriculum*: Report commissioned by Ministry of Education, July 2004.

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: the SOLO taxonomy*. New York: Academic Press.

Boyd, S., & Watson, V. (2006). *Shifting the frame: Exploring integration of the Key Competencies at six Normal Schools*. Wellington: New Zealand Council for Educational Research.

Carr, M. (2006). Dimensions of strength for key competencies. Retrieved 25 April, 2008, from http://nzcurriculum.tki.org.nz/curriculum_project_archives/developing_the_draft/key_competencies/background_reading

Carr, M. (2008, 13 March 2008). *Zooming in and Zooming Out: Challenges and choices in discussions about making progress.* Paper presented at the NZCER Conference: Making Progress - Measuring Progress, Wellington, March 2008.

Claxton, G. (2006). *Expanding the capacity to learn: A new end for education?* Paper presented at the British Education Research Association (BERA), Warwick, September 6.

Crick, R., McCombs, B., Haddon, A., Broadfoot, P., & Tew, M. (2007). The ecology of learning: factors contributing to learner-centred classroom cultures. *Research Papers in Education, 22*(3), 267-307.

Darr, C., & McDowall, S. (2008). *Standardised testing: Dilemmmas and possiblities.* Paper presented at the NZCER Conference: Making Progress - Measuring Progress, Wellington, March 2008.

Davis, B., Sumara, D., & Luce-Kapler, R. (2008). *Engaging minds: Changing teaching in complex times. Second Edition*. New York and London: Routledge.

Delandshere, G., & Petrosky, A. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher, 27*(2), 14-24.

Delors, J. (1996). *Learning: The Treasure Within*. Paris: UNESCO.

Gilbert, J. (2005). *Catching the Knowledge Wave? The Knowledge Society and the future of education.* Wellington: NZCER Press.

Harlen, W. (2007). *Assessment of Learning*. London, Thousand Oaks, New Dehli, Singapore: Sage Publications.

Harpaz, Y. (2005). Teaching and learning in a community of thinking. *Journal of Curriculum and Supervision, 20*, 136-157.

Harpaz, Y. (2007). Approaches to teaching thinking: towards a conceptual mapping of the field. *Teachers College Record 109*(8), 1845-1874.

Hipkins, R. (2006). *Learning to do research: Challenges for students and teachers*. Wellington: NZCER Press.

Hipkins, R. (2007). A closer look at the content/context dilemma. *New Zealand Science Teacher, 115*, 30-32.

Hipkins, R., Bull, A., & Joyce, C. (2008). The interplay of contexts and concepts in primary school children's systems thinking. *Journal of Biological Education, 42*(2), 73-77.

Hipkins, R., Roberts, J., & Bolstad, R. (2007). *Kick Starts series: Key competencies, the journey begins*. Wellington: NZCER Press.

Hipkins, R., Vaughan, K., Beals, F., & Ferral, H. (2004). *Shared pathways and multiple tracks. Interim research report - second year of Learning Curves, Meeting student learning needs in an evolving qualifications regime.* Wellington: New Zealand Council for Educational Research.

Hipkins, R., Vaughan, K., with Beals, F., Ferral, H., & Gardiner, B. (2005). *Shaping our futures: Meeting secondary students' learning needs in a time of evolving qualifications*. Wellington: New Zealand Council for Educational Research. http://www.nzcer.org.nz/pdfs/14349.pdf.

Kegan, R. (1994). *The Evolving Self: Problem and Process in Human Development.* Boston: Harvard University Press.

Kutzler, B. (2003). CAS as pedagogical tools for teaching and learning mathematics. In J. Fey, A. Cuoco, C. Kieran, L. McMullin & R. Zbiek (Eds.), *Computer algebra systems in secondary school mathematics education* (pp. 53-71). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Matters, G. (2006). *Australian Education Review: Using data to support learning in schools: Students, teachers and systems*. Melbourne: Australian Council for Educational Research.

Ministry of Education. (2007a). *The New Zealand Curriculum*. Wellington: Learning Media.

Ministry of Education. (2007b). Literacy learning progressions: Meeting the reading and writing demands of the curriculum. Retrieved 28 April, 2008, from http://www.literacyprogressions.org.nz/

Neill, A., & Maguire, T. (2006). *An evaluation of the CAS Pilot Project. Evaluations of the 2005 Secondary Numeracy Pilot project and the CAS Pilot Project*. Wellington: Ministry of Education.

OECD. (2003). PISA 2006 Scientific Literacy Framework. Retrieved June 17 2008, from http://www.fed.cuhk.edu.hk/~hkpisa/sample/files/2003_Sci_Framework.pdf

OECD. (2007a). *Understanding the Social Outcomes of Learning*. Paris: OECD.

OECD. (2007b). PISA 2006 Science Competencies for Tomorrow's World.   Retrieved June 2007, from
http://www.pisa.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html

Rivers, J., Wylie, C., Hodgen, E., Ferral, F., Dingle, R., Thompson, J., et al. (2006). *Growing independence : A summary of key findings from the Competent Learners @14 project*. Wellington: Ministry of Education.

Robertson, J. (2005). *Making sense of health promotion in context of health and physical education curriculum learning*. Wellington: Ministry of Education.

Rychen, D., & Salganik, L. (Eds.). (2003). *Key competencies for a successful life and a well-functioning society*. Cambridge, MA, USA: Hogrefe and Huber.

Sumara, D., & Davis, B. (2006). Correspondence, coherence, complexity: Theories of learning and their influences on processes of literary composition. *English teaching: Practice and Critique, 5* (2), 34-55

Wylie, C., Hipkins, R., & Hodgen, E. (2008). *On the edge of adulthood: young people's school and out-of-school experiences at 16.* Wellington: Ministry of Education.

Zohar, A. (2006). Connected knowledge in science and mathematics education. *International Journal of Science Education, 28*(13), 1579-1599.